Sparse Bayesian learning, beamforming techniques and asymptotic analysis for massive MIMO

Christo Kurisummoottil Thomas

Sorbonne Université

École Doctorale

Doctorat, Département Systèms de Communication

CHRISTO KURISUMMOOTTIL THOMAS

SPARSE BAYESIAN LEARNING, BEAMFORMING TECHNIQUES AND ASYMPTOTIC ANALYSIS FOR MASSIVE MIMO

Thèse dirigée par Dirk Slock, Professeur, EURECOM, France

Successfully defended on the 28 September 2020, before a committee composed of:

Rapporteur	**Prof. Martin HAARDT**	Ilmenau University of Technology, Germany
Rapporteur	**Prof. Antti TÖLLI**	University of Oulu, Finland
Jury	**Prof. David GESBERT**	EURECOM, France
Jury	**Prof. Laura COTTATELLUCCI**	Friedrich-Alexander University, Germany
Jury	**Prof. Mérouane DEBBAH**	CentraleSupélec, France
Jury	**Prof. Dirk SLOCK**	EURECOM, France

ABSTRACT

Multiple antennas at the base station side can be used to enhance the spectral efficiency and energy efficiency of the next generation wireless technologies. Indeed, massive multi-input multi-output (MIMO) is seen as one promising technology to bring the aforementioned benefits for fifth generation wireless standard, commonly known as 5G New Radio (5G NR). In this monograph, we will explore a wide range of potential topics in multi-user MIMO (MU-MIMO) relevant to 5G NR,

- Sum rate maximizing beamforming (BF) design and robustness to partial channel state information at the transmitter (CSIT)
- Asymptotic analysis of the various BF techniques in massive MIMO and
- Bayesian channel estimation methods using sparse Bayesian learning.

While massive MIMO has the aforementioned benefits, it makes the acquisition of the channel state information at the transmitter (CSIT) very challenging. Since it requires large amount of uplink (UL) pilots for channel estimation phase. Moreover, each antenna has associated with a radio frequency (RF) chain which in turn leads to high power consumption and hardware complexity at the base station (BS) side. One promising technology to overcome these issues is to utilize a hybrid beamforming (HBF) system. In HBF, the number of RF chains at the transmitter side is reduced significantly compared to number of antennas. Hence, it involves a two stage beamforming scheme. With the analog BF generates multiple beams in the spatial domain and thereby providing BF gain. The digital BF is used at the baseband for multiplexing the different user streams across the beams generated by the analog BF. Analog beamforming is implemented at the RF chain using phase shifters. One of our main focus in thesis is to propose efficient phase shifter design which can attain performance very close to that of the fully digital BF systems. For this purpose, we proposed an efficient scheme for analog phasor design using the technique of deterministic annealing.

Fully digital BF scheme becomes a special case of our HBF design and further for the performance analysis, we focus on fully digital BF schemes itself. In a fully digital massive MIMO system, it is important to consider low complexity BF solutions. With this direction in mind, we proposed a low complexity but close to optimal (linear minimum mean square error-LMMSE) BF solution termed as reduced order zero forcing (ZF). However, it is quite incomplete if we stop with the various BF designs, we do require extensive theoretical analysis to evaluate the spectral efficiency (SE) behaviour of the massive MIMO system which we consider in the next part of the thesis.

In the past decade, several academic research has been conducted on the asymptotic/large system analysis of massive MIMO systems. Large system analysis helps to avoid tedious Monte-Carlo simulations to evaluate the SE and provide simplified rate expressions as a function of the very few parameters such as channel second-order statistics, antenna dimensions and channel estimation error variance, etc. However, the majority of the existing research focus on simplified Rayleigh channel models or multiple of identity channel covariance matrices for different users to simplify the analysis. Moreover, those works which exploit distinct spatial channel covariance matrices for users lead to highly cumbersome expressions which are not intuitive and hence not much of use.

Motivated by the above issues, we propose a stochastic geometry inspired randomization of the user covariance subspaces (which indeed has strong intuitive justification under large system dimensions and for millimeter wave or massive MIMO systems). Our simplifications indeed lead to very intuitive and elegant rate expressions and hence that forms one of the landmark contributions in this thesis. We provide large system results for an upper bound of the expected weighted sum rate and several other suboptimal BF schemes under partial CSIT. Moreover, we analyze the SE under different channel estimation schemes such as least squares, LMMSE and subspace projected channel estimate. However, it has to be noted that the LMMSE or subspace projected channel estimates which give superior performance need the knowledge of the user covariance subspace (low rank). Note that we do not consider explicitly an estimation method for this subspace information. But we remark that our variational Bayesian inference techniques which form the final part of the thesis can be an efficient and accurate method to estimate the pathwise components in the MIMO channel. These pathwise information can be finally utilised to form an accurate estimate of the user channel covariance subspace.

Finally, we also looked at a Bayesian approach to sparse signal recovery problem. The sparse states can be considered to be either static or dynamic (where the temporal correlation is chosen to be modeled as an autoregressive process). Sparse Bayesian learning (SBL) algorithm focuses on formulating an appropriate hierarchical prior which can be modeled effectively the sparsity properties of the underlying signal. We consider a joint sparse signal plus hyperparameter (associated with the prior) estimation algorithm, which relies on variational Bayesian inference based methods. Here the motivation is to achieve lower complexity without sacrificing much on the signal recovery performance. One of the several applications of the SBL algorithm is in massive MIMO or milli meter wave channel estimation where the underlying wireless channel is sparse in angular or Doppler or delay domains. Apart from this, several applications exist in diverse fields (not only communication systems) such as data science or medical imaging, and hence this topic we considered assumes greater relevance.

A majority of the topics considered here indeed form a relevant contribution to massive MIMO research community both by the proposition of new innovative algorithms and theoretical analysis. However, much remains to be pursued and we hope that this throws up a few open questions too which can inspire at least a few others to follow the road not taken yet.

RÉSUMÉ

Plusieurs antennes du côté de la station de base peuvent être utilisées pour améliorer l'efficacité spectrale et l'efficacité énergétique des technologies sans fil de nouvelle génération. En effet, Multi-Input Multi-Output (MIMO) massif est considéré comme une technologie prometteuse pour apporter les avantages de la norme sans fil de cinquiéme génération, communément appelée 5G New Radio (5G NR).

Dans cette monographie, nous explorerons un large éventail de sujets potentiels en multi-utilisateurs MIMO (MU-MIMO) concernant la 5G NR,

- Conception de Techniques de Précodage Multi-Antenne maximisant la somme des débits et la robustesse à l'imprécision des connaissances partielles du canal au transmetteur (CSIT).
- Analyse asymptotique des différentes techniques de Précodage Multi-Antenne en Systémes MIMO Massifs et
- Méthodes d'estimation de canal Bayésien utilisant un apprentissage Bayésien Parcimonieux.

Bien que le MIMO massif présente les avantages susmentionnés, il permet l'acquisition de la connaissance du canal au transmetteur (CSIT) trés difficile. Puisqu'il nécessite une grande quantité de pilotes de liaison montante (UL) pour la phase d'estimation de canal. De plus, chaque antenne est associée à un chaîne de radiofréquence (RF) qui à son tour conduit à une consommation d'énergie élevée et à une complexité matérielle côté station de base (BS). Une technologie prometteuse pour surmonter ces problémes consiste à utiliser un systéme Hybride Techniques de Précodage (HBF). Dans HBF, le nombre de chaînes RF à l'émetteur côté est considérablement réduit par rapport au nombre d'antennes. Par conséquent, il s'agit d'une étape en deux schéma de précodage. Avec le BF analogique génére plusieurs faisceaux dans le domaine spatial et fournissant ainsi un gain BF. Le BF numérique est utilisé en bande de base pour multiplexer les différents l'utilisateur diffuse sur les faisceaux générés par le BF analogique. La formation de faisceaux analogique est mise en œuvre au niveau de la chaîne RF à l'aide de déphaseurs. L'un de nos principaux objectifs de thèse est de proposer une phase efficace conception de levier de vitesses qui peut atteindre des performances trés proches de celles des systémes BF entiérement numériques. Pour à cet effet, nous avons proposé un schéma efficace pour la conception de phaseurs analogiques utilisant la technique de recuit déterministe.

Le schéma BF entiérement numérique devient un cas particulier de notre conception HBF et, plus loin, pour l'analyse des performances, nous nous concentrons sur les schémas BF entiérement numériques eux-mêmes. Dans un systéme MIMO massif entièrement numérique, il est important d'envisager des solutions BF de faible complexité. Avec cette direction à l'esprit, nous

avons proposé une solution BF de faible complexité mais proche de la solution optimale (erreur quadratique moyenne minimale linéaire-LMMSE) appelée forçage d'ordre réduit (ZF). Cependant, il est assez incomplet si nous nous arrêtons aux différentes conceptions de BF, nous avons besoin d'une analyse théorique approfondie pour évaluer le comportement d'efficacité spectrale (SE) du systéme MIMO massif que nous considérons dans la partie suivante de la thése.

Au cours de la derniére décennie, plusieurs recherches universitaires ont été menées sur l'analyse asymptotique / des grands systémes de systémes MIMO massifs. L'analyse des grands systémes permet d'éviter les simulations Monte-Carlo fastidieuses pour évaluer le SE et de fournir des expressions de taux simplifiées en fonction des trés rares paramétres tels que les statistiques de second ordre de canal, les dimensions de l'antenne et la variance d'erreur d'estimation de canal, etc. Cependant, la majorité des recherches existantes se concentrent sur des modéles de canaux de Rayleigh simplifiés ou sur de multiples matrices de covariance de canaux d'identité pour différents utilisateurs afin de simplifier l'analyse. De plus, ces travaux qui exploitent des matrices de covariance de canal spatial distinctes pour les utilisateurs conduisent à des expressions trés lourdes qui ne sont pas intuitives et donc peu utiles.

Motivés par les problémes ci-dessus, nous proposons une randomisation inspirée de la géométrie stochastique des sous-espaces de covariance utilisateur (qui a en effet une forte justification intuitive sous de grandes dimensions de système et pour des systémes MIMO à ondes millimétriques ou massives). Nos simplifications conduisent en effet à des expressions de taux trés intuitives et élégantes, ce qui constitue donc l'une des contributions marquantes de cette thése. Nous fournissons de grands résultats de systéme pour une limite supérieure du taux de somme pondéré attendu et plusieurs autres schémas BF sous-optimaux sous CSIT partiel. De plus, nous analysons le SE sous différents schémas d'estimation de canal tels que les moindres carrés, LMMSE et l'estimation de canal projetée sous-espace. Cependant, il convient de noter que les estimations de canal projetées LMMSE ou sous-espace qui donnent des performances supérieures nécessitent la connaissance du sous-espace de covariance utilisateur (rang bas). Notez que nous ne considérons pas explicitement une méthode d'estimation pour ces informations de sous-espace. Mais nous remarquons que nos techniques d'inférence Bayésienne variationnelle qui forment la derniére partie de la thése peuvent être une méthode efficace et précise pour estimer les composantes pathwise dans le canal MIMO. Ces informations de cheminement peuvent finalement être utilisées pour former une estimation précise du sous-espace de covariance du canal utilisateur.

Enfin, nous avons également examiné une approche Bayésienne du probléme de récupération de signaux parcimonieux. Les états clairsemés peuvent être considérés comme statiques ou dynamiques (où la corrélation temporelle est choisie pour être modélisée comme un processus autorégressif). L'algorithme d'apprentissage Bayésien parcimonieux (SBL) se concentre sur la formulation d'un prior hiérarchique approprié qui peut être modélisé efficacement les propriétés de parcimonie du signal sous-jacent. Nous considérons un algorithme d'estimation conjoint signal clairsemé plus hyperparamétre (associé à l'ancien), qui repose sur des méthodes basées sur l'inférence Bayésienne variationnelle. Ici, la motivation est de réduire la complexité sans sacrifier beaucoup les performances de récupération du signal. Une des nombreuses applications de l'algorithme SBL est dans l'estimation massive de canal d'onde MIMO ou milli-métre où le canal sans fil sous-jacent est clairsemé dans les domaines angulaires ou Doppler ou de retard. En dehors de cela, plusieurs applications existent dans divers domaines (pas seulement les systémes de communication) tels que la science des données ou l'imagerie médicale, et par conséquent, ce sujet que nous avons considéré revêt une plus grande pertinence.

Une majorité des sujets abordés ici constituent en effet une contribution pertinente à la communauté de recherche massive du MIMO à la fois par la proposition de nouveaux algorithmes in-

novants et l'analyse théorique. Cependant, il reste encore beaucoup à faire et nous espérons que cela soulévera également quelques questions ouvertes qui pourront inspirer au moins quelques autres à suivre la voie qui n'est pas encore prise.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

TABLE **Page**

ABBREVIATIONS

The acronyms and abbreviations used throughout the manuscript are specified in the following. They are presented here in their singular form, and their plural forms are constructed by adding and s, for example, TX (transmitter) and TXs (transmitters). The meaning of an acronym is also indicated the first time that it is used.

3-D	Three Dimensional
3GPP	Third generation partnership project
5G	Fifth generation
ADC	Analog to Digital Converter
AoA	Angle of Arrival
AoD	Angle of Departure
AMP	Approximate Message Passing
AR	Autoregressive
BC	Broadcast Channel
BF	Beamforming or Beamformer
BP	Belief Propagation
BS	Base Station
CP	Cyclic Prefix
CPD	Canonical Polyadic Decomposition
CRB	Cramer-Rao Bound
CS	Compressed Sensing
CSI	Channel State Information
CSIT	Channel State Information at the Transmitter
CSIR	Channel State Information at the Receiver
CoCSIT	Covariance only CSIT
DA	Deterministic Annealing
DAC	Digital to Analog Converter
DCP	Difference of Convex functions Programming
DL	Downlink
DPC	Dirty Paper Coding
EM	Expectation Maximization
EP	Expectation Propagation
ESIP-WSR	Expected Signal and Interference Power based WSR
EWSMSE	Expected Weighted Sum Mean Squared Error
EWSR	Expected WSR
FD	Full Duplex
FDD	Frequency Division Duplexing
FG	Factor Graph
FIM	Fisher Information Matrix
FIR	Finite Impulse response
FFT	Fast Fourier Transform
GAMP	Generalized Approximate Message Passing
GSwAMP	Generalized Swept Approximate Message Passing
i.i.d.	Independent and identically distributed
IA	Interference Alignment
IBC	Interfering Broadcast Channels
iCSIT	Instantaneous CSIT
KF	Kalman Filtering
KL	Kullback Leibler
KLD	KL Divergence
KS	Kronecker Structured
LDR	Limited Dynamic Range
LoS	Line of Sight
LMMSE	Linear MMSE
LS	Least Squares
LSL	Large System Limit
LSA	Large System Approximation
LTE	Long Term Evolution
LTI	Linear time-invariant

MaMIMO	Massive MIMO
MaMISO	Massive MISO
mCRB	Mismatched CRB
MF	Matched Filter
MF	Mean Field
MP	Message Passing
MIMO	Multiple-Input Multiple-Output
MISO	Multiple-Input Single-Output
ML	Maximum Likelihood
MMSE	Minimum Mean Squared Error
mmWave	Millimeter Wave
MRC	Maximum Ratio Combining
MRT	Maximum Ratio Transmission
MSE	Mean Squared Error
MU	Multiuser
MU-MIMO	Multiuser MIMO
NLOS	Non-LoS
NMSE	Normalized MSE
NR	New Radio
OFDM	Orthogonal Frequency-Division Multiplexing
OTA	Over-the-air
pwCSIT	Pathwise CSIT
RCMM	Reciprocity Calibration for Massive MIMO
RF	Radio Frequency
RO-ZF	Reduced Order Zero Forcing
Rx	Receiver
R-ZF	Regularized ZF
SAVE	Space Alternating Variational Estimation
SBL	Sparse Bayesian Learning
SE	Spectral Efficiency
SI	Self-interference
SIMO	Single Input Multiple Output
SINR	Signal to Interference plus Noise Ratio
SNR	Signal to Noise Ratio
SoA	State of the Art
SSR	Sparse Signal Recovery
TDD	Time Division Duplexing
Tx	Transmitter
UE	User Equipment
UL	Uplink
ULA	Uniform Linear Array
VAMP	Vector Approximate Message Passing
VB	Variational Bayesian
WiMAX	Worldwide Interoperability for Microwave Access
WSMSE	Weighted Sum Mean Squared Error
WSR	Weighted Sum Rate
xAMP	Variants of AMP
ZF	Zero-Forcing

NOTATIONS

The next list describes an overview on the notation used throughout this manuscript. We use boldface uppercase letters ($\mathbf{X}$) for matrices, boldface lowercase letters for vectors ($\mathbf{x}$), and regular lowercase letters for scalars (x).

M, N_t	Represents the number of antennas at the base station side for fully digital and hybrid beamforming scenarios, respectively
M	Represents the number of RF chains at the base station for hybrid beamforming scenarios
n	Represents the subcarrier number
K, U	Represents the number of users in the network and in a particular cell, respectively
L	May represent either the rank of the channel covariance or the number of multipaths
$\mathbf{h}_{k,b_i}, \mathbf{H}_{k,b_k}$	Represent the MISO or MIMO channel matrix from UE k to base station b_i
b_k	Represents the base station to which user k is associated
$(x)^+$	Denotes $\max(x, 0)$
$p(x\|y)$	Represents the conditional distribution of x given y
$\mathscr{C}^M$	Represents the $M-$dimensional complex space
$\mathscr{R}^M$	Represents the $M-$ dimensional real space
$\mathbf{X}^H$	Conjugate transpose of matrix $\mathbf{X}$
$\mathbf{X}^T$	Transpose of matrix $\mathbf{X}$
$\mathbf{X}*$	Conjugate of a matrix $\mathbf{X}$
$\mathbf{X}^{1/2}$	Denotes the Hermitial Square root of any complex matrix $\mathbf{X}$
$\mathbf{x} \sim \mathscr{CN}(\boldsymbol{\mu}, \boldsymbol{\Theta})$	A circularly symmetric complex Gaussian random vector $\mathbf{x}$ with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Theta}$
$x \sim \mathscr{N}(\mu, \theta)$	A Gaussian random variable x with mean μ and variance θ
ln	Natural logarithm
arg max	Points or elements of the domain of some function at which the function values are maximized
arg min	Points or elements of the domain of some function at which the function values are minimized
$\mathbf{V}_{max}(\mathbf{A}, \mathbf{B})$ or $\mathbf{V}_{1:M}(\mathbf{A}, \mathbf{B})$	(Normalized) dominant generalized eigenvector or the matrix formed by M (normalized) dominant generalized eigenvectors of $\mathbf{A}$ and $\mathbf{B}$
$\lambda_{max}(\mathbf{A})$	Max eigen value of matrix $\mathbf{A}$
$vec(\mathbf{X})$	Vector obtained by stacking each of the columns of $\mathbf{X}$
$unvec(\mathbf{X})$	Matrix obtained by stacking the elements of the vector as columns of a matrix
$\mathbf{A} \succeq 0$	$\mathbf{A}$ is a positive semi-definite matrix
diag $(\mathbf{x})$	Diagonal matrix created by elements of a row or column vector $\mathbf{x}$
$< x >$	Represents the expectation of x with respect to the approximate posterior q
$\|\mathbf{X}\|$	Frobenius norm of $\mathbf{X}$
tr$\{\mathbf{X}\}$	Trace of a square matrix $\mathbf{X}$
E$[\cdot]$	Expectation operator
$\mathbf{x}_{\tilde{i}}$	Represents the vector $\mathbf{x}$ with i^{th} component removed
$\mathbf{x}_{i-}$	Represents the vector obtained from $\mathbf{x}$ with elements from 1 till $i-1$
$\mathbf{x}_{i+}$	Represents the vector obtained from $\mathbf{x}$ with elements from $i+1$ till end
$A_{i,j}$	Represents the $(i,j)^{th}$ element of the matrix $\mathbf{A}$
$\mathbf{e}_i$	Termed as the unit vector or the vector with all zero except i^{th} entry which is 1
$\mathbf{0}$	Represents a vector or a matrix (can be inferred from the context) of all zeros
$\mathbf{1}$	Represents a vector or a matrix (can be inferred from the context) of all ones
$\mathscr{X}$	The calligraphic font is used to represent any tensor.

Acknowledgement

For me, an aspiration to do a PhD started back in 2012, while doing my master's thesis with Professor P. Vijay Kumar at Indian Institute of Science (IISc), Bangalore. Moreover, my course work at IISc on telecommunication engineering was deeply rooted in mathematical concepts such as linear algebra, probability theory and random processes. This helped me to develop strong interests towards communication theory, which involves mathematical optimization to solve the basic problems therein. I am deeply indebted to the Professors at IISc for inspiring us towards the endless possibilities in communication technology. Their attitude and passion towards the work were a great inspiration for me to continue working on the communication research.

I was lucky enough to find a PhD position with Eurecom in 2017, after working in communication chip industry for 4.5 years. During the time at Intel, Bangalore, I was lucky to have been friends with Puneet and Murthy, who had also motivated me to pursue my dreams forward. Foremost, I would like to express my sincere gratitude to my supervisor Prof. Dirk Slock for guiding me in each of the contributions of this thesis and for his patience, motivation and immense knowledge. He was always available to discuss my research work and provided constructive criticism. His rich experience and expertise on massive MIMO technologies and statistical signal processing helped me to improve my understanding of the subject matter. I also really enjoyed doing lab or problem sessions to the master's students for the courses statistical signal processing (SSP) and signal processing for communication (SP4COM). I am also excited that i was able to present my thesis work in front of an esteemed jury comprising Prof. David Gesbert, Prof. Laura Cottatellucci, Prof. Mérouane Debbah and my reviewers Prof. Martin Haardt and Prof. Antti TÖLLI. Prof. Martin Haardt was kind enough to provide me the opportunity to interact with him during the thesis feedback sessions conducted through video conference. All his remarks on my thesis has helped immensely to improve this manuscript to the current state. I was also extremely lucky to have collaborated with Prof. Bruno Clerckx at Imperial College London and Prof. Luca Sanguinetti at University of Pisa, Italy, on the topic of rate splitting.

And I am always indebted to my parents whose love and prayers are with me in whatever I pursue. My gratitude also goes to my wife Jasmy, who was very supportive during the last three years of our stay in France. Also, grateful to my friends around here, Kalyan and his family, Chandan, Imene, Rakesh and others who have been very supportive during the time at Sophia Antipolis.

Part I

Motivation and Background

Chapter 1

INTRODUCTION

During the last three years, we got to know and contribute on a plethora of topics related to signal processing for next generation wireless communications. In short, we focus on three aspects of the research on massive multiple-input multiple-output (MIMO) or millimeter Wave (mmWave) systems, which are futuristic technologies, or novel paradigms in multi-antenna signal processing.

- Hybrid beamforming (BF) techniques in massive MIMO (MaMIMO),
- Asymptotic analysis of the various BF techniques.
- MaMIMO channel estimation using sparse Bayesian learning.

Hybrid beamforming (HBF) is a promising solution to reduce the complexity of a multi-antenna system by employing a reduced number of radio frequency (RF) chains compared to the number of antennas. In a multi-cell multi-user MaMIMO system, we propose HBF design, which maximizes the sum of the spectral efficiency across all the users. Unit magnitude constraints on the analog BF makes the problem highly non-convex and hence challenging to solve. Our objective here is to avoid the issue of local optima and the over-dependence on the initialization, which is plagued by the alternating optimization of the analog phasors in the conventional designs. As an efficient solution, we introduce an innovative solution based on the concept of deterministic annealing in machine learning. Further, we consider HBF design for more realistic scenarios like per-antenna or per-RF power constraints (which is the first time in the literature on HBF). Our algorithms perform significantly better than the state of the art designs, which are also aimed at spectral efficiency maximization. It was extremely fruitful research, which eventually won us the best student paper award at IEEE SPS conference, SPAWC 2018. However, it is not impossible that a fully digital transceiver design may become feasible in the coming years since we live at an age when the digital processor capabilities are increasing quickly. Still, HBF solutions may be the winner in the initial stages of next generation technologies while moving to mmWave deployments due to its energy efficiency. Moreover, in a full-duplex (FD) system, hybrid solutions may have much larger benefits since we can make spatial filtering before the signal hits the ADC. It can be useful, example, when we have strong interferers in spatial directions (self-interference), thereby reducing dynamic range problems. We also looked at the efficient hybrid or multi-stage BF design using the same optimization principle as discussed above for FD systems which make this work much more relevant.

It is to be noted that all the BF techniques discussed herein for HBF are applicable to fully digital system also, which is indeed a special case. Moreover, due to the reason described in the

paragraph, we further focus our attention on fully digital systems for our asymptotic analysis. We remark here that for the HBF design, our proposed techniques are applicable to any general channel models. In further related works, we start making certain assumptions on the wireless channel model to facilitate the asymptotic analysis. In MaMIMO/mmWave systems, the underlying user channel is sparse in angle, delay, and Doppler domains. For the pathwise MaMIMO channel model (which is a typical channel model where the propagation environment contains limited scattering), we first consider optimal BFs in the case of partial CSIT and noted that pathwise approach allows ZF of the interfering paths and that the ZF tasks get split between Transmitter and Receiver antennas. In the partial channel model, we assume that all the slow fading components such as AoA/AoD/delay/path attenuation remain constant over enough coherence time interval such that they can be estimated perfectly. Only the path phases (which are fast fading components) are assumed to change every channel use and hence are unknown. The estimation of the slow fading components can be done quite accurately using variational Bayesian (VB) inference techniques described later in this thesis.

Asymptotic analysis with HBF systems becomes a straightforward extension using the large system concepts described herein and it is left as future work. To start our asymptotic analysis, we look at a simplified BF scheme inspired by the extreme SNR region behavior of the MMSE BF. Moreover, it is to be mentioned that all the simplified results are for MaMISO case, even though our derived BF expressions do apply or easily extendable for the MaMIMO case also. For the MaMIMO (with multiple antenna users) case, we have some initial results under some simplifications on the MIMO channel model, but it is incomplete and requires further work. To reduce the complexity and simplify the beamforming design, we proposed a reduced order zero forcing (RO-ZF) BF solution, which has negligible performance loss compared to the optimal (linear) BF design. Using simple asymptotic analysis (requiring only law of large numbers), we evaluated the performance of RO-ZF, zero forcing (ZF), and optimal BFs for a realistic scenario of user channels with varying attenuation for varying levels of channel state information at the transmitter (CSIT). Further, using the techniques from random matrix theory, we evaluated the large system performance for various sub-optimal BF solutions in the case of partial CSIT. We formulated an upper bound of ergodic capacity, which is shown to be very tight in the MaMIMO regime and it is termed as Expected Signal and Interference Power based weighted sum rate (ESIP-WSR). We considered randomized user channel subspaces (stochastic geometry) to simplify the analysis and provided analytical insight into the problem compared to the state of the art solutions, which are cumbersome.

We obtained simplified sum rate expressions at low and high SNR regime, which gives intuitive insights into the spectral efficiency (SE) for a MaMIMO system. We analyzed both theoretically and numerically the system performance for various channel estimates (such as least squares, linear MMSE, and subspace projected) for different BF solutions. It is to be mentioned that linear MMSE or subspace projected channel estimates as expected give significant performance gains compared to just using least squares channel estimates. However, those schemes would require second order statistics of the user channels or channel subspace (and are low rank in MaMIMO or mmWave systems) estimation also. This channel subspace proves to be challenging due to a large number of antennas at the BS and hence require large overheads in the UL to obtain a reasonable estimation performance. However, if we take into account the specific structure of the pathwise channel model and exploit sparse signal processing techniques, this overhead for channel estimation can be reduced significantly. Here our further work on VB based sparse Bayesian learning algorithms described below becomes very useful.

Another interesting and relevant topic we started to work on is approximate Bayesian inference techniques for sparse signal processing. we focused on VB inference techniques whose un-

derlying concept is to find an approximate posterior which minimizes the Kullback Liebler (KL) divergence to the true posterior. Several existing techniques such as belief propagation (BP), mean field (MF) approximation, expectation propagation (EP), and approximate message passing (AMP) algorithms are different variants of the VB family of methods. In particular, we contributed to the theoretical performance analysis of these techniques using large system analysis derived from random matrix theory results and proposed extensions to sparse Bayesian learning using AMP. In the combined BP/MF/EP based approximate family of inference techniques, we conjectured that an optimal splitting of the unknown variables to estimated in the underlying factor graph considered can be formulated using the Fisher information matrix (FIM) based analysis. This indeed required us to consider mismatched Cramer Rao bound (CRB) based analysis, which deals with MSE bounds when the posterior information is not exact and note that the research of it is still in its infancy.

In a MaMIMO/mmWave system, low rank channel impulse response can be written as the Kronecker product of path components such as delay spread, Doppler, Transmitter (Tx) and Receiver (Rx) spatial multi-antenna dimensions. We proposed a Bayesian method based on Variational Bayesian inference called SAVE (space alternating Variational estimation), which has much lesser complexity and better convergence rate in terms of the mean square error than the existing state of the art low complexity solutions in sparse signal recovery. Moreover, our work is an instance of gridless compressive sensing, which also utilizes tensor decomposition methods such as canonical polyadic decomposition and Tucker decomposition. Hence this topic on approximate Bayesian inference has applications well outside the interest of communication society such as data science or many other signal processing applications including medical imaging or radar signal processing. To illustrate its significance, we refer to [3], where the authors talk about classification of hyperspectral images (HSI) using tensor decomposition techniques. HSI is a three way block of data, which is acquired when many two way images are taken over multiple continuous spectral bands. HSI is used in several applications including but not limited to astronomy, Earth and planetary observation, monitoring of natural resources, precision agriculture and biomedical imaging.

1.1 Motivation and State of the Art

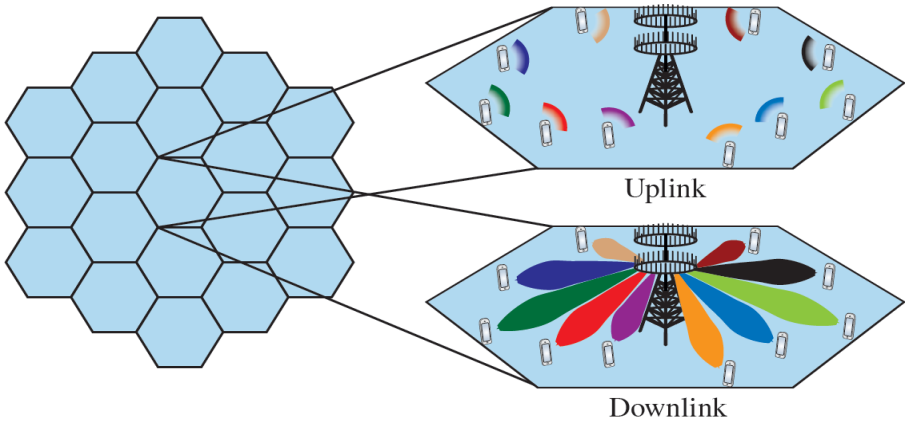


Figure 1.1: Illustration of uplink and downlink in Massive MIMO. Source of the figure [1]

MaMIMO [4] involves the use of a large number of antennas at the base station (BS) and access points to increase the system throughput for next generation technologies such as 5G and beyond. The main selling points of MaMIMO are increased spectral efficiency through highly directive nature of transmission and reduced power consumption. MaMIMO is also apt for communication in the mmWave frequency bands, which are characterized by different channel properties than that of the sub-6 GHz bands. Channel characteristics and hardware complexity differ from a conventional multi-antenna technology for a MaMIMO system, thus posing many challenges in the signal processing algorithm design. In this thesis, we try to address some of the challenges associated with the MaMIMO technologies and advance the state of the art a bit by proposing efficient solutions that are not only of interest to the academic community but possibly to the industry too.

Regardless of the numerous publications in the MaMIMO community, it has to be mentioned here that most of the existing works focus on suboptimal solutions for BF design or channel estimation schemes. Since the computational complexity usually scales with the system dimensions and it is not yet known in the literature how to find an optimal BF design for a multi-user MaMIMO system. Moreover, under practical scenarios, in the case of MaMIMO with a large number of antennas at the base station, it may not be feasible to have as many RF chains M as the number of antennas N_t. One promising solution is hybrid beamforming, which is a two-stage architecture where the beamformer (BF) is constructed by concatenation of a low-dimensional precoder (digital BF) and an analog BF, with the number of RF chains less than the number of antennas. This technique was first introduced in [5], with the analog precoder implemented using phase shifters.

An optimal BF design for a multi-user MIMO system under a fully digital precoding scheme itself is quite challenging. Several works have looked at this problem, for example, [6, 7]. Among them, in the pioneering work by Caire and Shamai [6], they show that an achievable rate region for MIMO broadcast channel (BC) can be obtained using Costa's Dirty Paper Coding (DPC) scheme [8] at the transmitter. Till now, this remains the best scheme in the literature. However, the complexity of the DPC scheme (which is nonlinear) is extremely high and not preferred in practice. Hence it is preferred to look at linear BF solutions, of which the optimal scheme which minimizes the MSE is the weighted sum MSE (WSMSE) based BF proposed by Christensen et. al. in [9]. The WSMSE BF is shown to converge to a local optimum of the weighted sum rate (WSR) maximization problem.

Moving to the hybrid BF case, the problem becomes highly challenging due to the nonconvex constraints (unit amplitude) on the analog BF coefficients. This indeed motivates most of the state of the art designs to use a suboptimal low complexity solution at the cost of degradation in performance compared to a fully digital multi-antenna system. However, we focus on the highly challenging weighted sum rate maximization problem itself. Nonconvex character of the resulting cost function implies that even if it is possible to show convergence to a local optimum [10], convergence to the global optimum cannot be guaranteed. To avoid the convergence to a local optimum, [11] proposed Deterministic Annealing (DA) for digital BF design in the MIMO interference channel. To avoid the issue of local optima plagued by alternating optimization of phasors, we propose to use deterministic annealing (for the first time in the literature) which has performance quite close to fully digital solution compared to the state of the art designs. Hence, deterministic annealing helps to track the globally optimal phasor design for a fixed set of digital BFs (of various users) and power allocation matrices. However, it remains to be mentioned that the overall WSR problem can only be shown to converge to a locally optimum solution. As shown in [11], deterministic annealing can be applied further to track the globally optimum solution for the actual WSR problem.

In massive multi-input multi-output (MISO) systems, the received signal and interference powers converge to their expected value due to the law of large numbers. Evaluation of the performance of the massive system usually involves extensive Monte-Carlo simulations to compute the spectral efficiency over large number of channel realizations. In a practical multi-user MaMIMO system, the expression for signal-to-interference-plus-noise-ratio (SINR) is quite cumbersome due to the presence of user channel covariance matrices present. Hence, the SINR expressions are not very intuitive and it is quite difficult to infer the system behavior from them. However, random matrix theory results [12, 13] helps to compute deterministic equivalents for the signal and interference power terms and helps to circumvent the need to do extensive Monte-Carlo simulations to evaluate the system performance. Major work on large system analysis for massive MISO (MaMISO) systems appears in [14]. The authors obtain deterministic (instead of channel realization dependent) expressions for various scalar quantities, facilitating the analysis and design of wireless systems. For example, it may allow us to evaluate beamforming performance without computing explicit beamformers. The analysis in [14] allowed for example, the determination of the optimal regularization factor in Regularized ZF (R-ZF) BF, both with perfect and partial CSIT. In [15], the authors investigate the deterministic limits for optimal beamformers, but only for the perfect CSIT MISO BC (broadcast channel) case. [16] proposes a large system analysis for optimized BF with partial CSIT as considered here. Furthermore, the channel, channel estimate, and channel error covariances can all be arbitrary and different for all users. However, the resulting deterministic analysis is quite cumbersome and does not allow much analytical insight. In this thesis, we introduce a stochastic geometry inspired randomization of the channel covariance eigenspaces, leading to much simpler analytical results, which depend only on some essential channel characteristics. We also introduce reduced complexity beamforming termed as reduced order ZF (RO-ZF) BF, which has negligible performance loss compared to the optimal BF and with reduced design complexity. This is motivated by the observation that optimal MMSE BF converges to the ZF BF [17] at high SNR and reduces to matched filtering at very low SNR. The form of the ZF BF considered herein is also motivated from the block diagonalization based precoding technique discussed in [18].

MaMIMO which is originally conceived for sub-6 GHz frequency band is ideal for mmWave frequency bands also, which is a potential target for future wireless technologies like 5G. The propagation characteristics at mmWave band is significantly different compared to the sub-6 GHz bands, with limited scattering clusters. An appropriate channel model, in that case, would be a pathwise MIMO channel model, characterized by few AoA/AoDs, path amplitudes, delay response, and Doppler shifts. Our work is aimed at providing efficient low complexity and high performance solutions for estimating the pathwise elements. We propose a Bayesian estimation approach, denoted as SAVE, which uses techniques from variational inference.

1.2 Organization of the thesis

In this section, we would like to provide a brief overview of the organization of the thesis. This thesis is divided into three parts justifying the title also. The Part II is about the research topic on beamforming techniques in MaMIMO. In Chapter 2, we start with an overview of the state of the art in HBF and motivates the various challenges in tackling HBF design at the BS side. We then describe in detail the different hybrid beamforming techniques proposed for half-duplex systems. We do consider a multi-cell multi-user MIMO system and the different BF techniques proposed are focussed on maximizing the downlink sum rate. In Chapter 3, we start looking at full-duplex systems, focus being on bidirectional backhaul link between two BSs. We state that

the existing very few hybrid beamforming designs in the literature avoid more practical hardware impairments at the RF side. We then derive multi-stage BF designs which can also tackle practical noise models in the RF chain. In both chapters, we do provide simulation results that depict the superiority of our BF designs compared to the state of the art.

In Chapter 4 of Part II, we look at robust BF designs for MaMIMO systems exploiting covariance CSIT. We start with a review of the BFs optimized using a weighted sum rate which is approximated using the difference of convex functions programming. Further, we look at the partial CSIT pathwise channel model, where only the path phases (fast fading components) are unknown. Further, using a MaMIMO limit of the expected weighted sum rate, we derive the transmit side and receive side BF expressions and finally illustrate the derived results with simulations. In Chapter 5, we look at a simple scheme to mitigate the pilot contamination effects in a single cell MaMIMO system. For this purpose, we first propose a rate splitting scheme (where the messages to UE is split into private and common message parts) and propose an efficient power allocation scheme by optimizing the sum rate using the difference of convex functions programming. We also derive exact expressions of SINR under maximum ratio precoding and the elaborate Monte-Carlo simulations for different scenarios to illustrate the effectiveness of rate splitting in mitigating pilot contamination (sum rate saturation at high SNR) to an extent.

In Part III, we move to the second aspect covered in this thesis, which is the asymptotic analysis of MaMIMO systems. In Chapter 6, we start looking at the deterministic equivalents of the SINR expressions for a simple channel model with multiple of identity covariance matrix, with the scalar factors (which represents the channel attenuation) are chosen as different for distinct users. An extreme case of the channel models, but one which allows to analyze the asymptotic limit expressions easily and provide very intuitive expressions. To analyze this scenario, we propose a simplified BF scheme called reduced order zero forcing (RO-ZF) and through simulations, we validate its sum rate performance which is very close to the MMSE based BF solutions. We derive asymptotic sum rate expressions of optimal BFs, RO-ZF, ZF and ZF-DPC under these simplified channel models. However, one remark here is that the assumed channel model with multiple of identity covariance matrices for all users is not practical. This particular channel modeling is more appropriate when the users are co-located. Moreover, the analysis (the resulting deterministic SINR expressions) becomes quite simple and provide intuitive rate expressions. We further extend the large system analysis to a more complex channel model with different covariance matrices for different users, which is more practical in a massive MIMO system.

In Chapter 7, we look at large system analysis based on stochastic geometry inspired randomization of the user covariance matrices. Firstly, we give a detailed description of the motivation behind our partial CSIT channel model and then review the BF expressions which are derived using an upper bound of the expected weighted sum rate. Further, we review and derive the necessary theory using random matrix theory results which will be used throughout the rest of the chapter for the large system analysis. Further, we derive deterministic equivalents of the signal and interference powers at various users side and then formulate the sum rate expressions. Finally, we provide very simplified sum rate expressions for certain special cases at high and low SNR and also for varying levels of CSIT. Monte-Carlo simulations are then provided which validate the accuracy of our large system expressions.

In the final Part IV, we look at approximate Bayesian inference techniques for sparse Bayesian learning (SBL). Our main motivation behind this section is to provide low complexity solutions for SBL and also advance the state of the art w.r.t the theoretical analysis in approximate Bayesian inference techniques which have wide applicability. In Chapter 8, we start with reviewing the original SBL algorithm and discuss its high complexity issues. Further, we derive the mean field variational Bayes based SBL algorithm called space alternating variational estimation (SAVE).

Through numerical simulations, we validate the superior signal recovery and fast convergence of SAVE compared to existing fast SBL algorithms. Further, we also extend the SAVE to a time varying sparse signal where the temporal correlation is modeled using a first order autoregressive process. In Chapter 9, we propose static and dynamic SBL algorithms using a combination of mean field, belief propagation, and expectaton propagation. Through, Fisher information matrix (FIM) analysis, we propose an optimal way to partition the variables in a factor graph such that optimal mean squared error (MSE) performance is obtained. Further, in Chapter 10, we go one step forward and look at the problem of joint dictionary learning and sparse state vector estimation. We consider that the dictionary can be structured (but non parameterized), with a Khatri-Rao or Kronecker factorization applied to the dictionary matrix. Further, combining tensor algebra and variational Bayesian inference, we propose novel estimation schemes for the considered problem. We also discuss identifiability issues under structured dictionary matrices.

1.3 Background Information

In this section, we would like to provide some background theory which will be useful also for anyone who is not an expert in the topics discussed here in, so that he can gain some prerequisites.

1.3.1 Background on Beamforming in MaMIMO

In this thesis, we use beamforming or precoding to represent the same concept. In short, they represent the usage of an antenna array to transmit one or more spatially directive signals. BF matrix is designed as a function of the estimated channel such that a directive signal (or a beam) is formed towards each user canceling the interference from other user's signals. First, we would like to mention that in this thesis we are specifically focused towards the weighted sum rate (WSR) maximization problem for the BF design. In fact, in the literature, we can find several optimization criteria such as WSR, SINR balancing (maximize the minimum SINR), weighted sum energy efficiency (energy efficiency of any user is defined as the ratio of rate and the power consumed per user), etc. In our case, we are specifically interested in maximizing the sum throughput across the entire network, with the weights chosen to assign certain priorities to users. Even though we are not interested in optimizing the weights in most of the work proposed here, it is considered to make it more general. By a broadcast channel (BC), we refer to a communication system where a single transmitter (BS) sends independent information through a shared medium to uncoordinated receivers (UEs). An interfering broadcast channel (IBC) refers to a multi-cell network where each BS serves multiple UEs in its network and the UEs in a particular cell are impacted by the inter-cell interference also. For a MaMISO IBC, the received signal at any UE k (assuming the channel between BS and UE k is represented as $\mathbf{h}_{k,b_k}$ and the BF for user k is $\mathbf{g}_k$, where b_k represents the BS to which user k is associated) can be written as

$$y_k = p_k \mathbf{h}_{k,b_k}^H \mathbf{g}_k x_k + \sum_{i \neq k} p_i \mathbf{h}_{k,b_i}^H \mathbf{g}_i x_i + v_k,\ v_k \sim \mathcal{N}(0, \sigma^2). \tag{1.1}$$

Further, the rate of any user k is defined as

$$\begin{aligned} R_k &= \ln(1+\gamma_k), \\ \text{where } \gamma_k &= \frac{p_k |\mathbf{h}_{k,b_k}^H \mathbf{g}_k|^2}{\sum_{i \neq k} p_i |\mathbf{h}_{k,b_i}^H \mathbf{g}_i|^2 + \sigma^2}, \end{aligned} \tag{1.2}$$

where γ_k denotes the instantaneous SINR (averaging only over the Gaussian Tx signal x_k and assuming that channel remains constant over the entire coherence interval) for user k. For a MaMIMO BC, the channel between user k and BS gets represented as $\mathbf{H}_k$ and the BF gets denoted as $\mathbf{G}_k$, assuming multiple streams can be decoded by the UE k. In this case, the capacity expression can be represented (assuming independent Gaussian signaling) as [19]

$$\begin{aligned} R_k &= \ln\det\left(\mathbf{R}_{\bar{k}}^{-1}\mathbf{R}_k\right) \\ &= \ln\det\left(\mathbf{I}+\mathbf{R}_{\bar{k}}^{-1}\mathbf{H}_{k,b_k}\mathbf{G}_k\mathbf{G}_k^H\mathbf{H}_{k,b_k}^H\right) \end{aligned} \tag{1.3}$$

where $\mathbf{R}_{\bar{k}}$ is the interference plus noise power and $\mathbf{R}_k$ is the total signal power received at user k

1.3.2 Alternating Minorization

First, we would like to mention the concept of alternating maximization. We define the vector $\mathbf{z}$ ($M-$length) to contain all the variables to be optimized by maximizing the scalar function $f(\mathbf{z})$. The generic iteration steps of an alternating maximization algorithm can be written as

$$\begin{aligned} &\text{Initialization: } \mathbf{z}^0 \text{ given} \\ &\qquad\qquad \text{For } t = 1,\ldots\text{till convergence} \\ &\qquad\qquad \text{For } i = 1,..,M \\ &\qquad\qquad z_i^t = \arg\max_{\mathbf{z}} f(\mathbf{z}_{i-}^t, z_i, \mathbf{z}_{i+}^{t-1}). \end{aligned} \tag{1.4}$$

The above algorithm iterates between the maximization of each of the variables z_i. If the objective function is concave w.r.t each of the variables (while fixing others), it can be easily verified that the alternating maximization algorithm leads to a local optimum solution, since

$$f(\mathbf{z}^t) \geq f\left(\mathbf{z}_{i-}^t, z_i^t, \mathbf{z}_{i+}^{t-1}\right) \geq f(\mathbf{z}^{t-1}) \tag{1.5}$$

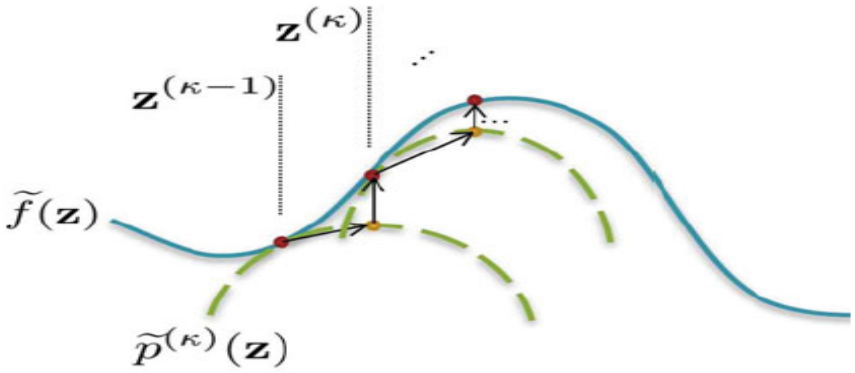


Figure 1.2: Concept of Alternating Minorization

The concept of minorization can be explained as follows. At every t^{th} iteration, let us say that we can find a function $g_t(\mathbf{z})$ which possess the following properties:

$$\begin{aligned} g_t(\mathbf{z}^t) &= f(\mathbf{z}^t), \\ g_t(\mathbf{z}) &\leq f(\mathbf{z}) \end{aligned} \tag{1.6}$$

The introduction of $g_t(\mathbf{z})$ is necessary since this surrogate function is much easier to optimize than the original function $f(\mathbf{z})$. Or in this thesis (for BF optimization), $f(\mathbf{z})$ is a non-concave function which cannot be optimized using standard convex optimization techniques. Hence, we formulate an approximate function $g_t(\mathbf{z})$ which is concave and hence can be optimized using conventional convex optimization methods. Moreover, optimization of this approximate problem leads to updates of $\mathbf{z}$ which monotonically increases the original objective function $f(\mathbf{z})$.

$$\begin{aligned} f(\mathbf{z}^t) = g_t(\mathbf{z}^t) &\leq g_t(\mathbf{z}^{t+1}) \\ &\leq f(\mathbf{z}^{t+1}). \end{aligned} \tag{1.7}$$

1.3.3 Background on Compressed Sensing

We start with a linear Gaussian model,

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{v}. \tag{1.8}$$

Given underdetermined $\mathbf{y}, \mathbf{A}$ ($\mathbf{A}$ is of dimension $M \times N$), the compressed sensing optimization problem can be written as

$$\min_{\mathbf{x}} \|\mathbf{x}\|_0 \text{ subject to } \mathbf{y} = \mathbf{A}\mathbf{x}.$$

We can recover $\mathbf{x}$ and its support for small $N - \|\mathbf{x}\|_0$ (small overdetermination if support were known). In the noisy case, it gets rewritten as

$$\min_{\mathbf{x}} \|\mathbf{x}\|_0 \text{ subject to } \left\|\mathbf{y} - \mathbf{A}\mathbf{x}\right\|_2 \leq \epsilon.$$

l_0 norm minimization is an NP-hard (non-deterministic polynomial-time hardness) problem. Hence, there exists other approximate solutions such as LASSO, basis pursuit which relax the l_0 to l_1 minimization and thus the problem becomes convex. This convex problems (though they may not have a closed form solution) can be solved using numerical methods or coordinate descent strategy.

We will now discuss briefly on the concepts of maximum a posteriori (MAP) or minimum mean squared error (MMSE) estimation here. MAP estimator is obtained by the global maximum of $p(\theta|\mathbf{y})$, where $p(\theta|\mathbf{y})$ represents the posterior distribution of θ.

$$\widehat{\theta}_{MAP} = \arg\max_{\theta} p(\theta|\mathbf{y}) \tag{1.9}$$

While an MMSE estimator is obtained as the θ which minimizes the mean squared error of the estimation, hence gets formulated as

$$\widehat{\theta}_{MMSE} = \arg\max_{\theta} \mathrm{E}_{\theta|\mathbf{y}} \left(\theta - \widehat{\theta}(\mathbf{y})\right)^2 \tag{1.10}$$

Further it can be derived as [20]

$$\widehat{\theta}_{MMSE} = \mathrm{E}(\theta|\mathbf{y}) \tag{1.11}$$

which is the mean of the posterior distribution.

Part II

Beamforming Techniques for Massive MIMO

Chapter 2

HYBRID BEAMFORMING

As Multi-Input Multi-Output (MIMO) systems allow spatial multiplexing, MaMIMO systems employ large numbers of antennas at the base station to increase the spectral efficiency of the system and possibly simplify beamforming techniques. With a large number of antennas though, it may not be feasible to have as many RF chains as the antennas due to the increased cost of the number of RF chains required (which includes Analog- to-Digital and Digital-to-Analog converters (ADCs/DACs), power amplifiers, and low noise amplifiers). So signal processing techniques called hybrid beamforming have been developed to take care of the case where the number of RF chains is less than the number of antennas.

In hybrid beamforming, the baseband precoder or the digital precoder is a low dimensional matrix which multiplexes the data streams to the number of RF chain which is much less than the number of antennas. The analog precoder further converts the output from the RF chains to the number of antennas. This technique was first introduced in [5]. In [5], a phase shifter constraint(unit modulus) is applied to the analog precoder elements. And the optimal analog precoder matrix is being derived for the case where there is only one data stream to be transmitted. In this case, the maximum performance is obtained only when there are at least 2 RF chains.

Most of the prior work on hybrid beamforming assume perfect channel knowledge (CSIT) at the transmitter. In practice, this is very difficult to obtain at fast fading rate, since for a MaMIMO system it increases the feedback in the uplink substantially reducing the spectral efficiency. In [21], a scheme called joint spatial division and multiplexing (which is a two-stage precoding scheme) is proposed such that a prebeamforming matrix is used which groups the users based on the spatial channel covariance. Users are separated in the spatial domain through a prebeamforming matrix. Users in the same group are further multiplexed through a linear MU-MIMO precoding which considers the effective channel as that including the prebeamforming matrix. Some of the prior works on designing the hybrid beamformers are described in [22–25]. In [22], some compressed sensing based schemes are proposed to estimate the channel in a hybrid structure utilizing the sparsity of the channel.

In [23], the phase shifter analog precoding matrix is seen as a two-step problem. In the first step, conditions for an unconstrained analog precoder matrix is derived assuming regularized ZF precoder for the baseband. Also, it is assumed that in the large system analysis limit, the SLNR for all users is equal. Also, the ZF precoding for digital beamforming makes it a suboptimal scheme. Once the unconstrained precoder is obtained, the phase shifter constrained analog precoder is obtained through an iterative algorithm by minimizing the Euclidean distance between unconstrained and the phase shifter constrained matrices. In [26], various architectures for the phase shifter matrix of analog precoder are given.

Herein, first we consider the design of analog precoder matrix where all the elements are phase elements by maximisation of the weighted sum rate. Each element of the phase shifter matrix (analog precoder) is obtained through an iterative process similar to [24]. But in [24], for the digital beamformers, the authors assume ZF precoders with effective channels including the analog precoder. With this approach, WSR problem simplifies to a water-filling algorithm for the power. For the analog precoder, taking the minimization of power as an objective, they optimize each element of the analog precoder matrix. Again, this is a suboptimal approach.

2.0.1 Summary of the Chapter

- In this chapter, firstly, for a multi-cell multi-user MIMO, we derive the hybrid digital and analog beamformers based on the maximization of weighted sum rate (under perfect CSIT) which is formulated as a weighted sum MSE (WSMSE) problem. The analog precoder is assumed to have all elements with unit modulus (or only phase elements) and each of the elements in this matrix are optimized iteratively in an alternating optimization fashion, as also the digital beamformers and auxiliary quantities (receivers and weights).

- Simulations are performed by alternating between joint updating of analog and digital beamformers using perfect CSIT. Further updating the digital beamformers on uncorrelated channel realizations (but with the same covariance) while the analog beamformers are frozen. Results show that even with the analog precoder based on outdated CSIT, close to optimal performance (WSR) is attained. This implies that CSIT feedback needs to be made much less for a large antenna array system. Practically, this is even more of an interest for OFDM system, where we cannot afford to loose throughput due to large feedback for CSIT. The results suggest indeed that we could afford updating the analog beamformers only once every so many channel uses across time or frequency in an OFDM system.

- Secondly, we propose a hybrid beamforming design based on the WSR criterion which is simplified using the minorization approach [27]. The advantage compared to the WSMSE solution [10] is that the iterative algorithm converges faster (no ping-pong between Tx and Rx optimization, and direct power optimization).

- We derive conditions under which the HBF can attain the fully digital performance with sufficient number of RF chains.

- To overcome the issue of local optima, we propose a deterministic annealing approach for the design of the analog phasors. Simulation results suggest that the proposed alternating optimization based WSR maximizing algorithm performs better than the state of the art solutions. Moreover, it is interesting to observe that the proposed DA based HBF design allows to narrow the gap to fully digital solutions [9].

- Finally, we also extend our HBF designs to more practical amplitude constraints such as per-RF or per-antenna power constraints (first time in the literature to our best knowledge) and show that our solutions have better scalability w.r.t the complexity compared to the existing few state of the art solutions (which are fully digital also).

2.0.2 Phase Shifter Architecture

Before going into the mathematical details of the HBF design, we would like to discuss here about the two different phase shifting architecture. In the fully connected phase shifting network, see

Figure 2.1, the feeding signal to each antenna is a weighted (by phasors) combination of the all the RF chain outputs. It requires $N_t M$ phase shifters. While for a partially connected phase shifting network as in Figure 2.2, it requires just N_t phase shifters. In a partially connected case, each RF chain is connected to only a subset of antennas ($L_t = \frac{N_t}{M}$ of them). However, the spectral efficiency of the partially connected HBF will be degraded compared to a fully connected case but a lower complexity and we validate this also in the coming simulation results section. It is also worth mentioning that there exist several other phase shifting architectures in the literature such as the one using switches instead of phase shifters [26]. However, our focus in this chapter is to provide a benchmark on the performance of the HBF systems and not look at reducing the complexity of such systems.

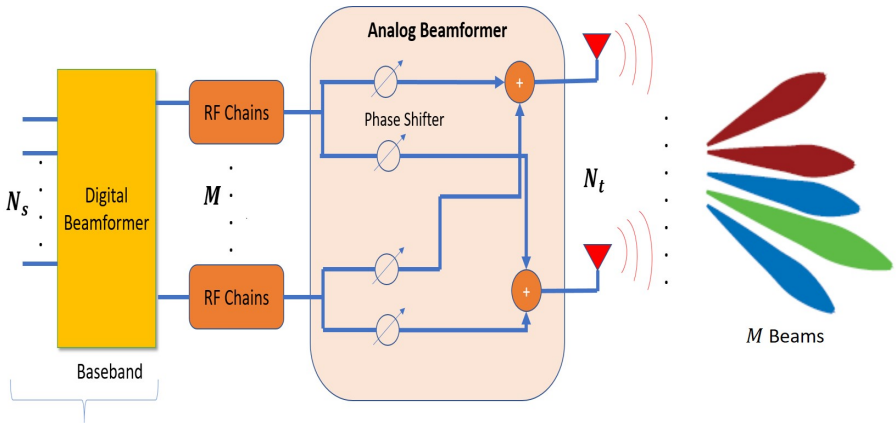


Figure 2.1: Fully Connected HBF

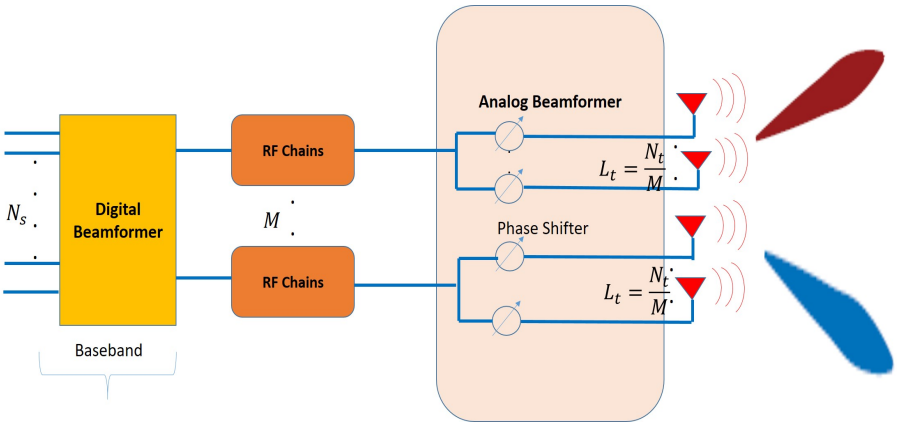


Figure 2.2: Partially Connected HBF

2.1 HBF Design using WSMSE for Multi-User MIMO

Consider a Multi-User MIMO system with N_t^c transmit antennas in cell c and K multi-antenna users. In this section, we shall consider a per stream approach (which in the perfect CSI case would be equivalent to per user). In an IBC formulation, one stream per user can be expected to be the usual scenario. In the development below, in the case of more than one stream per user, we shall treat each stream as an individual user. So, consider an IBC with C cells with a total of K

users. We shall consider a system-wide numbering of the users. User k is served by BS b_k. User k is equipped with N_k antennas. The $N_k \times 1$ received signal at user k in cell b_k can be written as

$$\mathbf{y}_k = \underbrace{\mathbf{H}_{k,b_k}\mathbf{V}^{b_k}\mathbf{g}_k\, s_k}_{\text{signal}} + \mathbf{H}_{k,b_k}\mathbf{V}^{b_k} \underbrace{\sum_{\substack{i\neq k \\ b_i=b_k}} \mathbf{g}_i\, s_i}_{\text{intracell interf.}} + \mathbf{H}_{k,c}\mathbf{V}^{c} \underbrace{\sum_{c\neq b_k}\sum_{i:b_i=c} \mathbf{g}_i\, s_i}_{\text{intercell interf.}} + \mathbf{v}_k \tag{2.1}$$

where s_k is the intended (white, unit variance) scalar signal stream, $\mathbf{H}_{k,b_k}$ is the $N_k \times N_t^{b_k}$ channel from BS b_k to user k. $\mathbf{H}_{k,b_i}$ represents the $N_k \times N_t^{b_i}$ channel from BS b_i to user k.

BS c serves $U_c = \sum_{i:b_i=c} 1$ users. We considered a noise whitened signal representation so that we get for the noise $\mathbf{v}_k \sim \mathscr{CN}(0, I_{N_k})$. The $N_t^{b_k} \times 1$ spatial Tx filter or beamformer (BF) is $\mathbf{g}_k$. The analog beamformer for base station c, V^c is of dimension $N_t^c \times M^c$. M^c is the number of RF chains at BS c. Treating interference as noise, user k will apply a linear Rx filter $\mathbf{f}_k$ (of dimension $N_k \times 1$) to maximize the signal power (diversity) while reducing any residual interference that would not have been (sufficiently) suppressed by the BS Tx. The Rx filter output is $\widehat{s}_k = \mathbf{f}_k^H \mathbf{y}_k$, hence

$$\widehat{s}_k = \mathbf{f}_k^H \mathbf{H}_{k,b_k}\mathbf{V}^{b_k}\mathbf{g}_k\, s_k + \sum_{i=1,\neq k}^{K} \mathbf{f}_k^H \mathbf{H}_{k,b_i}\mathbf{V}^{b_i}\mathbf{g}_i\, s_i + \mathbf{f}_k^H \mathbf{v}_k. \tag{2.2}$$

The transmit power constraints could be written as

$$\operatorname{tr}\left(\mathbf{V}^c \left(\sum_{i:b_i=c}^{K} \mathbf{g}_i\, \mathbf{g}_i^H\right) \mathbf{V}^{b_k\, H}\right) \leq P_c \tag{2.3}$$

where P_c is the transmit power constraint at BS c.

2.1.1 WSR Optimization in terms of WSMSE

We consider the problem of maximizing the weighted sum rate of the MIMO IBC system. This could be written as

$$\left[\mathbf{g}_1^{WSR}, ..., \mathbf{g}_K^{WSR}, \mathbf{V}^{1,WSR}, ..., \mathbf{V}^{C,WSR}\right] = \arg\max_{\mathbf{g},\mathbf{V}} \sum_{k=1}^{K} u_k R_k. \tag{2.4}$$

where the u_k are rate weights. In this thesis, we do not consider the optimization of the weights u_k and hence are known. These weights can be used to represent the priority assigned to certain users. For example, $u_k = 0$ for any user k means k^{th} user's rate is excluded from the objective function and $k = 1, \forall k$ means (2.4) reduces to sum rate maximization problem. In short, we have the following constraints on u_k.

$$0 \leq u_k \leq 1. \tag{2.5}$$

Here the optimization is over analog beamformers for C cells and digital beamformers for the K users. Under Gaussian signaling and optimal single user decoding, rate R_k for user k is defined as

$$R_k = \max_{\mathbf{f}_k} \ln\left(1 + \gamma_k\right), \tag{2.6}$$

where γ_k is the SINR (Signal to Interference plus Noise Ratio) for user k. It can be written as

$$\gamma_k = \frac{\left|\mathbf{f}_k \mathbf{H}_{k,b_k}\mathbf{V}^{b_k}\mathbf{g}_k\right|^2}{\sum_{i=1,i\neq k}^{K} \left|\mathbf{f}_k \mathbf{H}_{k,b_i}\mathbf{V}^{b_i}\mathbf{g}_i\right|^2 + ||\mathbf{f}_k||^2}. \tag{2.7}$$

Solving the weighted sum rate is equivalent to solving the weighted sum MSE problem from [9, 28]. Let us define w_k (≥ 0) as the weights associated with the MSE of user k. The augmented cost function for the WSMSE could be written as

$$WSMSE(\mathbf{V},\mathbf{g},\mathbf{f},\mathbf{w}) = \sum_{k=1}^{K} u_k \left(w_k e_k\left(\mathbf{f}_k,\mathbf{V},\mathbf{g}\right) - \ln\left(w_k\right)\right) \tag{2.8}$$

$$+ \sum_{c=1}^{C} \lambda_c \left[\mathrm{tr}\left(\mathbf{V}^{c\,H}\mathbf{V}^{c} \sum_{i:b_i=c} \mathbf{g}_i \mathbf{g}_i^H\right) - P_c\right], \tag{2.9}$$

where e_k is the MSE, w_k is the MSE weight and λ_c is the Lagrange multiplier associated with the power constraint at BS c. Let us define the transmit SNR as $\rho_c = P_c$. The MSE is

$$e_k\left(\mathbf{f}_k,\mathbf{V},\mathbf{g}\right) = E\left[\left(s_k - \mathbf{f}_k^H\mathbf{y}_k\right)\left(s_k - \mathbf{f}_k^H\mathbf{y}_k\right)^H\right] \tag{2.10}$$

$$= 1 - \mathbf{f}_k^H \mathbf{H}_{k,b_k}\mathbf{V}^{b_k}\mathbf{g}_k - \mathbf{g}_k^H\mathbf{V}^{b_k\,H}\mathbf{H}_{k,b_k}^H\mathbf{f}_k + \sum_{i=1}^{K}\mathbf{f}_k^H\mathbf{H}_{k,b_i}\mathbf{V}^{b_i}\mathbf{g}_i\mathbf{g}_i^H\mathbf{V}^{b_i\,H}\mathbf{H}_{k,b_k}^H\mathbf{f}_k + \|\mathbf{f}_k\|^2 \tag{2.11}$$

assuming $\mathrm{E}\left[|s_k|^2\right] = 1$. As in [9, 29], performing alternating optimization leads to solving simple quadratic or convex functions:

$$\min_{w_k} WSMSE \implies w_k = e_k^{-1}(\mathbf{f}_k,\mathbf{V},\mathbf{g}) = 1 + \gamma_k \tag{2.12}$$

$$\min_{\mathbf{f}_k} WSMSE \implies \mathbf{f}_k = \left(\sum_{i=1}^{K}\mathbf{H}_{k,b_i}\mathbf{V}^{b_i}\mathbf{g}_i\mathbf{g}_i^H\mathbf{V}^{b_i\,H}\mathbf{H}_{k,b_i}^H + \mathbf{I}\right)^{-1}\mathbf{H}_{k,b_k}\mathbf{V}^{b_k}\mathbf{g}_k \tag{2.13}$$

$$\min_{\mathbf{g}_k} WSMSE \implies \mathbf{g}_k = \left(\sum_{i=1}^{K}\frac{u_i}{e_i(\mathbf{f}_i,\mathbf{V},\mathbf{g})}\mathbf{V}^{b_k\,H}\mathbf{H}_{i,b_k}^H\mathbf{f}_i\mathbf{f}_i^H\mathbf{H}_{i,b_k}\mathbf{V}^{b_k} + \lambda_{b_k}\mathbf{V}^{b_k\,H}\mathbf{V}^{b_k}\right)^{-1}\mathbf{V}^{b_k\,H}\mathbf{H}_{k,b_k}^H\mathbf{f}_k\frac{u_k}{e_k(\mathbf{f}_k,\mathbf{V},\mathbf{g})} \tag{2.14}$$

where an analytical expression can be obtained for the Lagrange multipliers, viz.

$$\lambda_c = \frac{1}{P_c}\sum_{i:b_i=c}\frac{u_i}{e_i(\mathbf{f}_i,\mathbf{V},\mathbf{g})}||\mathbf{f}_i||^2, \tag{2.15}$$

and the beamformers are rescaled to make sure that the transmit power constraints are satisfied:

$$\begin{aligned}\mathbf{g}_k &\leftarrow \xi_{b_k}\mathbf{g}_k,\\ \xi_c &= \sqrt{P_c / \sum_{i:b_i=c}||\mathbf{V}^c\mathbf{g}_i||^2}.\end{aligned} \tag{2.16}$$

So the algorithm performs alternating optimization between the MSE weights, the Rx MMSE filters, and the digital and possibly the analog beamformers for which we shall discuss the optimization now.

2.1.2 Design of the Analog Beamformer with Perfect CSIT

Given g, f, and w, the analog beamformer V^c can be found by performing alternating optimization elementwise. Accounting of the unit modulus constraints of the entries of V^c can be done by parameterizing as

$$\left|\mathbf{V}^c_{m,n}\right| = 1 \Rightarrow \mathbf{V}^c_{m,n} = e^{j\theta^c_{m,n}}. \tag{2.17}$$

Now, as shown in the Appendix B, the WSMSE can be written as

$$WSMSE = 2\Re\{e^{j\theta^c_{m,n}} a^c_{m,n}\} + \text{"terms not containing } \theta^c_{m,n}\text{"} \tag{2.18}$$

where the scalar $a^c_{m,n}$ is defined in (24) of Appendix B. Then the minimization of the WSMSE w.r.t. $\theta^c_{m,n}$ yields

$$\theta^c_{m,n} = \pi - \angle a^c_{m,n}. \tag{2.19}$$

Note that one phase factor in V^c is undetermined, hence example, $\theta^c_{1,1} = 0$. Then (2.19) can be iterated for the other $m = 1,...,N^c_t$, $n = 1,...,M^c$. The complete derivation is given in the Appendix B. The steps of the complete iterative algorithm are given in table Algorithm 1. There the $\mathbf{V}^c$ are initialized from the M^c dominating generalized eigenvectors of the "signal" channel covariances $\sum_{k:b_k=c} \Theta^c_k$ and the "interference" channel covariances $\sum_{i:b_i\neq c} \Theta^c_i$ where $\Theta^c_k = \mathrm{E}\left[\mathbf{H}^H_{k,c}\mathbf{H}_{k,c}\right]$. Also, the operation $e^{j\angle\mathbf{A}}$ for a matrix $\mathbf{A}$ takes the elementwise phasors. Various variations on the alternating optimization updating schedules are possible. For instance, the elements $\theta^c_{m,n}$ could be updated only once in every sweep of updates of all quantities (as suggested in the table), or these elements could be iterated separately until convergence before updating again the other quantities.

2.1.3 Mixed Time Scale Adaptation

In this section, we consider two variants of the WSMSE iterative algorithm shown. In the first variant, Fast Time Scale Adaptation, the WSR is maximized straightforwardly using Algorithm 1, using (perfect) instantaneous CSIT for the computation of both analog and digital beamformers $\mathbf{V}$ and $\mathbf{g}$. This adaptation is repeated whenever the instantaneous CSIT (the channels $\mathbf{H}_{k,c}$) changes.

In the second variant, Mixed Time Scale Adaptation, the overall Fast Time Scale Adaptation just mentioned gets executed only from time to time, whenever the slow CSIT, here captured by the channel covariance matrices Θ^c_k, changes. In between those slow CSIT updates, the digital beamformers $\mathbf{g}$ and all auxiliary quantities, but not the analog beamformers, get updated whenever the fast CSIT changes. This can be done using Algorithm 1, in which step 5), the update of the analog beamformers, gets skipped. Hence, the values of the analog beamformers are frozen over the slow fading coherence time, whereas only the digital parts get updated at the fast fading rate. Whenever the slow fading CSIT is considered to have changed, all quantities are updated using the instantaneous CSIT available at such time instant. No dynamics of the fast or slow fading processes get exploited. When an update gets performed, all quantities to be updated get recomputed from scratch. The information in the previous updates gets ignored, except for providing the initialization values. The initialization mentioned in Algorithm 1 gets performed only once, at the very first initialization of the whole process.

Algorithm 1 WSMSE Iterative algorithm

Given: $P_c, H_{k,c}, u_k \forall k, c$.
Initialization: $\mathbf{V}^c = e^{j\angle \mathbf{V}_{1:M^c}(\sum_{k:b_k=c}\Theta_k^c, \sum_{i:b_i\neq c}\Theta_i^c)}$
The $\mathbf{f}_k$ are taken as the dominant left singular vector of $\mathbf{H}_{k,b_k}$.
The $\mathbf{g}_k$ are taken as the MMSEZF precoders for the effective channels $\mathbf{f}_k\mathbf{H}_{k,c}\mathbf{V}^c$.
Initialize SINR $\gamma_k^{(0)}$ from (2.7).
Iteration (j)

1. Update $\forall k$, $e_k^{(j)}$, $w_k^{(j)}$ from (2.12)
2. Update $\forall k$, $\mathbf{f}_k^{(j)}$ from (2.13)
3. Update $\forall c$, $\lambda_c^{(j)}$ from (2.15)
4. Update $\forall k$, $\mathbf{g}_k^{(j)}$ from (2.14),(2.16)
5. Update $\forall c$, $\forall (m,n)$, $\mathbf{V}_{m,n}^{c(j)}$ from (2.19)
6. Compute $\forall k$, $\gamma_k^{(j)}$, from (2.7)
7. Check for convergence of the WSR, if not go to step 1.

2.2 Hybrid Beamforming for Globally Converging Phasor Design

As we saw in Section 2.1, the main issue with WSR/WSMSE optimization for an HBF hybrid design is the high non-convexity of the cost function. This implies that even if it is possible to show convergence to a local optimum [10], convergence to the global optimum cannot be guaranteed. To avoid the convergence to a local optimum, [11] proposed Deterministic Annealing (DA) for digital BF design in the MIMO interference channel.

In this section, we go one step further and consider a multi-stream approach with d_k streams for user k. So, consider an Interfering BroadCast (IBC) (i.e. multi-cell MU downlink) system of C cells with a total of K users and N_t^c transmit antennas in cell c. User k is equipped with N_k antennas. $\mathbf{H}_{k,c}$ represents the $N_k \times N_t^c$ MIMO channel between user k and BS c and we define $\mathrm{E}\left[\mathbf{H}_{k,c}^H\mathbf{H}_{k,c}\right] = \Theta_k^c$. User k receives

$$\mathbf{y}_k = \mathbf{H}_{k,b_k}\mathbf{V}^{b_k}\mathbf{G}_k\mathbf{s}_k + \sum_{i\neq k}\mathbf{H}_{k,b_i}\mathbf{V}^{b_i}\mathbf{G}_i\mathbf{s}_i + \mathbf{v}_k, \tag{2.20}$$

where $\mathbf{s}_k$, of size $d_k \times 1$, is the intended signal stream vector (all entries are white, unit variance). BS c serves $U_c = \sum_{i:b_i=c} 1$ users. We are considering a noise whitened signal representation so that we get for the noise $\mathbf{v}_k \sim \mathcal{CN}(0, \mathbf{I}_{N_k})$. The analog beamformer $\mathbf{V}^c$ for base station c is of dimension $N_t^c \times M^c$ where M^c is the number of RF chains at BS c. The $M^c \times d_k$ digital beamformer is $\mathbf{G}_k$, where $\mathbf{G}_k = \left[\mathbf{g}_k^{(1)} \ldots \mathbf{g}_k^{(d_k)}\right]$ and $\mathbf{g}_k^{(s)}$ represents the beamformer for stream s of user k. The transmit power constraint at base station c can be written as $\mathrm{tr}\{\mathbf{V}^{c\,H}\mathbf{V}^c\sum_{i:b_i=c}^K \mathbf{G}_i\mathbf{G}_i^H\} \leq P_c$.

2.2.1 Alternating Minorization Approach

Consider the optimization of the hybrid beamforming design using WSR maximization of the Multi-cell MU-MIMO system:

$$\begin{aligned}[\mathbf{V}\ \mathbf{G}] &= \arg\max_{\mathbf{V},\mathbf{G}} WSR(\mathbf{G},\mathbf{V}) \\ &= \arg\max_{\mathbf{V},\mathbf{G}} \sum_{k=1}^{K} u_k \ln\det\left(\mathbf{R}_{\bar{k}}^{-1}\mathbf{R}_k\right),\end{aligned} \tag{2.21}$$

where the u_k are the rate weights, $\mathbf{G}$ represents the collection of digital BFs $\mathbf{G}_k$, $\mathbf{V}$ the collection of analog BFs $\mathbf{V}^{b_k}$. From [9], we can write,

$$\begin{aligned}\mathbf{R}_{\bar{k}} &= \sum_{i=1,i\neq k}^{K} \mathbf{H}_{k,b_i}\mathbf{Q}_i\mathbf{H}_{k,b_i}^H + \mathbf{I}_{N_k}, \\ \mathbf{R}_k &= \sum_{i=1}^{K} \mathbf{H}_{k,b_i}\mathbf{Q}_i\mathbf{H}_{k,b_i}^H + \mathbf{I}_{N_k}, \\ \mathbf{Q}_i &= \mathbf{V}^{b_i}\mathbf{G}_i\mathbf{G}_i^H\mathbf{V}^{b_i\,H}\end{aligned} \tag{2.22}$$

where $\mathbf{R}_{\bar{k}}$ is the interference plus noise covariance matrix.With the definition of the Tx covariance matrices $\mathbf{Q}_i$, the power constraints can be written as,

$$\textstyle\sum_{k:b_k=c} \mathrm{tr}\{\mathbf{Q}_k\} \leq P_c\,. \tag{2.23}$$

The WSR problem is non-concave in the $\mathbf{Q}_k$ due to the interference terms. Therefore finding the global optimum is challenging. In order to render a feasible solution, we consider the difference of convex functions (DC programming) approach as in [30] in which the WSR is written as the summation of a convex and a concave term. Consider the dependence of the WSR on $\mathbf{Q}_k$ alone:

$$\begin{aligned}WSR(\mathbf{G},\mathbf{V}) &= u_k \ln\det\left(\mathbf{R}_{\bar{k}}^{-1}\mathbf{R}_k\right) + WSR_{\bar{k}}, \\ WSR_{\bar{k}} &= \sum_{i=1,\neq k}^{K} u_i \ln\det\left(\mathbf{R}_{\bar{i}}^{-1}\mathbf{R}_i\right),\end{aligned} \tag{2.24}$$

where $\ln\det(\mathbf{R}_{\bar{k}}^{-1}\mathbf{R}_k)$ is concave in $\mathbf{Q}_k$ and $WSR_{\bar{k}}$ is convex in $\mathbf{Q}_k$. Since a linear function is simultaneously convex and concave, consider the first order Taylor series expansion of $WSR_{\bar{k}}$ in $\mathbf{Q}_k$ around $\widehat{\mathbf{Q}}$ (i.e. all $\widehat{\mathbf{Q}}_i$).

$$\begin{aligned}WSR_{\bar{k}}\left(\mathbf{Q}_k,\widehat{\mathbf{Q}}\right) &\approx WSR_{\bar{k}}\left(\widehat{\mathbf{Q}}_k,\widehat{\mathbf{Q}}\right) - \mathrm{tr}\left\{\left(\mathbf{Q}_k-\widehat{\mathbf{Q}}_k\right)\widehat{\mathbf{A}}_k\right\}, \\ \widehat{\mathbf{A}}_k &= -\left.\frac{\partial WSR_{\bar{k}}\left(\mathbf{Q}_k,\widehat{\mathbf{Q}}\right)}{\partial\mathbf{Q}_k}\right|_{\widehat{\mathbf{Q}}_k,\widehat{\mathbf{Q}}} \\ &= \sum_{i=1,\neq k}^{K} u_i\,\mathbf{H}_{i,b_k}^H\left(\widehat{\mathbf{R}}_{\bar{i}}^{-1} - \widehat{\mathbf{R}}_i^{-1}\right)\mathbf{H}_{i,b_k}\,.\end{aligned} \tag{2.25}$$

Note that the linearized tangent expression for $WSR_{\bar{k}}$ constitutes a lower bound for it and hence the DC approach (in $\mathbf{Q}$) is also a minorization approach (in $\mathbf{Q}$ or $\mathbf{G}$). Now, dropping constant

terms, reparameterizing the $\mathbf{Q}_k = \mathbf{G}_k\mathbf{G}_k^H$, performing this linearization for all users and augmenting the WSR cost function with the Tx power constraints, we get the Lagrangian,

$$\begin{aligned} WSR(\mathbf{G},\mathbf{V},\lambda) = &\sum_{k=1}^{K} u_k \ln\det\left(\mathbf{I} + \mathbf{G}_k^H \mathbf{V}^{b_k H}\widehat{\mathbf{B}}_k\mathbf{V}^{b_k}\mathbf{G}_k\right) \\ &- \mathrm{tr}\left\{\mathbf{G}_k^H\mathbf{V}^{b_k H}\left(\widehat{\mathbf{A}}_k + \lambda_{b_k}\mathbf{I}\right)\mathbf{V}^{b_k}\mathbf{G}_k\right\} + \sum_{j=1}^{C}\lambda_j P_j, \end{aligned} \tag{2.26}$$

where $\widehat{\mathbf{B}}_k = \mathbf{H}_{k,b_k}^H \widehat{\mathbf{R}}_{\bar{k}}^{-1}\mathbf{H}_{k,b_k}$. In what follows, we shall optimize the WSR with perfect CSIT by alternating optimization between digital and analog beamformers.

2.2.2 Digital BF Design

The gradient w.r.t. $\mathbf{G}_k$ of (2.26) (which is still the same as that of (2.4)) leads to the solution as d_k dominant generalized eigenvectors

$$\mathbf{G}_k^{'} = \mathbf{V}_{1:d_k}\left(\mathbf{V}^{b_k H}\widehat{\mathbf{B}}_k\mathbf{V}^{b_k}, \mathbf{V}^{b_k H}\left(\widehat{\mathbf{A}}_k + \lambda_{b_k}\mathbf{I}\right)\mathbf{V}^{b_k}\right), \tag{2.27}$$

with associated generalized eigenvalues $\Sigma_k = \Sigma_{1:d_k}\left(\mathbf{V}^{b_k H}\widehat{\mathbf{B}}_k\mathbf{V}^{b_k}, \mathbf{V}^{b_k H}\left(\widehat{\mathbf{A}}_k + \lambda_{b_k}\mathbf{I}\right)\mathbf{V}^{b_k}\right)$. Let $\mathbf{\Sigma}_k^{(1)} = \mathbf{G}_k^{'H}\mathbf{V}^{b_k H}\widehat{\mathbf{B}}_k\mathbf{V}^{b_k}\mathbf{G}_k^{'}$ and $\mathbf{\Sigma}_k^{(2)} = \mathbf{G}_k^{'H}\mathbf{V}^{b_k H}\widehat{\mathbf{A}}_k\mathbf{V}^{b_k}\mathbf{G}_k^{'}$. The advantage of formulation (2.26) is that it allows straightforward power adaptation: introducing stream powers in the diagonal matrices $\mathbf{P}_k \geq 0$ and substituting $\mathbf{G}_k = \mathbf{G}_k^{'}\mathbf{P}_k^{\frac{1}{2}}$ in (2.26) yields

$$WSR(\mathbf{P},\lambda) = \sum_{j}^{C}\lambda_j P_j + \sum_{k=1}^{K}\left[u_k \ln\det\left(\mathbf{I} + \mathbf{P}_k\mathbf{\Sigma}_k^{(1)}\right) - \mathrm{tr}\{\mathbf{P}_k\left(\mathbf{\Sigma}_k^{(2)} + \lambda_{b_k}\mathbf{V}^{b_k H}\mathbf{V}^{b_k}\right)\}\right], \tag{2.28}$$

the optimization of which leads to the following interference leakage aware water filling (WF) (jointly for the $\mathbf{P}_k$ and λ_c)

$$\mathbf{P}_k = \left(u_k\left(\mathbf{\Sigma}_k^{(2)} + \lambda_{b_k}\mathbf{V}^{b_k H}\mathbf{V}^{b_k}\right)^{-1} - \mathbf{\Sigma}_k^{-(1)}\right)^{+}, \tag{2.29}$$

where $(x)^+ = \max(0,x)$ is applied to all diagonal elements, and the Lagrange multipliers are adjusted to satisfy the power constraints. This can be done by bisection and gets executed per BS. Given the digital BFs, we update the analog beamformers $\mathbf{V}^c$. First, we consider the case in which the analog beamformer is unconstrained.

2.2.3 Design of Unconstrained Analog BF

To optimize $\mathbf{V}^c$, we set the gradient of (2.26) w.r.t. $\mathbf{V}^c$ equal to zero. Using the results $\nabla \ln\det\mathbf{X} = \mathrm{tr}\left(\mathbf{X}^{-1}\nabla\mathbf{X}\right)$ and $\det(\mathbf{I}_M + \mathbf{XY}) = \det(\mathbf{I}_N + \mathbf{YX})$ from [31], we get

$$\sum_{k:b_k=c}\widehat{\mathbf{B}}_k\mathbf{V}^c\mathbf{G}_k\mathbf{G}_k^H\mathbf{W}_k - \sum_{k:b_k=c}\left(\widehat{\mathbf{A}}_k + \lambda_c\mathbf{I}\right)\mathbf{V}^c\mathbf{G}_k\mathbf{G}_k^H = 0,$$

$$\text{where } \mathbf{W}_k = u_k\left(\mathbf{I} + \mathbf{G}_k\mathbf{G}_k^H\mathbf{V}^{b_k H}\widehat{\mathbf{B}}_k\mathbf{V}^{b_k}\right)^{-1}.$$

Now using $vec\,(\mathbf{AXB}) = \left(\mathbf{B}^T \otimes \mathbf{A}\right) vec\,(\mathbf{X})$ from [31], we get

$$\begin{aligned}
&\mathbf{V}^c \overset{(a)}{=} unvec\,(\mathbf{V}_{max}\,(\mathbf{B}_c, \mathbf{A}_c)) \text{ with} \\
&\mathbf{B}_c = \sum_{k:b_k=c} \left(\left(\mathbf{G}_k \mathbf{G}_k^H \mathbf{W}_k\right)^T \otimes \widehat{\mathbf{B}}_k\right), \\
&\mathbf{A}_c = \sum_{k:b_k=c} \left(\left(\mathbf{G}_k \mathbf{G}_k^H\right)^T \otimes \left(\widehat{\mathbf{A}}_k + \lambda_c \mathbf{I}\right)\right).
\end{aligned} \tag{2.30}$$

In (a) above in (2.70), $\mathbf{V}_{max}\,(\mathbf{B}_c, \mathbf{A}_c)$ represents the dominant generalized eigenvector of $\mathbf{B}_c, \mathbf{A}_c$. $vec(\mathbf{X})$ represents the vectorization of any matrix $\mathbf{X}$ (by stacking the columns on top of each other) and $unvec(\mathbf{x})$ represents the reverse operation which converts the vector $\mathbf{x}$ to the matrix $\mathbf{X}$. The unconstrained BF derived here is used in Section 2.2.4.3 to design the deterministic annealing based analog phasors.

2.2.4 Design of Phase Shifter Constrained Analog Beamformer

Given the digital BFs, the phase shifter analog beamformer $\mathbf{V}^c$ for BS c can be found by performing alternating optimization elementwise. Accounting for the unit modulus constraints of the entries of $\mathbf{V}^c$ can be done by parameterizing as

$$\left|\mathbf{V}^c_{p,q}\right| = 1 \implies \mathbf{V}^c_{p,q} = e^{j\theta^c_{p,q}}. \tag{2.31}$$

Since the analog BF is common to all users in a cell c, from (2.26) we can write the WSR as a function of $\theta^c_{p,q}$ as

$$\begin{aligned}
f\left(\theta^c_{p,q}\right) = \sum_{k:b_k=c} \Big[u_k \ln \det \Big(&\mathbf{I} + \mathbf{C}^k_{p,q}\, e^{j\theta^c_{p,q}} + \mathbf{D}^k_{p,q}\, e^{-j\theta^c_{p,q}} \\
&+ \mathbf{T}^{k,1}_{\overline{p},\overline{q}} \Big) - \mathrm{tr}\left(\mathbf{E}^k_{p,q}\, e^{j\theta^c_{p,q}} + \mathbf{F}^k_{p,q}\, e^{-j\theta^c_{p,q}} + \mathbf{T}^{k,2}_{\overline{p},\overline{q}}\right) \Big] + c_{\overline{p},\overline{q}},
\end{aligned} \tag{2.32}$$

where $c_{\overline{p},\overline{q}}$ are terms that are independent of $\theta^c_{p,q}$. Also, the summation in (2.32) is over all the users in cell c. The steps leading to these expressions are derived in Appendix C. The matrices $\mathbf{C}^k_{p,q}, \mathbf{D}^k_{p,q}$ are defined in equation (30) of the Appendix C, respectively. The definition of $\mathbf{E}^k_{p,q}, \mathbf{F}^k_{p,q}$ is similar to the matrices $\mathbf{C}^k_{p,q}, \mathbf{D}^k_{p,q}$, with $\widehat{\mathbf{B}}_k$ replaced by $(\widehat{\mathbf{A}}_k + \lambda_c \mathbf{I})$. Here $\mathbf{T}^{k,1}_{\overline{p},\overline{q}}, \mathbf{T}^{k,2}_{\overline{p},\overline{q}}$ are matrices with terms independent of $\theta^c_{p,q}$. Setting the derivative of (2.32) w.r.t. $\theta^c_{p,q}$ to zero we get

$$\begin{aligned}
e^{j\theta^c_{p,q}} \sum_{k:b_k=c} \mathrm{tr}\{\widetilde{\mathbf{W}}_k \mathbf{C}^k_{p,q} - \mathbf{E}^k_{p,q}\} &= e^{-j\theta^c_{p,q}} \sum_{k:b_k=c} \mathrm{tr}\{\widetilde{\mathbf{W}}_k \mathbf{D}^k_{p,q} - \mathbf{F}^k_{p,q}\} \\
\text{where} \quad \widetilde{\mathbf{W}}_k &= u_k \left(\mathbf{I} + \mathbf{G}_k^H \mathbf{V}^{b_k\,H} \widehat{\mathbf{B}}_k \mathbf{V}^{b_k} \mathbf{G}_k\right)^{-1}.
\end{aligned} \tag{2.33}$$

This leads to two extrema for $\theta^c_{p,q}$ of which the best one needs to be chosen:

$$\begin{aligned}
\theta^c_{p,q} &= \arg \max_{\theta^{c1}_{p,q}, \theta^{c2}_{p,q}} f\left(\theta^c_{p,q}\right), \\
\theta^{c1}_{p,q} &= -\frac{\angle a}{2}, \\
\theta^{c2}_{p,q} &= \pi - \frac{\angle a}{2}, \\
a &= \frac{\sum_{k:b_k=c} \mathrm{tr}\{\widetilde{\mathbf{W}}_k \mathbf{C}^k_{p,q} - \mathbf{E}^k_{p,q}\}}{\sum_{k:b_k=c} \mathrm{tr}\{\widetilde{\mathbf{W}}_k \mathbf{D}^k_{p,q} - \mathbf{F}^k_{p,q}\}}.
\end{aligned} \tag{2.34}$$

Alternating WSR maximization between digital and analog BF now leads to Algorithm 2.

Algorithm 2 Hybrid BF Design via Alternating Minorizer

Given: $P_c, \mathbf{H}_{k,c}, u_k \, \forall k, c$.

Initialization: $\mathbf{V}^c = e^{j\angle \mathbf{V}_{1:M^c} \left(\sum_{k:b_k=c} \Theta_k^c, \sum_{i:b_i \neq c} \Theta_i^c \right)}$,
The $\mathbf{G}_k$ are taken as the ZF precoders for the effective channels $\mathbf{H}_{k,b_k}\mathbf{V}^{b_k}$ with uniform powers.

Iteration (j) :

1. Compute $\widehat{\mathbf{B}}_k, \widehat{\mathbf{A}}_k, \ \forall k$ from (2.26).
2. Update $\mathbf{G}_k^{'(j)}, \ \forall k$, from (2.27).
3. Update $\mathbf{P}_k$ and $\lambda_c, \ \forall k, c$ from (2.29).
4. Update $\left(\mathbf{V}_{p,q}^c \right)^{(j)}, \ \forall c, \forall (p,q)$, from (2.34) (phasor constrained) or from (2.30) (unconstrained).
5. Check for convergence of the WSR: if not go to step 1.

2.2.4.1 Mixed time scale HBF

In our first work on hybrid beamforming [10], we considered the hybrid beamforming design using a weighted sum MSE (WSMSE) based approach and derived a similar iterative algorithm as above. We consider two variants of the WSMSE iterative algorithm. In the first variant, Fast Time Scale Adaptation, the WSR is maximized straightforwardly using Algorithm 1, using (perfect) instantaneous CSIT for the computation of both analog and digital beamformers V and g. This adaptation is repeated whenever the instantaneous CSIT (the channels $H_{k,c}$) changes.

In the second variant, Mixed Time Scale Adaptation, the overall Fast Time Scale Adaptation just mentioned gets executed only from time to time, whenever the slow CSIT, here captured by the channel covariance matrices Θ_k^c, changes. In between those slow CSIT updates, the digital beamformers g and all auxiliary quantities, but not the analog beamformers, get updated whenever the fast CSIT changes. This can be done using Algorithm 1, in which step 5., the update of the analog beamformers, gets skipped. Hence, the values of the analog beamformers are frozen over the slow fading coherence time, whereas only the digital parts get updated at the fast fading rate. Whenever the slow fading CSIT is considered to have changed, all quantities are updated using the instantaneous CSIT available at such time instant. No dynamics of the fast or slow fading processes get exploited. When an update gets performed, all quantities to be updated get recomputed from scratch. The information in the previous updates gets ignored, except for providing the initialization values. The initialization mentioned in Algorithm 1 gets performed only once, at the very first initialization of the whole process.

2.2.4.2 Hybrid Beamformer Capabilities

In this section, we analyze to what extent a hybrid BF can achieve the same performance as a fully digital BF. In particular, we shall see that this is possible for a sufficient number of RF chains and with the antenna array responses being phasors. Consider a specular or pathwise channel model with say L multi-paths per link. For notational simplicity, we shall consider a uniform L and $N_k = N_r, \forall k$. Let the antenna array response for BS c be $\mathbf{h}_t^c(\phi)$ for Angle of Departure (AoD)

ϕ. We assume that all entries of $\mathbf{h}_t^c(\phi)$ have the same magnitude. This assumption is necessary for the following theorem to be valid and it is necessitated by the unit magnitude constraints on the analog BF. Then the collective $N_t \times L$ multipath Tx array response $\mathbf{H}_{t,k}$ for the downlink channel of user k is

$$\mathbf{H}_{t,k}^c = \left[\mathbf{h}_t^c\left(\phi_{k,1}\right)\, \mathbf{h}_t^c\left(\phi_{k,2}\right)\, ...\, \mathbf{h}_t^c\left(\phi_{k,L}\right)\right]^*, \tag{2.35}$$

and the concatenated antenna array response matrix to all users can be written as,

$$\overline{\mathbf{H}}_t^c = \left[\mathbf{H}_{t,1}^c\, \mathbf{H}_{t,2}^c\, ...\, \mathbf{H}_{t,K}^c\right], \tag{2.36}$$

of dimension $N_t \times N_p$, where we denote the total number of paths $N_p = LK$. Similarly, we define $\overline{\mathbf{H}}_r^c$ and $\overline{\mathbf{A}}^c$ for the concatenated Rx antenna array responses and complex path amplitudes. $\overline{\mathbf{A}}^c$ is an $N_p \times N_p$ block diagonal matrix with blocks of size $L \times L$ and $\overline{\mathbf{H}}_r^c$ is a $KN_r \times N_p$ block diagonal matrix with blocks of size $N_r \times L$. Finally, we can write the $KN_r \times N_t$ MIMO channel from BS c to all a users as

$$\mathbf{H}^{c\,H} = \overline{\mathbf{H}}_t^c \overline{\mathbf{A}}^{c\,H} \overline{\mathbf{H}}_r^{c\,H}. \tag{2.37}$$

Theorem 1. *For a multi-cell MU MIMO system with $M \geq N_p$ and phasor antenna responses, to achieve optimal all-digital precoding performance, the analog beamformer can be chosen as the Tx side concatenated antenna array response.*

Proof: From [9] or [11, eq. (13)], the optimal all-digital beamformer is of the form

$$\begin{aligned}(\mathbf{H}^{c\,H}\mathbf{D}_1^c\mathbf{H}^c + \lambda_c\mathbf{I})^{-1}\mathbf{H}^{c\,H}\mathbf{D}_2^c &= \mathbf{H}^{c\,H}\mathbf{B}^c \\ &= \overline{\mathbf{H}}_t^c\, \overline{\mathbf{A}}^{c\,H} \overline{\mathbf{H}}_r^{c\,H}\mathbf{B}^c \\ \text{where}\quad \mathbf{B}^c &= \left(\lambda_c\mathbf{I} + \mathbf{D}_1^c\mathbf{H}^c\mathbf{H}^{c\,H}\right)^{-1}\mathbf{D}_2^c,\end{aligned} \tag{2.38}$$

$\mathbf{D}_1^c$, $\mathbf{D}_2^c$ are block diagonal matrices and we used the identity $(\mathbf{I}+\mathbf{XY})^{-1}\mathbf{X} = \mathbf{X}(\mathbf{I}+\mathbf{YX})^{-1}$. Under the Theorem assumptions we can then separate the BFs as

$$\begin{aligned}\mathbf{V}^c &= \overline{\mathbf{H}}_t^c, \\ \mathbf{G}^c &= \overline{\mathbf{A}}^{c\,H}\overline{\mathbf{H}}_r^{c\,H}\mathbf{B}^c.\end{aligned} \tag{2.39}$$

Hence $\mathbf{V}$ depends only on the Tx antenna array responses. □

Note that whereas the digital BF $\mathbf{G}$ in (2.27) is a function of the instantaneous CSIT, the analog BF $\mathbf{V}^c$ is only a function of AoDs, hence only of the slow fading channel components. This explains why the outdated CSIT based update for $\mathbf{V}$ in a mixed time scale scenario in [10] has a performance close to that of an instantaneous CSIT update based $\mathbf{V}$. Also, the theorem above motivates us to use the concatenated antenna array response matrix as the initialization of the analog BF for the Algorithm 1, when the number of RF chains M is greater than N_p or even when it is not, by taking the M strongest paths.

2.2.4.3 Deterministic Annealing for Global Convergence

In this section, we analyze how to improve the performance of the alternating optimization algorithm proposed (Algorithm 2) in the scenario in which the number of specular paths across all users exceeds the number of RF chains. In the previous sections, we considered the hybrid beamforming design using the WSR cost function which is a non-convex function. Due to which the

algorithm will converge to different local optima depending on the initialization. So we consider here one approach called deterministic annealing (DA) to avoid the problem of local optima. In DA, we use a temperature parameter to track the global optimum with a homotopy method starting from a convex problem. Starting with a high temperature, where we know the optimal solution, we slowly decrease the temperature to reach the desired solution. If at the high temperature we know the global optimum value, then if the temperature variations are slow, at the next value the global optimum will have the previous solution in its region of attraction. For the analog beamforming design using phasors, simulation results show that it converges to a local optimum and that it is very sensitive to the initialization used. In DA, we start from the optimal unconstrained $\mathbf{V}$ (note that HBF with factored digital and analog BFs has its own convexity issues that can be resolved with a separate DA strategy as in [11]). Then the gradual forcing of the amplitude of the unconstrained $\mathbf{V}$ entries to 1 allows to approach the global optimum. Here the amplitude relaxation parameter of each $\mathbf{V}$ entry is related to the temperature parameter. Note that in resulting Algorithm 2, d is some constant smaller than 1, say 0.9. The number of iterations required is a number of time constants of $e^{d\,t}$.

Algorithm 3: Deterministic Annealing for Analog Beamformer

Let $\mathbf{V}^c_{i,j} = |\mathbf{V}^c_{i,j}|e^{j\theta^c_{i,j}}$. Let the unconstrained $\mathbf{V}^c$ design (joint $\mathbf{V}^c$ and all $\mathbf{G}_k$) using Algorithm 2 converge first.

1. Scale $\forall\,(i,j) : |\mathbf{V}^c_{i,j}| \leftarrow e^{d\ln|\mathbf{V}^c_{i,j}|}$.
2. Reoptimize all $\theta^c_{i,j}$ and all digital BFs using Algorithm 2.
3. Update stream powers and Lagrange multipliers.
4. Go to step 1 for a number of iterations.
5. Finally redo steps 2-3 a last time with all $|\mathbf{V}^c_{i,j}| = 1$ in 1.

2.2.5 Simulation results

In this section, we evaluate the proposed algorithm using simulation results, which will be limited to a single cell MISO system. We used a simple channel model similar to [23], which is based on a uniform linear array (ULA). Assume that there are L multi-path components between a user and the base station. The channel between the base station and k^{th} user can be written as $h_k = \sum_{i=1}^{L} \alpha_{k,i}\, a_t\left(\phi_{k,l}\right)$. Assuming $\lambda/2$ as the antenna spacing, the antenna array response is written as

$$a_t\left(\phi\right) = \left[1\; e^{j\pi\sin(\phi)} \ldots e^{j\pi(N_t-1)\sin(\phi)}\right]^T. \tag{2.40}$$

The complex path amplitudes are modeled as Rayleigh fading $\alpha_{k,l} \sim \mathscr{CN}(0,\sigma^2_{\alpha_l})$, where $\sigma^2_{\alpha_l}$ is assumed to be from an exponential distribution with parameter 1. The $\phi_{k,l}$ are taken from a Laplacian distribution with an angular spread as 10 degrees. These angle values and path powers are fixed for all channel realizations (slow fading). In each channel realization, what changes is the complex path gains $\alpha_{k,l}$ (fast fading).

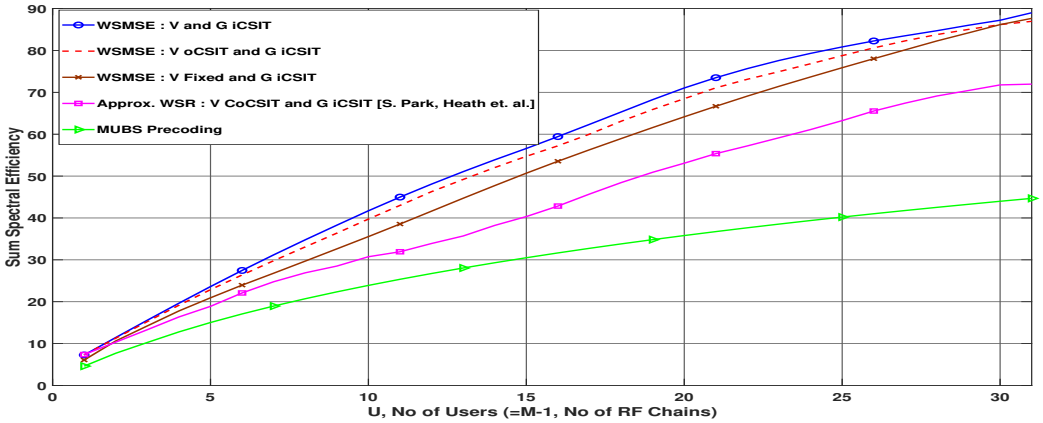


Figure 2.4: Sum Spectral Efficiency vs No of Users, $U = M - 1$, M is no of RF Chains. $N_t = 32$, $SNR = 20\,dB$ and $L = 6$ paths.

Notations used for the figures: oCSIT refers to outdated CSIT, iCSIT means instantaneous CSIT and CoCSIT implies covariance CSIT. V Fixed in the figures refers to the case where V is fixed to be the M dominant eigenvectors of the sum of the users channel covariance matrix.

2.2.5.1 BF initializations for the algorithm

In this thesis, for the simulations, corresponding to fully or hybrid beamforming schemes, the BFs are initialized as follows. We use the concept of deterministic annealing proposed for a fully digital solution in [32]. At low SNR, an optimal BF solution corresponds to matched filtering. Starting with this solution for the fully digital BF at low SNR, we optimize the BFs using the alternating minorization concept proposed here. Further, at any SNR, the converged values from the previous iterations are used as the initialization point. This process is followed in the simulations of all other chapters (where BF techniques are discussed) in this thesis.

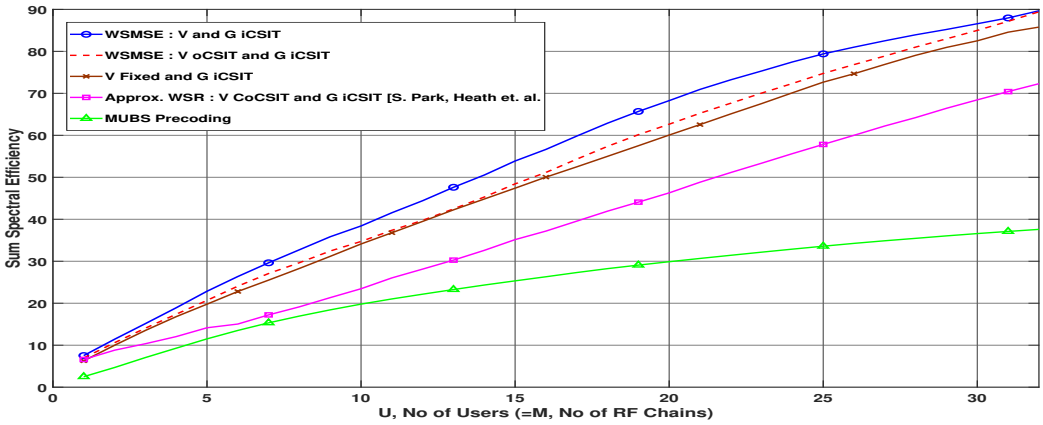


Figure 2.3: Sum Spectral Efficiency vs No of Users U = M , M is no of RF Chains. $N_t = 32$, $SNR = 20\,dB$ and $L = 6$ paths.

2.2.5.2 Simulations for WSMSE based HBF

In all the figures, the simulations are done for $SNR = 20dB$, $L = 6$. Since single cell, $C = 1$ and the number of users is denoted by U.

The sum spectral efficiency (sum of the rates of the U users) is plotted versus the number of users and compared with the sub-optimal algorithms proposed in [23] and [25] ("'MUBS Precoding"', MUBS refers to a multi-user beam steering scheme). In [23], V is computed using covariance CSIT and the g are updated with instantaneous CSIT. The interesting part about our work is in showing that the analog beamformer need not be adapted at the fast fading rate. For the comparison to be fair, the proposed update of the analog beamformer with outdated CSIT and digital beamformer with instantaneous CSIT is compared with prior works which use covariance CSIT for V and instantaneous CSIT for g. It is evident from all the figures that our approach based on Mixed Time Scale WSMSE Adaptation outperforms those of [23] and [25]. In Figure 1, $N_t = 32$ and the number of users is equal to the number of RF chains ($U = M$). Simulations are done for the two variants of CSIT. In the first case, V and g are iteratively updated w.r.t the instantaneous channel. In the second case, we consider two realizations of the channels $(h_k(1), h_k(2))$. For $h_k(1)$, V and g are updated as per Algorithm 1. For $h_k(2)$, the result obtained with $h_k(1)$ is used for V and is not updated, whereas the g are updated. So this is the case of outdated CSIT for V. The results are averaged over 40 such pairs of channel realizations, each time with the same channel covariances. From Figure 2.3 we can see that in the second case with outdated CSIT for V, performance is slightly degraded but still much better than the suboptimal algorithms of [23] and [25].

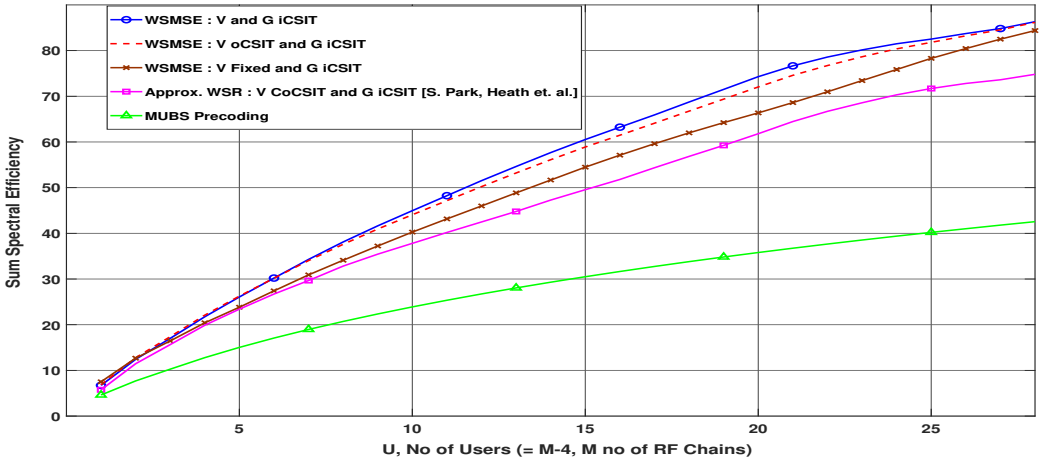


Figure 2.5: Sum Spectral Efficiency vs No of Users, U= M -4, M is no of RF Chains. $N_t = 32$, $SNR = 20\,dB$ and $L = 6$ paths.

In Figures 2.4 and 2.5, we repeat the same simulation as described above with $U = M - 1$ and $U = M - 4$ respectively. It can be seen by comparing Figures 2.3 and 2.4 that when there is an excess of RF chains over the number of users, the spectral efficiency increases. In Figure 2.6, we have considered the case where $N_t = 64$ and $U = M$. In Figures 2.7 and 2.8, we consider $N_t = 64$ antennas with $U = M - 1$ and $U = M - 4$ respectively.

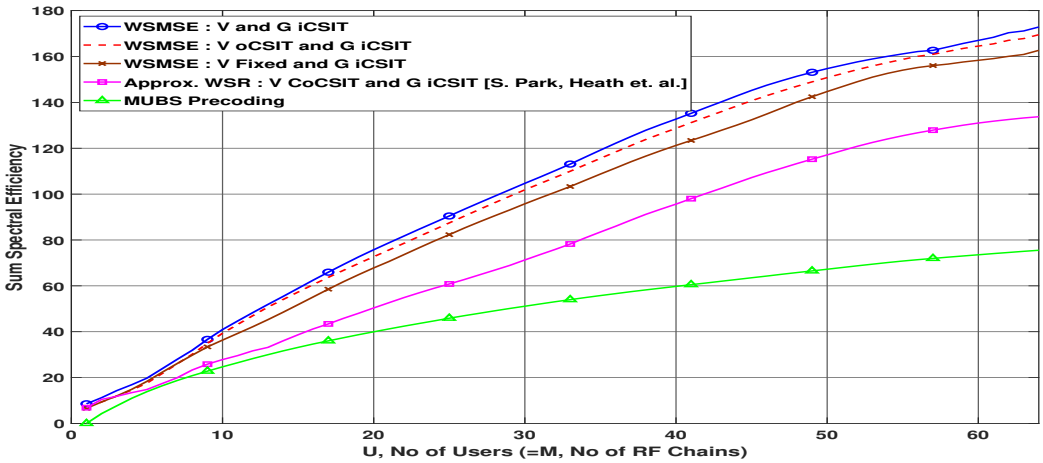


Figure 2.6: Sum Spectral Efficiency vs No of Users, U = M, M is no of RF Chains. $N_t = 64$, $SNR = 20\,dB$ and $L = 6$ paths.

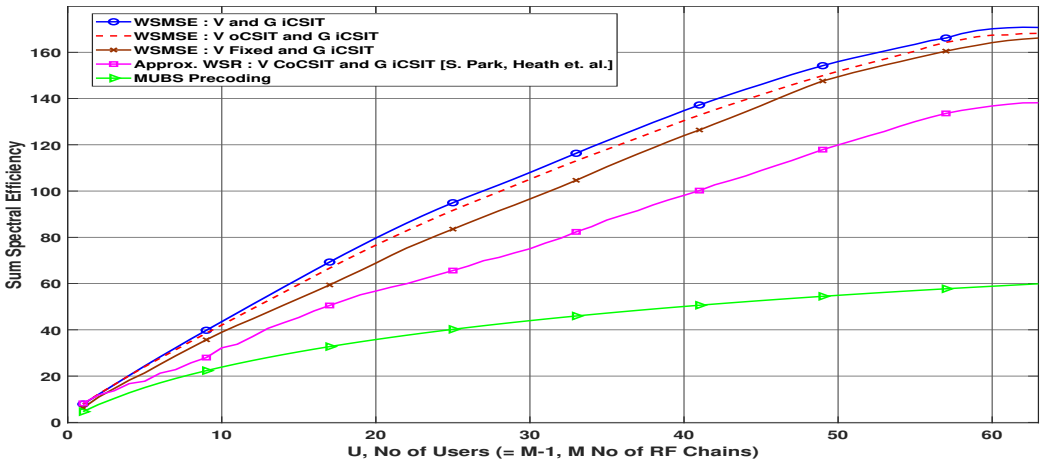


Figure 2.7: Sum Spectral Efficiency vs No of Users, U = M-1, M is no of RF Chains. $N_t = 64$, $SNR = 20\,dB$ and $L = 6$ paths.

2.2.5.3 Deterministic Annealing based HBF

In this subsection, we validate the simulation results for our proposed deterministic annealing based HBF design compared to other state of the art solutions. Notations used in the figure: CoCSIT refers to covariance CSIT and EV refers to dominant eigenvectors of the sum of the channel covariance matrices of all users. We compare the performance of the proposed algorithms with the WSMSE based fully digital BF [9] (referred to as "WSMSE Fully Digital [Christensen et al]"), approximate WSR based hybrid design [24] (referred to as "Approximate WSR [Sohrabi, Wei Yu]", WSMSE based alternating optimization [10] (referred to as "WSMSE HBF") and the covariance CSIT based scheme [23] (referred to as "V CoCSIT and G R-ZF [S.Park et al]"). "V Random Initialization" refers to the case when Algorithm 1 starts with random phases for the analog BF.

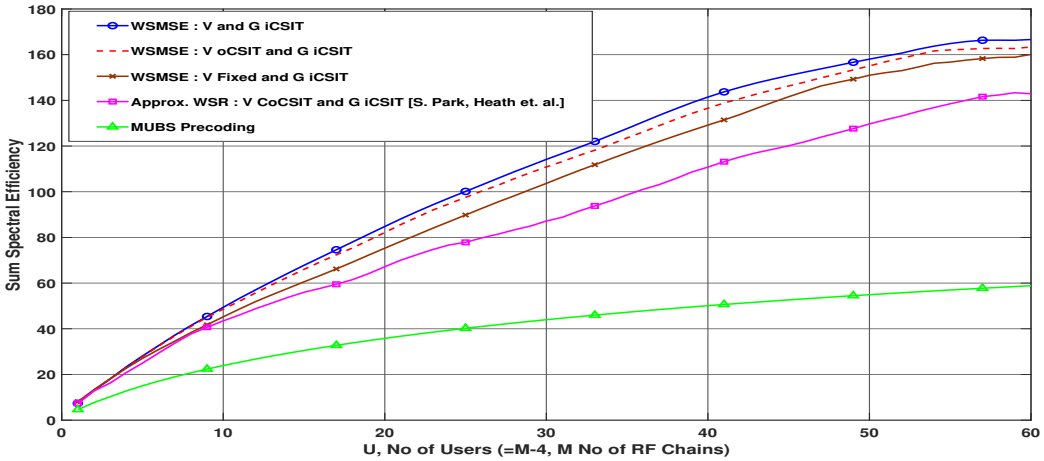


Figure 2.8: Sum Spectral Efficiency vs No of Users, U = M-4, M is no of RF Chains. $N_t = 64$, $SNR = 20\,dB$ and $L = 6$ paths.

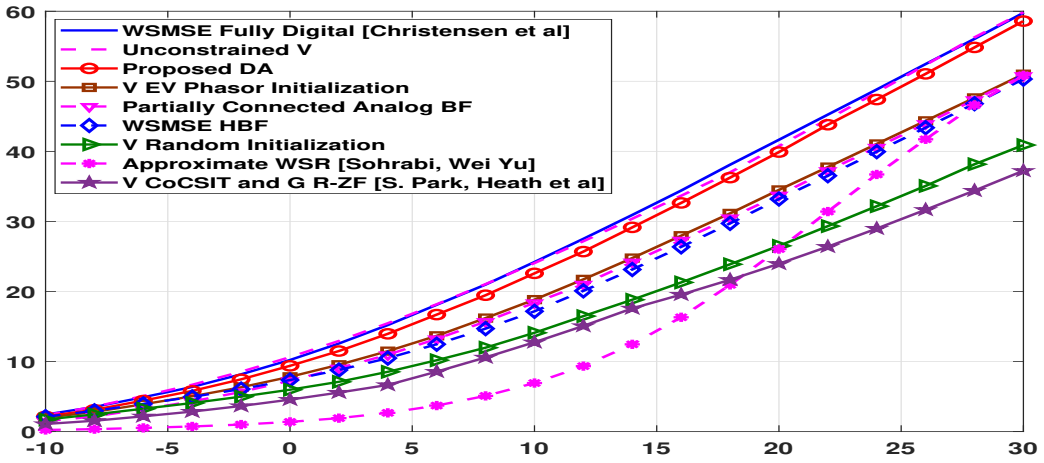


Figure 2.9: Sum Rate comparisons for, $N_t = 32, M = 16, K = 8, C = 1, L = 4$ paths.

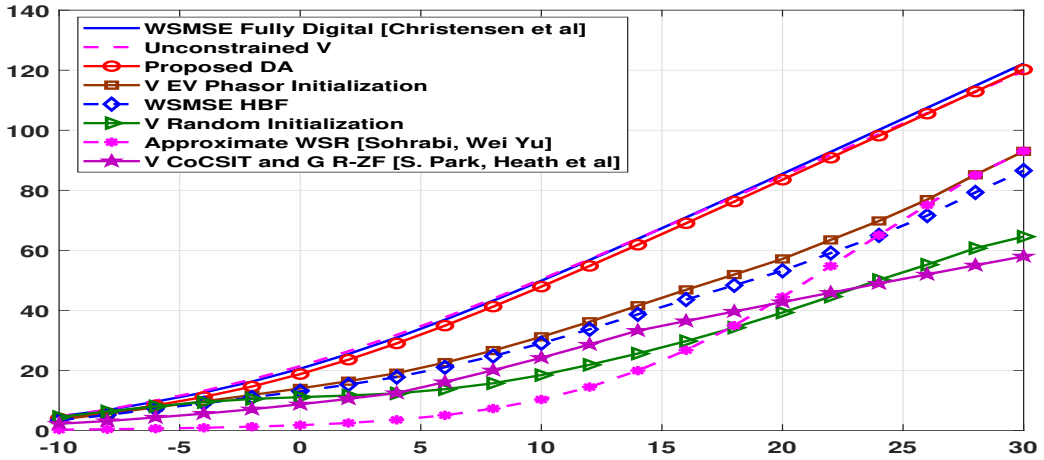


Figure 2.10: Sum Rate comparisons for, $N_t = 64, M = 16, K = 16, C = 1, L = 2$ paths.

It is evident from the figures that the DA based approach (Algorithm 3) performs significantly better than just alternating optimization (Algorithm 2) and also the state of the art methods. DA also has a performance close to the fully digital performance.

2.3 Hybrid Beamforming under Realistic Power Constraints

In contrast to the conventional (sum-)power constraint (SPC) on the base station (BS), this section considers a more realistic scenario with additionally per-RF or per-antenna power constraints (PRFPC/PAPC). In practice, each RF chain is equipped with a power amplifier and its linear range of the PA combined with Peak to Average Power Ratio (PAPR) considerations lead to a power constraint per power amplifier. Another scenario is the case of a distributed system where a central BS is connected via a high speed backbone network to remote antennas. Fully digital BF designs with PAPC can be found in [33–36]. [33] focuses on the design of BF vectors for a MISO system to minimize the per-antenna power while enforcing a set of SINR constraints for each user. ZF BF design with PAPC are discussed in [34], while [35] utilizes UL/DL duality of the sum MSE for the precoder design. Existing approaches for this problem are based on either interior point methods that do not favorably scale with the problem size or subgradient methods [37] that have a very slow convergence rate. We propose a novel HBF design (for both fully or partially connected structures) based on the WSR criterion which is simplified using minorization and alternating optimization. To our best knowledge, this is the first work to propose HBF design under the more realistic scenario of per-RF or per-antenna power constraints. We propose a novel interference leakage aware water-filling (ILA-WF) for the stream power optimization, even for just SPC, but also augmented with PRFPC or PAPC. We propose to solve the resulting convex Lagrange dual problem by alternating bisection but may other solutions can be considered. The ILA-WF allows automatic discovery of the sustainable number of streams per user in MIMO channels.

There exist two types of phased arrays at mmWave frequencies: (i) passive phased arrays and (ii) active phased arrays [38]. Though passive phase shifters incur some power loss, they require only the same number of power amplifiers as RF units, leading to PRFPC considerations. Since there is a clear trend towards active systems, we also consider PAPC in section 2.3.4. Although we do not model the power loss (which would complete the picture), simulations show that due to the reduced number of power constraints, passive systems with PRFPC have some power efficiency gain over active systems with PAPC. The per-RF power constraints (PRFPC) at BS c can be written as

$$\sum_{k:b_k=c} \left[\mathbf{G}_k \mathbf{G}_k^H\right]_{i,i} \leq a_i^c,\ i = 1,...,M^c, \tag{2.41}$$

where $\left[\mathbf{G}_k \mathbf{G}_k^H\right]_{i,i}$ represents the i^{th} diagonal element of $\mathbf{G}_k \mathbf{G}_k^H$. Further, also total Tx power constraints need to be satisfied, $\sum_{k:b_k=c} \mathrm{tr}\{Q_k\} \leq P^c$. The WSR problem is non-concave in the $\mathbf{Q}_k$ due to the interference terms. Therefore finding the global optimum is challenging. To render a feasible solution, we consider constructing a minorizer based on the difference of convex functions (DC programming) approach. Consider the dependence of WSR on $\mathbf{Q}_k$ alone.

$$\begin{aligned} WSR &= u_k \ln\det\left(\mathbf{R}_{\bar{k}}^{-1}\mathbf{R}_k\right) + WSR_{\bar{k}}, \\ WSR_{\bar{k}} &= \sum_{i=1,\neq k}^{K} u_i \ln\det\left(\mathbf{R}_{\bar{i}}^{-1}\mathbf{R}_i\right), \end{aligned} \tag{2.42}$$

where $\ln\det\left(\mathbf{R}_{\overline{k}}^{-1}\mathbf{R}_k\right)$ is concave in $\mathbf{Q}_k$ and $WSR_{\overline{k}}$ is convex in $\mathbf{Q}_k$. Since a linear function is simultaneously convex and concave, DC programming [30] introduces the first order Taylor series expansion of $WSR_{\overline{k}}$ in $\mathbf{Q}_k$ around $\widehat{\mathbf{Q}}$ (i.e. all $\widehat{\mathbf{Q}}_i$).

$$\begin{aligned}\underline{WSR}_{\overline{k}}\left(\mathbf{Q}_k,\widehat{\mathbf{Q}}\right) &= WSR_{\overline{k}}\left(\widehat{\mathbf{Q}}_k,\widehat{\mathbf{Q}}\right) - \mathrm{tr}\left\{\left(\mathbf{Q}_k-\widehat{\mathbf{Q}}_k\right)\widehat{\mathbf{A}}_k\right\},\\ \widehat{\mathbf{A}}_k &= -\left.\frac{\partial WSR_{\overline{k}}\left(\mathbf{Q}_k,\widehat{\mathbf{Q}}\right)}{\partial \mathbf{Q}_k}\right|_{\widehat{\mathbf{Q}}_k}\\ &= \sum_{i=1,\neq k}^{K} u_i \mathbf{H}_{i,b_k}^H\left(\widehat{\mathbf{R}}_{\overline{i}}^{-1}-\widehat{\mathbf{R}}_i^{-1}\right)\mathbf{H}_{i,b_k}.\end{aligned} \tag{2.43}$$

Note that the linearized tangent expression $\underline{WSR}_{\overline{k}}$ constitutes a (touching) lower bound for $WSR_{\overline{k}}$ via $-\mathrm{tr}\{\mathbf{R}^{-1}\mathbf{\Delta}\} \le -\ln\det(\mathbf{R}^{-1}(\mathbf{R}+\mathbf{\Delta}))$ and $\mathbf{R}_k \succeq \mathbf{R}_{\overline{k}}$. Hence the DC approach is also a minorization approach [27], regardless of the (re)parameterization of $\mathbf{Q}$. Now let $\widehat{\mathbf{B}}_k = \mathbf{H}_{k,b_k}^H\widehat{\mathbf{R}}_{\overline{k}}^{-1}\mathbf{H}_{k,b_k}$, $\mathbf{\Psi}_c = \mathrm{diag}\left(\Psi_{c,1},....,\Psi_{c,M^c}\right)$ represents the Lagrange multipliers associated with the per-RF power constraints $\mathbf{\Phi}_c = \mathrm{diag}\left(a_1^c,...,a_{M^c}^c\right)$. $\mathbf{\Psi}$ represents the set of all $\mathbf{\Psi}_c$ and $\mathbf{\Lambda} = \mathrm{diag}\,(\lambda_1,...,\lambda_C)$. Then, dropping constant terms, reparameterizing the $\mathbf{Q}_k$ as in (2.22), performing this linearization for all users, and augmenting the WSR cost function with the Tx power constraints, we get the Lagrangian (2.44) which gets maximized alternatingly [27] between digital and analog BF

$$\begin{aligned}\mathscr{L}\left(\mathbf{V},\mathbf{G},\mathbf{\Lambda},\mathbf{\Psi}\right) = &\sum_{c=1}^{C}\lambda_c P^c + \sum_{c=1}^{C}\mathrm{tr}\{\mathbf{\Psi}_c\mathbf{\Phi}_c\} + \sum_{k=1}^{K} u_k \ln\det\left(\mathbf{I}+\mathbf{G}_k^H\mathbf{V}^{b_k H}\widehat{\mathbf{B}}_k\mathbf{V}^{b_k}\mathbf{G}_k\right)\\ &-\mathrm{tr}\left\{\mathbf{G}_k^H\left(\mathbf{V}^{b_k H}\left(\widehat{\mathbf{A}}_k+\lambda_{b_k}\mathbf{I}\right)\mathbf{V}^{b_k}+\mathbf{\Psi}_{b_k}\right)\mathbf{G}_k\right\}.\end{aligned} \tag{2.44}$$

2.3.1 Digital BF Design

Maximizing (2.44) w.r.t. $\mathbf{G}_k$ leads to the KKT conditions

$$\mathbf{V}^{b_k H}\widehat{\mathbf{B}}_k\mathbf{V}^{b_k}\mathbf{G}_k = \left(\mathbf{V}^{b_k H}\left(\widehat{\mathbf{A}}_k+\lambda_{b_k}\mathbf{I}\right)\mathbf{V}^{b_k}+\mathbf{\Psi}_{b_k}\right)\mathbf{G}_k\tfrac{1}{u_k}\left(\mathbf{I}+\mathbf{G}_k^H\mathbf{V}^{b_k H}\widehat{\mathbf{B}}_k\mathbf{V}^{b_k}\mathbf{G}_k\right) \tag{2.45}$$

with solution d_k dominant generalized eigenvectors (g.e.v.)

$$\mathbf{G}_k^{'} = \mathbf{V}_{1:d_k}\left(\mathbf{V}^{b_k H}\widehat{\mathbf{B}}_k\mathbf{V}^{b_k},\mathbf{V}^{b_k H}\left(\widehat{\mathbf{A}}_k+\lambda_{b_k}\mathbf{I}\right)\mathbf{V}^{b_k}+\mathbf{\Psi}_{b_k}\right) \tag{2.46}$$

with eigenvalues $\mathbf{\Sigma}_k = \frac{1}{u_k}\left(\mathbf{I}+\mathbf{G}_k^H\mathbf{V}^{b_k H}\widehat{\mathbf{B}}_k\mathbf{V}^{b_k}\mathbf{G}_k\right)$. The gradient in (2.45), which would be the same with $\underline{WSR}$ replaced by WSR, leads to g.e.v. conditions whereas maximizing $\mathscr{L}$ in (2.44) leads to select the dominant g.e.v. Let $\mathbf{S}_k = \mathbf{G}_k'^H\mathbf{V}^{b_k H}\widehat{\mathbf{B}}_k\mathbf{V}^{b_k}\mathbf{G}_k'$, $\mathbf{W}_k = \mathbf{G}_k'^H\mathbf{V}^{b_k H}\widehat{\mathbf{A}}_k\mathbf{V}^{b_k}\mathbf{G}_k'$, and $\mathbf{T}_k(\lambda_{b_k},\mathbf{\Psi}_{b_k}) = \mathbf{W}_k + \mathbf{G}_k'^H\left(\lambda_{b_k}\mathbf{V}^{b_k H}\mathbf{V}^{b_k}+\mathbf{\Psi}_{b_k}\right)\mathbf{G}_k'$. Note that g.e.v. diagonalize $\mathbf{S}_k$, $\mathbf{T}_k\left(\lambda_{b_k}^{(j-1)},\mathbf{\Psi}_{b_k}^{(j-1)}\right)$ and $\mathbf{\Sigma}_k$ (see further for iteration index (j)). As g.e.v. are normalized, the stream powers $\mathbf{P}_k \succeq 0$ (diagonal) need to be optimized separately. But this is straightforward from (2.44): substituting $\mathbf{G}_k = \mathbf{G}_k'\mathbf{P}_k^{\frac{1}{2}}$ in (2.44) yields,

$$\begin{aligned}\mathscr{L}\left(\mathbf{V},\mathbf{G}',\mathbf{P},\mathbf{\Lambda},\mathbf{\Psi}\right) = &\sum_{c=1}^{C}\left(\lambda_c P^c + \mathrm{tr}\{\mathbf{\Psi}_c\mathbf{\Phi}_c\}\right) +\\ &\sum_{k=1}^{K}\left[u_k\ln\det\left(\mathbf{I}+\mathbf{S}_k\mathbf{P}_k\right) - \mathrm{tr}\{\mathbf{T}_k\left(\lambda_{b_k},\mathbf{\Psi}_{b_k}\right)\mathbf{P}_k\}\right].\end{aligned} \tag{2.47}$$

2.3.2 Optimization of Power Variables

The optimization of (2.47) w.r.t. $\mathbf{P}_k$ leads to the following interference leakage aware water-filling (ILA-WF)

$$\begin{aligned}\left(u_k\left(\mathbf{W}_k+\mathbf{G}_k'^H\left(\lambda_{b_k}\mathbf{V}^{b_k H}\mathbf{V}^{b_k}+\mathbf{\Psi}_{b_k}\right)\mathbf{G}_k'\right)^{-1}-\mathbf{S}_k^{-1}\right)^+ &= \mathbf{P}_k^o\left(\lambda_{b_k},\mathbf{\Psi}_{b_k}\right)\\ &= \left(u_k\mathbf{T}_k^{-1}\left(\lambda_{b_k},\mathbf{\Psi}_{b_k}\right)-\mathbf{S}_k^{-1}\right)^+\end{aligned} \tag{2.48}$$

where $(\mathbf{X})^+$ denotes the positive semi-definite part of Hermitian $\mathbf{X}$. We substitute the optimized power distribution $\mathbf{P}_k^o(\lambda_{b_k},\mathbf{\Psi}_{b_k})$ in (2.47) yielding the Lagrange dual function

$$\begin{aligned}g\left(\mathbf{\Lambda},\mathbf{\Psi}\right) &= \mathscr{L}\left(\mathbf{V},\mathbf{G}',\mathbf{P}^o\left(\mathbf{\Lambda},\mathbf{\Psi}\right),\mathbf{\Lambda},\mathbf{\Psi}\right)\\ &= \sum_{c=1}^{C} g_c\left(\lambda_c,\mathbf{\Psi}_c\right)\\ g_c(\lambda_c,\mathbf{\Psi}_c) &= \lambda_c P^c+\mathrm{tr}\{\mathbf{\Psi}_c\mathbf{\Phi}_c\} \;+\sum_{k:b_k=c}\left[u_k\ln\det\left(\mathbf{I}+\mathbf{S}_k\mathbf{P}_k^o\right)-\mathrm{tr}\{\mathbf{T}_k\left(\lambda_{b_k},\mathbf{\Psi}_{b_k}\right)\mathbf{P}_k^o\}\right]\end{aligned} \tag{2.49}$$

where we omitted the dependence of $g()$ on $\mathbf{V},\mathbf{G}'$, which are currently fixed in the alternating optimization process, as we maximize over $\mathbf{P}$. $\mathbf{\Lambda},\mathbf{\Psi}$ should be chosen such that $g(\mathbf{\Lambda},\mathbf{\Psi})$ is finite. Further, the non-negativity of $\mathbf{\Lambda}$ and $\mathbf{\Psi}$ imposes constraints on the dual objective function. Formally, the Lagrangian dual problem per cell can be stated as follows:

$$\min_{\lambda_c,\mathbf{\Psi}_c} g_c\left(\lambda_c,\mathbf{\Psi}_c\right) \text{ subject to } \lambda_c\geq 0,\mathbf{\Psi}_c\succeq 0\,,\ \forall c\,. \tag{2.50}$$

Since the dual function $g_c(\lambda_c,\mathbf{\Psi}_c)$ is the pointwise supremum of a family of functions of $\lambda_c,\mathbf{\Psi}_c$, it is convex [39] and the globally optimal value $\lambda_c,\mathbf{\Psi}_c$ can be found by a multitude of convex optimization techniques. We propose to use the alternating bisection method as in Algorithm 4. This requires to specify search ranges. We can take the lower bounds $\left(\underline{\lambda}_c,\underline{\Psi}_{c,i}\right)=(0,0)$. The upper bounds are obtained by finding the largest value over users such that the strongest mode of that user loses power with the corresponding power constraint being the only active one: $\overline{\lambda}_c=\max_{k:b_k=c}\left(u_k\mathbf{S}_k-\mathbf{W}_k\right)_{1,1}/\left(\mathbf{G}_k'^H\mathbf{V}^{cH}\mathbf{V}^c\mathbf{G}_k'\right)_{1,1}$ and $\overline{\Psi}_{c,i}=\max_{k:b_k=c}\left(u_k\mathbf{S}_k-\mathbf{W}_k\right)_{1,1}/|\left(\mathbf{G}_k'\right)_{i,1}|^2$. To simplify the description of the method in Algorithm 4, we introduce $\Psi_{c,0}=\lambda_c$. Also, $\mathbf{\Psi}_{c,\bar{i}}$ denotes all components of $\mathbf{\Psi}_c$ except for $\Psi_{c,i}$ and we take some liberty in ordering arguments of $g_c()$. The complexity could be reduced by reducing the bisection search ranges in consecutive sweeps of overall alternating optimization sweeps.

With the optimized λ_{b_k} and $\mathbf{\Psi}_{b_k}$, $\mathbf{P}_k^o\left(\lambda_{b_k},\mathbf{\Psi}_{b_k}\right)$ is no longer diagonal. So consider its eigen decomposition $\mathbf{P}_k^o=\mathbf{U}_k\mathbf{P}_k\mathbf{U}_k^H$ leading to the new diagonal $\mathbf{P}_k$ and absorb the unitary $\mathbf{U}_k$: $\mathbf{G}_k'\leftarrow\mathbf{G}_k'\mathbf{U}_k$. Note that the minorization approach, which avoids introducing Rxs, can at every BF update allow to introduce an arbitrary number of streams per user by determining multiple dominant generalized eigenvectors, and then let the ILA-WF operation decide how many streams can actually be sustained. Given the digital BFs and the Lagrange multipliers, the analog BF $\mathbf{V}^c$ can be found by alternating optimization.

2.3.3 Design of Unconstrained Analog BF

At first, we consider the case in which the analog BF is unconstrained. Hence the resulting design would also apply to more general two-stage BF design [40] in which the outer BF stage ($\mathbf{V}^c$) is in common to all users in a cell.

Algorithm 4: Alternating bisection for Lagrange multipliers

Initialization: $\underline{\Psi}_{c,i} = 0, \overline{\Psi}_{c,i}, \forall c, i.$

for $c = 1, ..., C$
  Repeat until convergence
    for $i = 0, 1, ..., M^c$
      $\Psi_{c,i} = \left(\underline{\Psi}_{c,i} + \overline{\Psi}_{c,i}\right)/2$
      if $g_c\left(\mathbf{\Psi}_{c,\bar{i}}, \underline{\Psi}_{c,i}\right) < g_c\left(\mathbf{\Psi}_{c,\bar{i}}, \overline{\Psi}_{c,i}\right)$, $\overline{\Psi}_{c,i} = \Psi_{c,i}$,
      else $\underline{\Psi}_{c,i} = \Psi_{c,i}$
    end for
end for

2.3.3.1 Fully Connected Case

To optimize $\mathbf{V}^c$, we equate the gradient of (2.44) w.r.t. $\mathbf{V}^c$ to zero. Using $\partial \ln \det \mathbf{X} = \mathrm{tr}\left(\mathbf{X}^{-1}\partial \mathbf{X}\right)$ and $\det(\mathbf{I}_M + \mathbf{AB}) = \det(\mathbf{I}_N + \mathbf{BA})$ from [31], we get

$$\sum_{k:b_k=c} \left(\widehat{\mathbf{B}}_k \mathbf{V}^c \mathbf{G}_k \zeta_k \mathbf{G}_k^H - \left(\widehat{\mathbf{A}}_k + \lambda_c \mathbf{I}\right) \mathbf{V}^c \mathbf{G}_k \mathbf{G}_k^H\right) = 0, \quad \text{with } \zeta_k = u_k \left(\mathbf{I} + \mathbf{G}_k^H \mathbf{V}^{b_k H} \widehat{\mathbf{B}}_k \mathbf{V}^{b_k} \mathbf{G}_k\right)^{-1}. \tag{2.51}$$

Now with $vec(\mathbf{AXB}) = (\mathbf{B}^T \otimes \mathbf{A})\, vec(\mathbf{X})$ [31], we get

$$\mathbf{V}^c = unvec(\mathbf{V}_{max}(\mathbf{B}_c, \mathbf{A}_c)), \text{ with} \tag{2.52}$$

$$\begin{aligned} \mathbf{B}_c &= \sum_{k:b_k=c} \left(\left(\mathbf{G}_k \zeta_k \mathbf{G}_k^H\right)^T \otimes \widehat{\mathbf{B}}_k\right), \\ \mathbf{A}_c &= \sum_{k:b_k=c} \left(\left(\mathbf{G}_k \mathbf{G}_k^H\right)^T \otimes \left(\widehat{\mathbf{A}}_k + \lambda_c \mathbf{I}\right)\right). \end{aligned} \tag{2.53}$$

2.3.3.2 Partially Connected Case

As in [26], in a partially connected phase shifting network, each RF chain is connected to a subset of antennas. Assuming each RF chain is connected to $L_t^c = N_t^c / M^c$ antennas, the analog precoder matrix can be written as a block diagonal matrix

$$\mathbf{V}^c = \begin{bmatrix} \mathbf{v}_1^c & 0 & \dots & 0 \\ 0 & \mathbf{v}_2^c & & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \mathbf{v}_{M^c}^c \end{bmatrix} \tag{2.54}$$

where $\mathbf{v}_i^c \in \mathbb{C}^{L_t^c \times 1}$ (with unit magnitude elements in the phasor case). The advantage of a partially connected structure is that we need only N_t^c phase shifters. But at the cost of degradation in performance compared to a fully connected structure where there is more phase control. We define $\widetilde{\mathbf{B}}_{c,k}$ as the $N_t^c M^c \times N_t^c$ matrix obtained by concatenating the following subsets of columns:

$(i-1)N_t^c+1:(i-1)N_t^c+L_t^c, i=1,...,M^c$ of $\left(\mathbf{G}_k\boldsymbol{\zeta}_k\mathbf{G}_k^H\right)^T\otimes\widehat{\mathbf{B}}_k$. We define $\widetilde{\mathbf{A}}_{c,i}$ similarly and let $\widetilde{\mathbf{V}}^c=\left[\mathbf{v}_1^{cT},\mathbf{v}_2^{cT},...,\mathbf{v}_M^{cT}\right]^T$. Then optimizing (2.44) w.r.t. $\widetilde{\mathbf{V}}^c$ yields

$$\widetilde{\mathbf{V}}^c=\mathbf{V}_{max}\left(\sum_{k:b_k=c}\widetilde{\mathbf{B}}_{c,k},\sum_{k:b_k=c}\widetilde{\mathbf{A}}_{c,k}\right). \tag{2.55}$$

Alternating WSR maximization between digital BF and an unconstrained analog BF now leads to Algorithm 5.

Algorithm 5: Hybrid BF Design via Alternating Minorizer

Given: $P^c,\boldsymbol{\Phi}_c,\mathbf{H}_{k,c},u_k,\forall k,c.$
Initialization: $(\mathbf{V}^c)^{(0)}=\mathbf{V}_{1:M^c}(\sum_{k:b_k=c}\Theta_k^c,\sum_{i:b_i\neq c}\Theta_i^c)$,
The $\mathbf{G}_k^{(0)}$ are taken as the ZF precoders for the effective channels $\mathbf{H}_{k,b_k}\mathbf{V}^{b_k}$ with uniform powers (from SPC).
Iteration (j) :

1. Compute $\widehat{\mathbf{Q}}_k^{(j)},\widehat{\mathbf{B}}_k^{(j)},\widehat{\mathbf{A}}_k^{(j)},\ \forall k$ from (2.22), (2.43), (2.44).
2. Update $\mathbf{G}_k^{'(j)},\ \forall k$, from (2.46).
3. Update $\lambda_c^{(j)},\boldsymbol{\Psi}_c^{(j)}\ \forall c$ using Algorithm 4 and thus $\mathbf{P}_k^{(j)}\ \forall k$, from (2.48).
4. Update $(\mathbf{V}^c)^{(j)},\ \forall c$, from (2.52) for fully connected case or from (2.55) for partially connected case.
5. Check for convergence of the WSR: if not go to step 1.

2.3.4 Hybrid Beamforming Design with Per-Antenna Power Constraints

Per-antenna power constraints for HBF can be written as

$$\sum_{k:b_k=c}\left[\mathbf{V}^c\mathbf{G}_k\mathbf{G}_k^H\mathbf{V}^{cH}\right]_{i,i}\leq a_i^c,\ i=1,...,N_t^c. \tag{2.56}$$

Substituting the above modified power constraints, WSR alternating maximization through minorization leads to the following expressions for the BFs (fully connected case)

$$\begin{aligned}\mathbf{G}_k^{'}&=\mathbf{V}_{1:d_k}\left(\mathbf{V}^{b_kH}\widehat{\mathbf{B}}_k\mathbf{V}^{b_k},\mathbf{V}^{b_kH}\left(\widehat{\mathbf{A}}_k+\lambda_{b_k}\mathbf{I}+\boldsymbol{\Psi}_{b_k}^{'}\right)\mathbf{V}^{b_k}\right)\\ \mathbf{V}^c&=unvec\left(\mathbf{V}_{max}\left(\mathbf{B}_c,\mathbf{A}_c^{'}\right)\right),\\ \text{where }\mathbf{A}_c^{'}&=\sum_{k:b_k=c}\left(\left(\mathbf{G}_k\mathbf{G}_k^H\right)^T\otimes\left(\widehat{\mathbf{A}}_k+\lambda_c\mathbf{I}+\boldsymbol{\Psi}_{b_k}^{'}\right)\right),\\ \text{and }\ \boldsymbol{\Psi}_c^{'}&=\text{diag}\left(\Psi_{c,1},....,\Psi_{c,N_t^c}\right).\end{aligned} \tag{2.57}$$

The ILA-WF can be modified similarly. Note that as for PRFPC, the maximum number of power constraints that can be satisfied with equality is the number of streams (stream powers).

2.3.5 Algorithm Convergence

The convergence proof of [28] does not apply directly because the power constraints here are not separable in the BF variables. The ingredients required are minorization [27], alternating or cyclic optimization [27] (also called block coordinate descent), Lagrange dual function [39], saddle-point interpretation [39] and KKT conditions [39]. For the WSR cost function $WSR(\mathbf{Q})$ in (2.42) we construct the minorizer as in (2.43), (2.44) leading to

$$WSR(\mathbf{Q}) \geq \underline{WSR}\left(\mathbf{Q},\widehat{\mathbf{Q}}\right) = \sum_{k=1}^{K}\left[u_k \ln\det\left(\mathbf{I}+\widehat{\mathbf{B}}_k\mathbf{Q}_k\right) - \mathrm{tr}\{\widehat{\mathbf{A}}_k\left(\mathbf{Q}_k-\widehat{\mathbf{Q}}_k\right)\}\right] \tag{2.58}$$

where $\underline{WSR}\left(\widehat{\mathbf{Q}},\widehat{\mathbf{Q}}\right) = WSR\left(\widehat{\mathbf{Q}}\right)$. The minorizer, which is concave in $\mathbf{Q}$, still has the same gradient as $WSR(\widehat{\mathbf{Q}})$ and hence KKT conditions are not affected. Now reparameterizing $\mathbf{Q}$ in terms of $\mathbf{P},\mathbf{G}',\mathbf{V}$ as in (2.22), and adding the power constraints to the minorizer, we get the Lagrangian (2.47). Every alternating update of $\mathscr{L}$ w.r.t. $\mathbf{V}$, $\mathbf{G}'$, or $(\mathbf{P},\mathbf{\Lambda},\mathbf{\Psi})$ leads to an increase of the WSR, ensuring convergence (within each of these 3 parameter groups, we further alternate between each user or BS). For the KKT conditions, at the convergence point, the gradients of $\mathscr{L}$ w.r.t. $\mathbf{V}$ or $\mathbf{G}'$ corresponds to the gradients of the Lagrangian of the original WSR. For fixed $\mathbf{V}$ and $\mathbf{G}'$, $\mathscr{L}$ is concave in $\mathbf{P}$, hence we have strong duality for the saddle point $\max_{\mathbf{P}}\min_{\Lambda,\mathbf{\Psi}}\mathscr{L}$. Also, at the convergence point the solution to $\min_{\Lambda,\mathbf{\Psi}}\mathscr{L}\left(\mathbf{V}^o,\mathbf{G}'^o,\mathbf{P}^o,\mathbf{\Lambda},\mathbf{\Psi}\right)$ satisfies the gradient KKT condition for $\mathbf{P}$ and the complementary slackness conditions for $c=1,\ldots,C$

$$\begin{aligned}\lambda_c^o\left(P^c - \sum_{k:b_k=c}\mathrm{tr}\{\mathbf{V}^{co}\mathbf{G}_k'^o\mathbf{P}_k^o\mathbf{G}_k'^{oH}\mathbf{V}^{coH}\}\right) = 0,\\ \mathrm{tr}\{\mathbf{\Psi}_c^o\left(\mathbf{\Phi}_c^o - \sum_{k:b_k=c}\mathbf{G}_k'^o\mathbf{P}_k^o\mathbf{G}_k'^{oH}\right)\} = 0\end{aligned} \tag{2.59}$$

where all individual factors in the products are nonnegative (and for $\mathbf{\Psi}_c^o$, the sum of nonnegative terms being zero implies all the terms being zero).

In the proposed approach, $g\left(\mathbf{\Lambda},\mathbf{\Psi}|\mathbf{V},\mathbf{G}'\right) = \max_{\mathbf{P}}\mathscr{L}\left(\mathbf{V},\mathbf{G}',\mathbf{P},\mathbf{\Lambda},\mathbf{\Psi}\right)$. In contrast, in [30], Lagrangian duality and alternating optimization are interchanged with dual function $g(\mathbf{\Lambda}) = \max_{\mathbf{V},\mathbf{G}',\mathbf{P}}\mathscr{L}\left(\mathbf{V},\mathbf{G}',\mathbf{P},\mathbf{\Lambda}\right)$ (no PRFPC or PAPC), leading to more complex iterations and a power optimization that is further away from classical water filling.

2.3.5.1 Simulation results

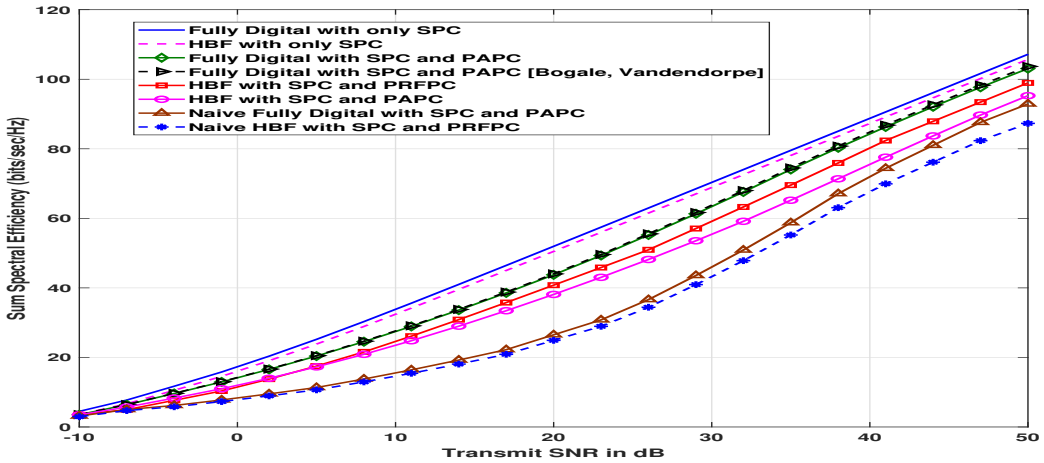


Figure 2.11: Sum rates, $N_t = 64, M = 16, K = 8, C = 1, L = 4$.

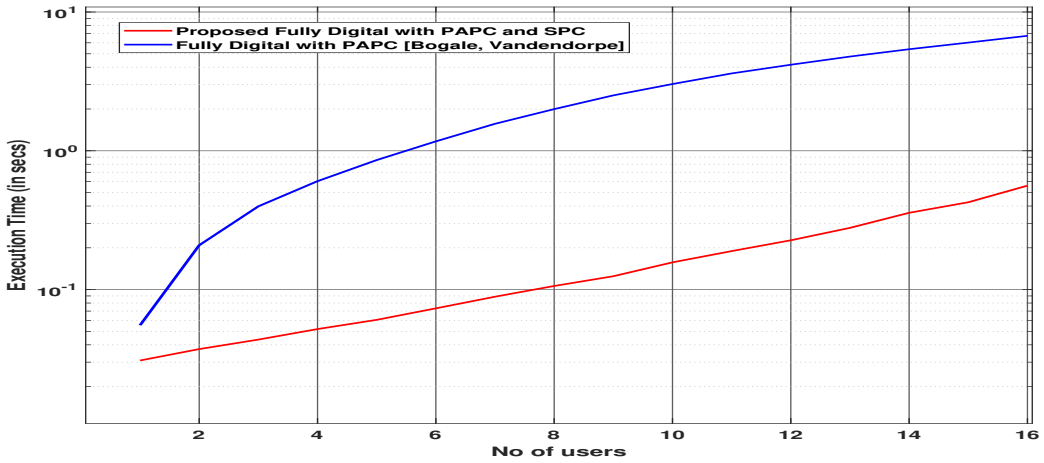


Figure 2.12: Execution time comparison.

It is clear that the proposed unconstrained HBF solution has the same performance as the fully digital solution. With phase shifter constrained analog precoder, the proposed DA based design narrows the gap to the fully digital performance and performs much better than state of the art solutions such as WSMSE which suffer from the issue of local optima. In Fig. 2.11, we compare our fully digital and HBF designs based on SPC and/or PAPC and/or PRFPC. Imposing PAPC or PRFPC in addition to the SPC degrades the sum rate but less for PRFPC as there are fewer constraints. Our digital SPC+PAPC designs performs identically to that in [35]. The optimized designs for PAPC or PRFPC outperform naive designs in which the SPC BF is scaled down to satisfy the PAPC or PRFPC constraints, esp. at intermediate SNR.

In Fig. 2.12, for the fully digital PAPC, we compare the execution time in Matlab for the proposed solution to that of the geometric programming (GP) approach in [35] for the power allocation (which is solved using interior point methods). The digital BF computation has similar complexity ($\mathcal{O}(N_t^3)$) between SMSE in [35] and the proposed solution. The complexity $(N_t + 1)x$ of the alternating bisection is linear in the number of power constraints, where x represents the

complexity associated with the evaluation of $g(\mathbf{\Lambda}, \mathbf{\Psi})$). GP has a worst case polynomial time complexity. Faster convergence of the minorization approach compared to the SMSE solution and the reduced complexity of the alternating bisection vs the GP lead to a much shorter execution time for the proposed algorithm as shown in Fig. 2.12.

2.3.6 Conclusion

We presented a WSR maximizing algorithm for HBF, with unconstrained amplitude or phasor analog BF, fully or partially connected, in a Multi-User Multi-Cell MIMO system. First we considered a mixed time scale approach for HBF with analog BF varied according to the slow fading and digital BF being changed w.r.t the fast fading or instantaneous CSIT. Later we proved using theorem 1 that, to reach that of fully digital performance (WSMSE based), analog BF can be chosen as the concatenated antenna array reponse matrix (slow fading components) if the number of RF chains are greater than the total number of user paths. We considered for the first time the more realistic scenario of per-RF or per-antenna power constraints for a HBF system. Convergence of the alternating minorization approach was shown and adding deterministic annealing allowed to attain the global optimum.

2.4 Hybrid Beamforming Design for Multi-User MIMO-OFDM Systems

We consider a multi-cell MU downlink (i.e. Interfering BroadCast Channel (IBC)) OFDM system of C cells with a total of K users. We constrain the total transmit power to be P_c at BS c and N_t^c transmit antennas in cell c. N_s represents the total number of subcarriers which is shared across all the users. User k is equipped with N_k antennas. The number of streams intended for user k is d_k. Let $\mathbf{H}_{k,c}[n]$ represents the $N_k \times N_t^c$ MIMO downlink channel between user k and BS c and we define the channel covariance to be $\mathrm{E}\left(\mathbf{H}_{k,c}^H[n]\mathbf{H}_{k,c}[n]\right) = \Theta_k^c[n]$. n represents the subcarrier index throughout the chapter. It is important to emphasize here that the analog precoder is assumed to be frequency flat (same for all subcarriers) and digital precoder to be frequency selective. Note that we consider the Rx to be a fully digital system since N_k is not very high at the UE. User k receives

$$\begin{aligned}\mathbf{y}_k[n] = \mathbf{H}_{k,b_k}[n]\mathbf{V}^{b_k}\mathbf{G}_k[n]\,\mathbf{s}_k[n] + \\ \sum_{i\neq k}\mathbf{H}_{k,b_i}[n]\mathbf{V}^{b_i}\mathbf{G}_i[n]\,\mathbf{s}_i[n] + \mathbf{v}_k[n],\end{aligned} \tag{2.60}$$

where $\mathbf{s}_k[n]$, of size $d_k \times 1$, is the transmit symbol vector with $\mathbf{s}_k[n] \sim \mathscr{CN}(\mathbf{0}, \mathbf{I})$. b_i refers to the serving base station of user i. BS c serves U_c users and $K = \sum_{c=1}^{C} U_c$. Assuming that $\mathbf{y}_k[n]$ represent a noise whitened signal model, we get for the noise $\mathbf{v}_k \sim \mathscr{CN}(\mathbf{0}, \mathbf{I}_{N_k})$ (circularly complex Gaussian random vector). The analog BF which is same across all the subcarriers, $\mathbf{V}^c$ for base station c is of dimension $N_t^c \times M^c$ where M^c is the number of RF chains at BS c. The $M^c \times d_k$ digital beamformer is $\mathbf{G}_k[n]$, where $\mathbf{G}_k[n] = \left[\mathbf{g}_k^{(1)}[n] \ldots \mathbf{g}_k^{(d_k)}[n]\right]$ and $\mathbf{g}_k^{(s)}[n]$ represents the beamformer for stream s of user k.

2.4.1 MIMO OFDM Channel Model

In this sub-section, we omit the user and cell indices for simplicity. We consider a geometric channel model for a mmWave propagation environment [41] with L_s scattering clusters and L_r

scatterers or rays in each cluster. In a more compact form, we can represent the channel matrix at a subcarrier n as,

$$\mathbf{H}[n] = \mathbf{H}_r \sum_{d=1}^{D} \mathbf{A}_d[n]\mathbf{H}_t^H, \text{ where} \tag{2.61}$$

where

$$\begin{aligned} \mathbf{H}_r &= \left[\mathbf{h}_r(\theta_{1,1}), ..., \mathbf{h}_r(\theta_{L_s,L_r})\right], \\ \mathbf{H}_t &= \left[\mathbf{h}_t(\phi_{1,1}), ..., \mathbf{h}_t(\phi_{L_s,L_r})\right], \\ \mathbf{A}_d[n] &= \text{diag}\left(\alpha_{1,1}p(dTs - \tau_1 - \tau_{r1}), ..., \alpha_{L_s,L_r}p(dTs - \tau_{L_s} - \tau_{rL_r})\right) e^{-j2\pi\frac{nd}{N_s}}. \end{aligned} \tag{2.62}$$

Here $\phi_{s,l}, \theta_{s,l}$ represent the angle of departure (AoD) and angle of arrival (AoA), respectively for the l^{th} path in the s^{th} cluster. $\mathbf{h}_r(\cdot), \mathbf{h}_t(\cdot)$ represent the antenna array responses at Rx and Tx respectively. The complex path gain which is an indicator of the channel power in each path is modeled as, $\alpha_{s,l} \sim \mathscr{CN}(0, \frac{N_t N_r}{L_s L_r})$ and $p(\tau)$ is the band-limited pulse shaping filter response evaluated at τ seconds. Each cluster has a time delay $\tau_s \in \mathscr{R}$ and each ray has a relative time delay τ_{rl}. Note that our HBF design which follows, is applicable for general MIMO channel models and the channel model outlined here is utilized for the simulations in Section 2.4.7. Another remark here is that, even though for an HBF system, at the baseband we have access to only the low-dimensional effective channel resulting from the combination of the propagation channel and the analog precoder, it is still possible to estimate the individual components in a pathwise channel model as we consider here, for example [22, 42].

2.4.2 WSR Maximization via Minorization and Alternating Optimization

For the convenience of analysis, we define the Tx covariance matrix as $\mathbf{Q}_i[n] = \mathbf{V}^{b_i}\mathbf{G}_i[n]\mathbf{G}_i[n]^H\mathbf{V}^{b_i H}$. HBF design using WSR maximization of the multi-cell MU-MIMO OFDM system can be formulated as follows,

$$\begin{aligned} [\mathbf{V}\ \mathbf{G}] &= \arg\max_{\mathbf{V},\mathbf{G}} WSR(\mathbf{G}, \mathbf{V}) \\ &= \arg\max_{\mathbf{V},\mathbf{G}} \sum_{k=1}^{K} u_k \sum_{n=1}^{N_s} \ln\det\left(\mathbf{R}_{\bar{k}}[n]^{-1}\mathbf{R}_k[n]\right), \\ \text{s.t.} &\sum_{k:b_k=c} \sum_{n=1}^{N_s} \text{tr}\{\mathbf{Q}_k[n]\} \le P_c. \end{aligned} \tag{2.63}$$

where the u_k being the weight for user k (can represent the priority), $\mathbf{G}$ represents the collection of digital BFs $\mathbf{G}_k[n]$ and $\mathbf{V}$ the collection of analog BFs $\mathbf{V}^{b_k}$. From [9, 28], we can write,

$$\begin{aligned} \mathbf{R}_{\bar{k}}[n] &= \sum_{i=1, i\neq k}^{K} \mathbf{H}_{k,b_i}[n]\mathbf{Q}_i[n]\mathbf{H}_{k,b_i}^H[n] + \mathbf{I}_{N_k}, \\ \mathbf{R}_k[n] &= \sum_{i=1}^{K} \mathbf{H}_{k,b_i}[n]\mathbf{Q}_i[n]\mathbf{H}_{k,b_i}^H[n] + \mathbf{I}_{N_k}, \end{aligned} \tag{2.64}$$

where $\mathbf{R}_{\bar{k}}[n]$ is the interference plus noise covariance matrix. Further, we utilize the alternating minorization concept outlined in Section 2.2. The derivation follows the same steps as before (optimization across the subcarriers can be decoupled) and hence due to repetition can be

skipped here. We define the following auxiliary variables which appear in the BF expressions.

$$\begin{aligned}\widehat{\mathbf{A}}_k[n] &= \sum_{i=1,\neq k}^{K} u_i \mathbf{H}_{i,b_k}^H[n]\left(\widehat{\mathbf{R}}_{\bar{i}}[n]^{-1} - \widehat{\mathbf{R}}_i[n]^{-1}\right)\mathbf{H}_{i,b_k}[n]\,.\\ \widehat{\mathbf{B}}_k[n] &= \mathbf{H}_{k,b_k}^H[n]\widehat{\mathbf{R}}_{\bar{k}}^{-1}[n]\mathbf{H}_{k,b_k}[n].\end{aligned} \tag{2.65}$$

Also, the resulting Lagrangian from the alternating minorization can be written as

$$\begin{aligned}\mathscr{L}(\mathbf{G},\mathbf{V},\mathbf{\Lambda}) = \sum_{k=1}^{K}\sum_{n=1}^{N_s}\Big[u_k \ln\det\Big(\mathbf{I} + \mathbf{G}_k^H[n]\mathbf{V}^{b_k H}\widehat{\mathbf{B}}_k[n]\mathbf{V}^{b_k}\mathbf{G}_k[n]\Big)\\ - \mathrm{tr}\{\mathbf{G}_k^H[n]\mathbf{V}^{b_k H}\left(\widehat{\mathbf{A}}_k[n] + \lambda_{b_k}\mathbf{I}\right)\mathbf{V}^{b_k}\mathbf{G}_k[n]\}\Big] + \sum_{j=1}^{C}\lambda_j P_j,\end{aligned} \tag{2.66}$$

2.4.3 Digital BF Design

By Hadamard's inequality [43, p. 233], it can be seen that for the maximization problem above, $\mathbf{G}_k^H[n]\mathbf{V}^{b_k H}\widehat{\mathbf{B}}_k[n]\mathbf{V}^{b_k}\mathbf{G}_k[n]$ should be diagonal and thus maximizing w.r.t $\mathbf{G}_k[n]$ leads to the following dominant generalized eigenvector solution. Also, note that the gradient w.r.t. $\mathbf{G}_k[n]$ of (2.66) is still the same as that of (2.63).

$$\mathbf{G}_k^{'}[n] = \mathbf{V}_{1:d_k}\left(\mathbf{V}^{b_k H}\widehat{\mathbf{B}}_k[n]\mathbf{V}^{b_k}, \mathbf{V}^{b_k H}\left(\widehat{\mathbf{A}}_k[n] + \lambda_{b_k}\mathbf{I}\right)\mathbf{V}^{b_k}\right), \tag{2.67}$$

with associated generalized eigenvalues $\mathbf{\Sigma}_k[n] = \mathbf{\Sigma}_{1:d_k}\left(\mathbf{V}^{b_k H}\widehat{\mathbf{B}}_k[n]\mathbf{V}^{b_k}, \mathbf{V}^{b_k H}\left(\widehat{\mathbf{A}}_k[n] + \lambda_{b_k}\mathbf{I}\right)\mathbf{V}^{b_k}\right)$. λ_{b_k} represents the Lagrange multiplier associated with the power constraint at BS b_k. Let $\mathbf{\Sigma}_k^{(1)}[n] = \mathbf{G}_k^{'H}[n]\mathbf{V}^{b_k H}\widehat{\mathbf{B}}_k[n]\mathbf{V}^{b_k}\mathbf{G}_k^{'}[n]$ and $\mathbf{\Sigma}_k^{(2)}[n] = \mathbf{G}_k^{'H}[n]\mathbf{V}^{b_k H}\widehat{\mathbf{A}}_k[n]\mathbf{V}^{b_k}\mathbf{G}_k^{'}[n]$. Intuitively, (2.67) represents a compromise between increasing the signal part and reducing the interference. Now we introduce stream powers in the diagonal matrices $\mathbf{P}_k[n] \geq 0$. The Lagrangian formulation (2.66) allows us to optimize the stream powers. Further substituting $\mathbf{G}_k[n] = \mathbf{G}_k^{'}[n]\,\mathbf{P}_k^{\frac{1}{2}}[n]$ in (2.66) yields the following interference leakage aware water filling (WF) (jointly for the $\mathbf{P}_k[n]$ and λ_c)

$$\mathbf{P}_k[n] = \left(u_k\left(\mathbf{\Sigma}_k^{(2)}[n] + \lambda_{b_k}\mathbf{V}^{b_k H}\mathbf{V}^{b_k}\right)^{-1} - \mathbf{\Sigma}_k^{-(1)}[n]\right)^{+}, \tag{2.68}$$

where $(\mathbf{X})^+$ denotes the positive semi-definite part of Hermitian $\mathbf{X}$ (so by removing the terms with negative eigenvalues to zero) and the Lagrange multipliers (per BS) are computed using bisection to satisfy the power constraints.

2.4.4 Design of Unconstrained Analog BF

At first, we investigate the case in which the analog BF is unconstrained. One remark here is that the resulting HBF design is also applicable to general two-stage BF design [40], where the higher dimensional outer BF stage ($\mathbf{V}^c$) is common to all users in a cell. To optimize $\mathbf{V}^c$, we equate the gradient of (2.66) w.r.t. $\mathbf{V}^c$ to zero. Using the result $\partial \ln\det\mathbf{X} = \mathrm{tr}(\mathbf{X}^{-1}\partial\mathbf{X})$ and $\det(\mathbf{I}_M + \mathbf{AB}) = \det(\mathbf{I}_N + \mathbf{BA})$ from [31], we get

$$\begin{aligned}\sum_{k:b_k=c}\sum_{n=1}^{N_s}\left(\widehat{\mathbf{B}}_k[n]\mathbf{V}^c\mathbf{G}_k[n]\zeta_k[n]\mathbf{G}_k^H[n] - \left(\widehat{\mathbf{A}}_k[n] + \lambda_c\mathbf{I}\right)\mathbf{V}^c\mathbf{G}_k[n]\mathbf{G}_k^H[n]\right) = 0,\\ \text{with } \zeta_k[n] = u_k\left(\mathbf{I} + \mathbf{G}_k^H[n]\mathbf{V}^{b_k H}\widehat{\mathbf{B}}_k[n]\mathbf{V}^{b_k}\mathbf{G}_k[n]\right)^{-1}.\end{aligned} \tag{2.69}$$

Now with $vec(\mathbf{AXB}) = \left(\mathbf{B}^T \otimes \mathbf{A}\right) vec(\mathbf{X})$ [31], where $\otimes$ represents the Kronecker product between the two matrices, we get

$$\begin{aligned} \mathbf{V}^c &= unvec\left(\mathbf{V}_{max}\left(\mathbf{B}_c[n], \mathbf{A}_c[n]\right)\right), \text{ with} \\ \mathbf{B}_c[n] &= \sum_{k:b_k=c} \sum_{n=1}^{N_s} \left(\mathbf{G}_k[n]\zeta_k[n]\mathbf{G}_k^H[n]\right)^T \otimes \widehat{\mathbf{B}}_k[n], \\ \mathbf{A}_c[n] &= \sum_{k:b_k=c} \sum_{n=1}^{N_s} \left(\mathbf{G}_k[n]\mathbf{G}_k^H[n]\right)^T \otimes \left(\widehat{\mathbf{A}}_k[n] + \lambda_c \mathbf{I}\right). \end{aligned} \tag{2.70}$$

We emphasize here that the extension to the partially connected HBF architecture is quite straightforward and we include the comparison of both in Section 2.4.7.

2.4.5 Algorithm Convergence

The convergence proof follows in the same direction as in [44]. For the WSR cost function for a wideband system, we construct the minorizer as in (2.65), (2.66) leading to

$$WSR(\mathbf{Q}) \geq \underline{WSR}(\mathbf{Q}, \widehat{\mathbf{Q}}) \quad = \sum_{k=1}^{K} \sum_{n=1}^{N_s} \left[u_k \ln\det\left(\mathbf{I} + \widehat{\mathbf{B}}_k[n]\mathbf{Q}_k[n]\right) - \mathrm{tr}\{\widehat{\mathbf{A}}_k[n]\left(\mathbf{Q}_k[n] - \widehat{\mathbf{Q}}_k[n]\right)\}\right], \tag{2.71}$$

where $\underline{WSR}\left(\widehat{\mathbf{Q}}, \widehat{\mathbf{Q}}\right) = WSR\left(\widehat{\mathbf{Q}}\right)$. The resulting minorizer above is a concave function in $\widehat{\mathbf{Q}}$ and has the same gradient as $WSR(\widehat{\mathbf{Q}})$. Hence the KKT conditions are unaffected. Now reparameterizing $\mathbf{Q}$ in terms of $\mathbf{P}, \mathbf{G}', \mathbf{V}$ as in (2.64) and adding the power constraints to the minorizer, we get the Lagrangian (2.66). Every alternating update of $\mathscr{L}$ w.r.t. $\mathbf{V}$, $\mathbf{G}'$, or $(\mathbf{P}, \mathbf{\Lambda})$ increases the WSR since the approximate problem is a concave function, which ensures convergence within each of these 3 parameter groups and we further alternate between each user or BS. Also, at the convergence point, the gradients of $\mathscr{L}$ w.r.t. $\mathbf{V}$ or $\mathbf{G}'$ corresponds to the gradients of the Lagrangian of the original WSR and hence the KKT conditions remain unaffected. For fixed $\mathbf{V}$ and $\mathbf{G}'$, $\mathscr{L}$ is concave in $\mathbf{P}$, hence strong duality is satisfied for the saddle point $\max_{\mathbf{P}} \min_{\mathbf{\Lambda}} \mathscr{L}$. Also, at the convergence point, the solution to $\min_{\mathbf{\Lambda}} \mathscr{L}(\mathbf{V}^o, \mathbf{G}'^o, \mathbf{P}^o, \mathbf{\Lambda})$ satisfies the gradient KKT condition for $\mathbf{P}$ and the complementary slackness conditions for $c = 1, \ldots, C$

$$\lambda_c^o \left(P^c - \sum_{k:b_k=c} \sum_{n=1}^{N_s} \mathrm{tr}\{\mathbf{V}^{co}\mathbf{G}_k'^o[n]\mathbf{P}_k^o[n]\mathbf{G}_k'^{oH}[n]\mathbf{V}^{coH}\} = 0, \right. \tag{2.72}$$

where all individual factors in the products are nonnegative. In the proposed approach, $g\left(\mathbf{\Lambda}|\mathbf{V}, \mathbf{G}'\right) = \max_{\mathbf{P}} \mathscr{L}\left(\mathbf{V}, \mathbf{G}', \mathbf{P}, \mathbf{\Lambda}\right)$.

2.4.6 Analysis on the number of RF Chains and HBF Performance

In this section, we derive an analytical solution for the analog phasors to achieve a fully digital BF performance. In short, we prove that it is possible to achieve using a sufficient number of RF chains under certain conditions on the MaMIMO channel being considered. For notational simplicity, we shall consider a uniform $L = L_s L_r$ and $N_k = N_r, \forall k, N_t^c = N_t, M^c = M, \forall c$. Let us represent the concatenated antenna array response matrix of all user channel from BS c as, $\overline{\mathbf{H}}_t^c = \left[\mathbf{H}_{t,1}^c \; \mathbf{H}_{t,2}^c \ldots \mathbf{H}_{t,K}^c\right]$, of dimension $N_t \times N_p$, where we denote the total number of paths $N_p = LK$. We define $\mathbf{A}_{d,k}^c[n]$ as the diagonal path amplitude matrix for the channel from BS c to user k for subcarrier n. Similarly, we define $\overline{\mathbf{H}}_r^c$ and $\overline{\mathbf{A}}^c[n] = \mathrm{diag}\left(\sum_{d=1}^{D} \mathbf{A}_{d,1}^c[n], \ldots, \sum_{d=1}^{D} \mathbf{A}_{d,K}^c[n]\right)$ of size

$N_p \times N_p$ for the concatenated Rx antenna array responses and complex path amplitudes. $\overline{\mathbf{H}}_r^c$ is a $KN_r \times N_p$ block diagonal matrix with blocks of size $N_r \times L$. Finally, we can write the $KN_r \times N_t$ MIMO channel from BS c to all a users as $\mathbf{H}^{cH}[n] = \overline{\mathbf{H}}_t^c \overline{\mathbf{A}}^{cH}[n] \overline{\mathbf{H}}_r^{cH}$.

Theorem 2. *Consider a multi-cell MU MIMO OFDM system with the number of RF chains being less than the total number of paths across all user channels from any BS and assume phasor antenna responses. In order to achieve optimal all-digital precoding performance, an analytical solution for the analog beamformer can be obtained as the Tx side concatenated antenna array response and thus frequency flat assuming no beam squint effect.*

Proof: From [9], the optimal all-digital beamformer for any subcarrier n is of the form

$$\begin{aligned} \left(\mathbf{H}^{cH}[n]\mathbf{D}_1^c[n]\mathbf{H}^c[n] + \lambda_c \mathbf{I}\right)^{-1} \mathbf{H}^{cH}[n]\mathbf{D}_2^c[n] & \\ = \mathbf{H}^{cH}\mathbf{B}^c & \\ = \overline{\mathbf{H}}_t^c\, \overline{\mathbf{A}}^{cH}[n]\, \overline{\mathbf{H}}_r^{cH} \mathbf{B}^c[n], & \end{aligned} \tag{2.73}$$

where $\mathbf{B}^c[n] = \left(\lambda_c \mathbf{I} + \mathbf{D}_1^c[n]\mathbf{H}^c[n]\mathbf{H}^{cH}[n]\right)^{-1} \mathbf{D}_2^c[n]$, $\mathbf{D}_1^c[n]$, $\mathbf{D}_2^c[n]$ are block diagonal matrices and we used the identity $(\mathbf{I}+\mathbf{XY})^{-1}\mathbf{X} = \mathbf{X}(\mathbf{I}+\mathbf{YX})^{-1}$. Under the Theorem assumptions we can then separate the BFs as

$$\mathbf{V}^c = \overline{\mathbf{H}}_t^c,\ \mathbf{G}^c[n] \quad = \overline{\mathbf{A}}^{cH}[n] \overline{\mathbf{H}}_r^{cH} \mathbf{B}^c[n]\,. \tag{2.74}$$

Hence $\mathbf{V}$ is a function of only the frequency flat antenna array responses which are slow fading components. So it is independent of the subcarrier number and this explains why it is optimal to consider a frequency flat design for analog BF. However, note that the digital BF $\mathbf{G}$ in (2.67) is a function of the instantaneous CSIT and needs to be updated every channel use in the time and frequency domain. Also, while the spatial angles in antenna array responses may include a frequency dependency called beam-squint in the literature [45], we do not consider this factor at the moment.

For the case when $M < N_p$, we utilize the DA based approach proposed earlier in our work [46]. We refer the reader for a more detailed discussion on this to our paper. In the below table Algorithm 6, we describe in detail the HBF algorithm which combines minorization and DA.

2.4.7 Simulation Results

In this section, we validate the performance of the proposed HBF algorithms for a single cell and multi-cell system (Figure 2.14) with K single antenna users and for an OFDM system with $N_s = 32$ subcarriers using extensive Monte-Carlo simulations. We use the pathwise channel model in (2.61). We consider a Uniform Linear Array (ULA) of antennas with $\mathbf{h}_{t,k}\left(\phi_{c,l}\right)$, the AoD $\phi_{c,l}$ are restricted to the interval $[0^o, 30^o]$ and uniformly distributed. For the multi-cell case in Figure 2.14, the parameters used are the same for both the cells, i.e. $M^1 = M^2 = M, N_t^1 = N_t^2 = N_t, U_1 = U_2 = K/2, L_s = 1, L_r = 4, D = L_s L_r$. Our system dimensions are such that the number of RF chains satisfies the condition, $M < LK$ such that alternating optimization of phasors results in local optima issues. For simplicity, a rectangular pulse shape is used. Notations used in the figure: "ABF" refers to the analog BF. "EV phasors" refers to the HBF design with analog phasors being chosen as the projection of eigenvectors of the sum of the user channel covariance matrices onto the unit modulus constraints. We compare the performance of the proposed algorithms with the WSMSE based fully digital BF [9] (referred to as "WSMSE Fully Digital"), approximate WSR based

Algorithm 6: Minorization and DA based HBF design

Given: $P_c, \mathbf{H}_{k,c}[n], u_k \, \forall k, c, n$, b is a constant < 1, say 0.9.
Initialization: $\mathbf{V}^c = e^{j\angle \mathbf{V}_{1:M^c}(\sum_{k:b_k=c}\Theta_k^c[n], \sum_{i:b_i\neq c}\Theta_i^c[n])}$,
The $\mathbf{G}_k^{(0)}[n]$ are initialized to be ZF precoders for the effective channels $\mathbf{H}_{k,b_k}[n]\mathbf{V}^{b_k}$, with uniform power distribution across the streams. **Iteration** (j) :

1. Compute $\widehat{\mathbf{B}}_k[n], \widehat{\mathbf{A}}_k[n], \, \forall k, n$ from (2.65), (2.66).
2. Update $\mathbf{G}_k^{'(j)}[n]$ from (2.67), and $\mathbf{P}_k[n]$ from (2.68), $\forall k, n$.
3. Update $(\mathbf{V}_{p,q}^c)^{(j)}$, $\forall c, \forall (p,q)$, using DA (phasor constrained) or from (2.70) (unconstrained).
4. If the algorithm is converged, exit the loop, otherwise go to step 1.
5. Scale $\forall \, (i,j) : |\mathbf{V}_{i,j}^c| \leftarrow e^{b \ln |\mathbf{V}_{i,j}^c|}$ $(\mathbf{V}_{i,j}^c = |\mathbf{V}_{i,j}^c| e^{j\theta_{i,j}^c})$.
6. Reoptimize all $\theta_{i,j}^c$ and all digital BFs using steps 1-4.
7. Update stream powers and Lagrange multipliers.
8. Go to step 5 for a number of iterations.
9. Finally redo steps 6-7 a last time with all $|\mathbf{V}_{i,j}^c| = 1$ in step 5.

hybrid design [47] (referred to as “HBF with ABF based on Channel Average". For the multi-cell version of [47], channel average with only the direct user channels in a cell are considered. “HBF with Alternating Optimization of Phasors" refers to our algorithm in the paper [10], but extended to an OFDM system.

It is evident from Figure 2.13 and Figure 2.14 that our proposed unconstrained HBF has almost the same performance as that of the WSMSE based fully digital BF. With phase shifter constrained analog precoder, the proposed DA based design narrows the gap to the fully digital performance and performs much better than state of the art solutions such as WSMSE which suffer from the issue of local optima for analog phasors. Also, the performance degrades for a partially connected architecture compared to the fully connected system. However, it is to be noted that the complexity of the proposed HBF design is slightly on the higher side and it is $O(N_t^3 N_{it})$, where N_{it} represents the number of iterations required for Algorithm 1 to converge.

In Figure 2.15, we demonstrate the sum SE per subcarrier for 256 subcarriers and with the delay tap of the frequency selective channel increased to $D = 12$. Note that we see a slight performance degradation for the HBF system w.r.t the fully digital case when the frequency selectivity is increased. Hence, it is of interest to relook at the HBF design for wideband OFDM system which can close the gap to the fully digital case.

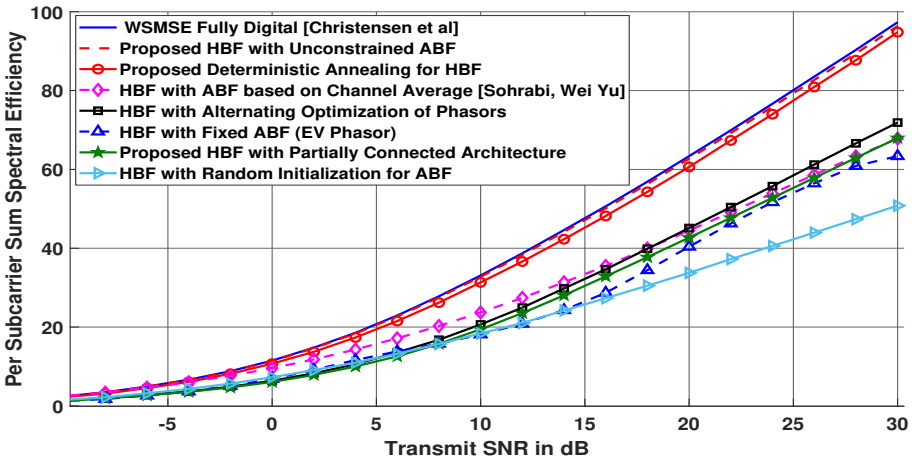


Figure 2.13: Sum rate, $N_t = 32, M = 16, K = 16, C = 1, L = 4, N_s = 32$.

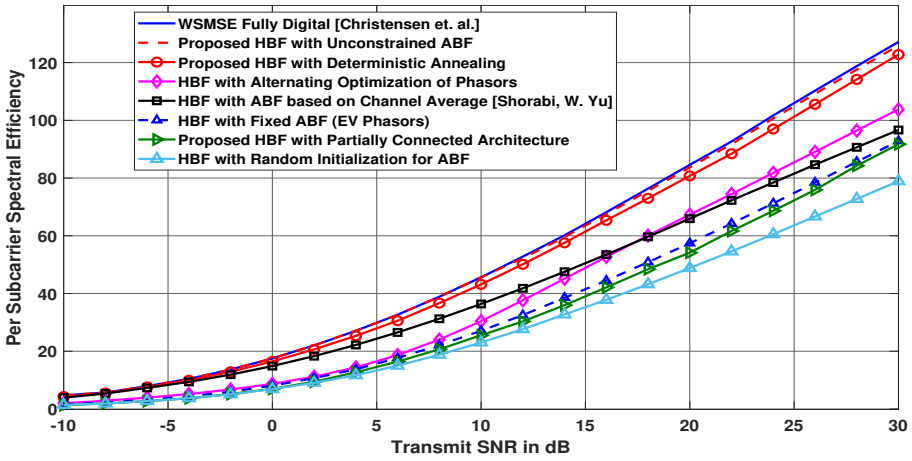


Figure 2.14: Sum rate, $N_t = 64, M = 16, K = 16, C = 2, L = 4, N_s = 32$.

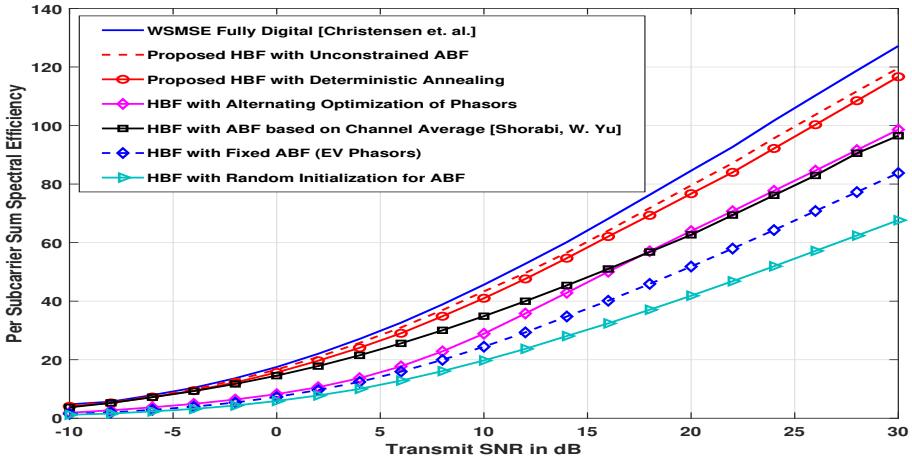


Figure 2.15: Sum rate, $N_t = 64, M = 16, K = 16, C = 2, L = 12, N_s = 256$.

2.4.8 Conclusions and Perspectives

Conclusions and Perspectives 1

- In this chapter, we derived and presented an optimal BF algorithm for the HBF scenario in a multi-cell MU-MIMO single carrier and OFDM system.
- We optimized the WSR objective function using a difference of convex functions approach (which is also an instance of minorization) and the BF solutions are alternatively computed till convergence.
- We noted that alternating optimization of analog phasors using WSMSE or WSR objective function leads to bad local optima, convergence depends on the initialization used. Hence, to arrive at a globally optimal phasor design, we proposed an innovative solution based on the concept of deterministic annealing.
- Convergence to a local optimum is shown and through extensive simulations, we show that our DA based approach for analog BF design performs far better than the existing state of the art solutions based on WSMSE or other suboptimal objective functions.
- We would like to remind here that the complexity of our proposed alternating minorization algorithm is very high and may not be much appreciated as a practical solution. Nevertheless, our solution can act as a performance benchmark for other suboptimal solutions in the literature. Moreover, it would be advisable to look at a low complexity solution and under imperfect channel knowledge scenario, which is left as future work.

Chapter 3

HYBRID BEAMFORMING FOR FULL-DUPLEX SYSTEMS

3.1 Introduction

In-band full-duplex (FD) wireless, which allows each node to transmit and receive simultaneously has the potential to double the spectral efficiency and is one of the prominent candidates for 5G. It avoids the use of two independent channels for bi-directional communication, by allowing more flexibility in spectrum utilization, improving data security, and reduces the air interface latency and delay issues. Unfortunately, it suffers from severe self-interference (SI) which could be 110 dB higher than the Rx signal power, and canceling it is not a trivial task due to nonlinearities and imperfections in the Tx chains, as identified in [48].

However, advancement in cancellation techniques has made FD operation possible. A combination of analog, digital, and passive SIC techniques is required to reduce SI near the noise floor, by allowing signal reception with a high signal-to-self-interference-plus-noise ratio. The first design and implementation of FD WiFi radio were introduced in [49]. In [50], SIC in FD is investigated experimentally and a practical FD system is proposed. In [51], the authors combine analog and digital SIC techniques and study the effect of residual SI together with clipping plus-quantization noise due to the limited dynamic range (LDR) of ADCs is studied. The analog cancellation stage is fundamental to reduce the SI sufficiently to ensure that it does not saturate the ADCs in the RX chains. Its complexity remains a serious challenge for upcoming massive MIMO FD scenarios, as it scales very poorly with the number of antennas. As discussed in [52], the next-generation base stations (BS) will deploy 64-256 antenna elements. Therefore, the analog cancellation stage may become infeasible for upcoming communication scenarios, due to the large complexity associated. Also, the cost of hardware components required to mimic the SI signal may become unattractive.

The use of separate Tx/Rx antenna arrays combined with various spatial precoding techniques has also been proposed to mitigate SI. In [53], two sequential convex programming (SCP) based algorithms for the joint optimization of beamforming (BF) and SIC are proposed. Recent studies on fully digital BF schemes under LDR using weighted sum rate (WSR) criteria for FD systems can be found in [54, 55].

3.1.1 Summary of the Chapter

- We propose a two-stage BF design for a bidirectional FD MIMO OFDM system based on the WSR criterion which is solved using the alternating minorization approach the main

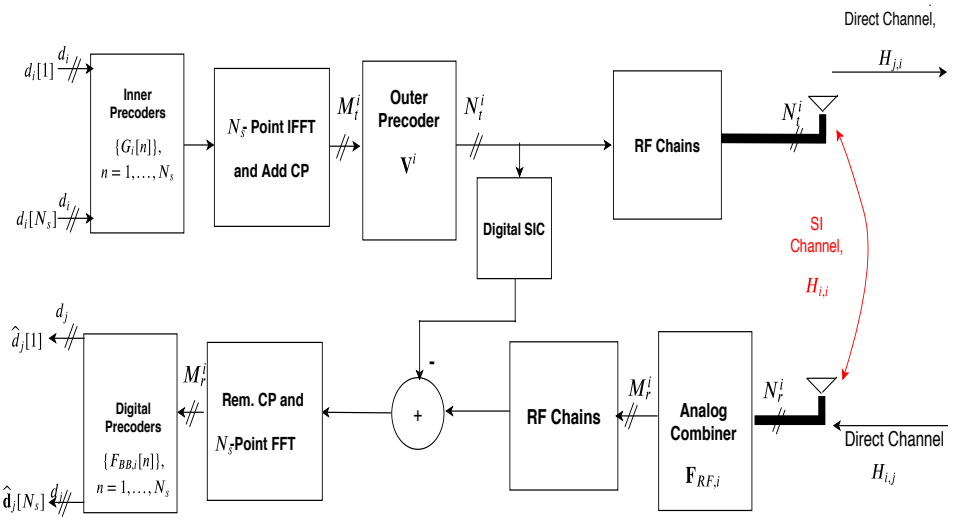


Figure 3.1: Bidirectional FD MIMO OFDM System with Multi-Stage/Hybrid BF. Only a single node is shown for simplicity in the figure.

advantage of which compared to the weighted sum mean square error (WSMSE) approach is its faster convergence. The minorization approach also involves user stream power optimization which also implicitly selects the number of supportable streams for a user.

- At the Tx side, we propose to use a two-stage BF at the baseband where the higher dimensional precoder is applied to the time domain signal which aims to mitigate the SI and the lower dimensional precoder in the OFDM domain provides spatial multiplexing gain. At the Rx side, we introduce an HBF design. The objective of the time domain phase shifter analog BF stage is to suppress the SI before the ADC while preserving the dimension of the desired signal space.

- Compared to the only existing state of the art design on HBF for FD systems [56], we consider a more realistic LDR noise model at both the Tx and Rx. Our previous work [57] on multi-stage BF design for FD systems uses weighted sum mean square error (WSMSE) based method to design the BFs at Tx/Rx. However, we observe that alternating minorization based approach as we consider here leads to much faster convergence than WSMSE based methods. Moreover, our proposed approach also is readily extendable to a partial CSIT case and leads to much efficient design than WSMSE under imperfect channel knowledge. Also, we would like to remark that our previous works on HBF [44] are for half-duplex systems and does not take into account the more practical noise model as LDR which is considered herein.

- Through Monte Carlo simulations, we validate the performance of our proposed multi-stage/HBF design. Simulations demonstrate that using an analog combiner stage at Rx (which operates before the Rx side LDR noise) has better sum rate performance compared to using a two-stage BF at Tx side for SI nulling.

3.2 Full-Duplex Bidirectional MIMO System Model

In this chapter, we shall consider a multi-stream approach with d_j streams intended for the base station (BS) j. Two BSs are represented by the indices i and j respectively. So, consider a single user bidirectional FD backhaul system as depicted in Figure 3.1, with N_t^i or N_t^j Tx antennas at the BS i or j, respectively. We may also use index 1 or 2 instead of i or j in the chapter. Furthermore, we consider an OFDM system with N_s subcarriers. BSs are equipped with N_r^1 or N_r^2 receive antennas. $\mathbf{H}_{i,j}$, $i \neq j$ represents the $N_r^i \times N_t^j$ MIMO direct channel between node i and node j. Let $\mathbf{H}_{i,i}$ represent the SI channel from the Tx of node i to the Rx of node i. User i receives

$$\begin{aligned}\mathbf{y}_i[n] &= \mathbf{F}_{RF,i}\mathbf{H}_{i,j}[n](\mathbf{V}^j\,\mathbf{G}_j[n]\mathbf{d}_j[n] + \mathbf{c}_j[n]) \\ &\quad + \mathbf{F}_{RF,i}\mathbf{H}_{i,i}[n](\mathbf{V}^i\,\mathbf{G}_i[n]\mathbf{d}_i[n] + \mathbf{c}_i[n]) + \mathbf{e}_i[n] + \mathbf{F}_{RF,i}\mathbf{n}_i[n],\end{aligned} \tag{3.1}$$

where $\mathbf{d}_j[n]$, of size $d_j \times 1$, is the intended signal stream vector (all entries are white, unit variance) to node i. At the Tx side, we have a two-stage beamformer (inner BF, $\mathbf{G}_j$ of lower dimension, and an outer BF, $\mathbf{V}^j$ of higher dimension), both the beamformers being at the digital (baseband) side. The outer BF will be applied to the time domain signal at the Tx side, so after the IFFT and it will be common to all the subcarriers. The inner BF will be different for different subcarriers. We are considering a noise whitened signal representation so that we get for the noise $\mathbf{n}_i \sim \mathscr{CN}(0, \mathbf{I}_{N_r^i})$. The higher dimensional outer precoder $\mathbf{V}^j$ at Tx of node j is of dimension $N_t^j \times M_t^j$. The digital beamformer is $\mathbf{G}_j$ which has dimensions $M_t^j \times d_j$, where $\mathbf{G}_j = \left[\mathbf{g}_j^{(1)} \ldots \mathbf{g}_j^{(d_j)}\right]$ and $\mathbf{g}_j^{(s)}$ represents the beamformer for stream s. $\mathbf{c}_i, \mathbf{e}_i$ represents the noise at the Tx or Rx antennas of node i respectively, which models the effect of LDR. LDR noise at Tx or Rx closely approximates the effects of non-ideal amplifiers, oscillators, and ADCs/DACs. The covariance matrix of $\mathbf{c}_i$ is given by $\alpha_i (\alpha_i \ll 1)$ times the energy of the transmitted signal at each antenna. $\mathbf{c}_i$ is approximated as the Gaussian model, $\mathbf{c}_i[n] \sim \mathscr{CN}(\mathbf{0}, \frac{\alpha_i}{N_s}\,\mathrm{diag}(\sum_{n=1}^{N_s}\mathbf{Q}_i[n]))$, where $\mathbf{Q}_i[n]$ is the Tx signal covariance matrix at subcarrier n of node i and can be written as $\mathbf{Q}_i[n] = \mathbf{V}^i\mathbf{G}_i[n]\mathbf{G}_i^H[n]\mathbf{V}^{i\,H}$ and $\mathbf{c}_i[n]$ is statistically independent of $\mathbf{x}_i[n]$. $\mathbf{e}_i[n]$ is the LDR noise at the Rx side and can be approximated as $\mathbf{e}_i[n] \sim \mathscr{CN}(\mathbf{0}, \frac{\beta_i}{N_s}\,\mathrm{diag}(\mathbf{Z}))$, where $\mathbf{Z}$ is the sum of the covariance matrix of the undistorted Rx signal across all subcarriers [58] assuming the subcarrier signals are decorrelated, $\mathbf{Z} = \sum_{n=1}^{N_s}\mathrm{E}(\mathbf{z}_i[n]\mathbf{z}_i^H[n])$, $\mathbf{z}_i[n] = \mathbf{y}_i[n] - \mathbf{e}_i[n]$ and $\mathbf{e}_i[n]$ is statistically independent of $\mathbf{z}_i[n]$. Also, $\beta_i \ll 1$. The Tx power (sum of all subcarrier powers) constraint at node j can be written as $\sum_{n=1}^{N_s}\mathrm{tr}\{\mathbf{V}^{j\,H}\mathbf{V}^j\,\mathbf{G}_j[n]\,\mathbf{G}_j^H[n]\} \leq P_j$. We introduce a digital self-interference canceller at the baseband which subtracts the residual interference signal $\mathbf{H}_{i,i}\mathbf{x}_i$ from the received signal. Assuming that $\mathbf{H}_{i,i}$ is perfectly estimated at the baseband and since $\mathbf{x}_i$ is already known to node i, we can rewrite the received signal at the baseband as

$$\begin{aligned}\mathbf{y}_i'[n] &= \mathbf{y}_i[n] - \mathbf{F}_{RF,i}\mathbf{H}_{i,i}[n]\mathbf{x}_i[n] \\ &= \mathbf{F}_{RF,i}\mathbf{H}_{i,j}[n]\mathbf{x}_j[n] + \mathbf{v}_i[n],\end{aligned} \tag{3.2}$$

where

$$\mathbf{v}_i[n] = \mathbf{F}_{RF,i}\mathbf{H}_{i,j}[n]\mathbf{c}_j[n] + \mathbf{F}_{RF,i}\mathbf{H}_{i,i}[n]\mathbf{c}_i[n] + \mathbf{e}_i[n] + \mathbf{F}_{RF,i}\mathbf{n}_i[n] \tag{3.3}$$

is the unknown interference plus noise component after SI cancellation. In this chapter, for our BF design, we assume that all the channel matrices and scaling factors in (3.1) are known. Also,

another point worth noting here is that the dependence of the signal model (3.2) on the SI power is only through the LDR noise and the BF design in the next section try to reduce the LDR noise significantly.

3.2.1 Channel Model

In this sub-section, we omit the node indices for simplicity. Considering a delay-d geometric direct channel model for a mmWave propagation environment [41] with L_s scattering clusters and L_r scatterers or rays in each cluster, we have

$$\mathbf{H}_d = \sum_{s=1}^{L_s} \sum_{l=1}^{L_r} \alpha_{s,l} \mathbf{h}_r(\theta_{s,l}) \mathbf{h}_t(\phi_{s,l})^H p(dT_s - \tau_s - \tau_{rl}) \tag{3.4}$$

Here $\theta_{s,l}, \phi_{s,l}$ represent the angle of arrival (AoA) and angle of departure (AoD) respectively for the l^{th} path in the s^{th} cluster. $\mathbf{h}_r(\cdot), \mathbf{h}_t(\cdot)$ represent the antenna array responses at Rx and Tx respectively. The complex path gain, $\alpha_{s,l} \sim \mathscr{CN}(0, \frac{N_t N_r}{L_s L_r})$ and $p(\tau)$ represents the band-limited pulse shaping filter response evaluated at τ seconds. Each cluster has a time delay $\tau_s \in \mathscr{R}$ and each ray $l = 1, ..L_r$ has a relative time delay τ_{rl}. $\frac{1}{T_s}$ represents the sampling rate. The total delay of any path is $dT_s - \tau_s - \tau_{rl}$. Now, we write the $(m,n)-$the element of the channel in the subcarrier n as

$$\mathbf{H}[n] = \sum_{d=1}^{D} \mathbf{H}_d e^{-j2\pi \frac{nd}{N_s}}. \tag{3.5}$$

In a more compact form, this can be represented as,

$$\begin{aligned} \mathbf{H}[n] &= \mathbf{H}_r \sum_{d=1}^{D} \mathbf{A}_d[n] \mathbf{H}_t^H, \text{ where} \\ \mathbf{H}_r &= [\mathbf{h}_r(\theta_{1,1}), ..., \mathbf{h}_r(\theta_{L_s,L_r})], \\ \mathbf{H}_t &= [\mathbf{h}_t(\phi_{1,1}), ..., \mathbf{h}_t(\phi_{L_s,L_r})], \\ \mathbf{A}_d[n] &= \text{diag}(\alpha_{1,1} p(dTs - \tau_1 - \tau_{r1}), \alpha_{L_s,L_r} p(dTs - \tau_{L_s} - \tau_{rL_r})) e^{-j2\pi \frac{nd}{N_s}}. \end{aligned} \tag{3.6}$$

Note that our HBF design which follows, is applicable for general MIMO channel models and the channel model outlined here is utilized for the simulations in Section VI. Further considering the SI channel, as the distance between the transmit and receive arrays does not satisfy the far-field range condition, we need to employ the near-field model which has a spherical wavefront. In such a case, the SI channel coefficients highly depend on the placement of the transmit and receive arrays and can be written as

$$(\mathbf{H}_{i,i})_{m,n} = \frac{\rho}{r_{m,n}} \exp(-j2\pi \frac{r_{m,n}}{\lambda}), \tag{3.7}$$

where $r_{m,n}$ is the distance between $m-$th element of the receive array and $n-$the element of the transmit array and ρ being the SI channel power normalization factor. Note that, (3.7) is a simple model which does not take into account the mutual antenna coupling or signal reflections in the SI channel.

3.3 WSR maximization through WSMSE

Consider the optimization of the two-stage BF/hybrid combiner design using WSR maximization of the Multi-cell MU-MIMO system:

$$\begin{aligned}[\mathbf{V}\,\mathbf{G}\,\mathbf{F}_{RF}\,\mathbf{F}_{BB}] &= \arg\max_{\substack{\mathbf{V},\mathbf{G},\\ \mathbf{F}_{RF},\mathbf{F}_{BB}}} WSR(\mathbf{G},\mathbf{V},\mathbf{F}_{RF},\mathbf{F}_{BB}) \\ &= \arg\max_{\mathbf{V},\mathbf{G}} \sum_{i=1}^{2}\sum_{n=1}^{N_s} u_i \ln\det\left(\mathbf{R}_{\bar{i}}^{-1}[n]\mathbf{R}_i[n]\right),\end{aligned} \tag{3.8}$$

where the u_i are the rate weights, $\mathbf{G}$ represents the collection of digital BFs $\mathbf{G}_i[n]$, $\mathbf{V}$ the collection of analog BFs $\mathbf{V}^i$. For the BF design considered in this section, the underlying assumption is that all the channels are perfectly known at the Rx and Tx side. In addition to this, the Tx signal covariance matrix $\mathrm{E}(\mathbf{d}_j[n]\mathbf{d}_j[n]^H) = \mathbf{I}$ is assumed to be known at the Rx side. At the receiver, we apply a hybrid combiner with analog BF denoted by $\mathbf{F}_{RF,i}$ of size $M_r^i \times N_r^i$, where M_r^i represents the number of RF chains at the Rx side. $\mathbf{F}_{BB,i}[n]$ represent the baseband digital combiner of size $d_j \times M_r^i$. For notational convenience, we define the received signal covariance matrices $\mathbf{\Theta}_{i,j}[n] = \mathbf{H}_{i,j}[n]\mathbf{Q}_j[n]\mathbf{H}_{i,j}^H[n], \Phi_{i,j}[n] = \mathbf{H}_{i,j}[n]\,\mathrm{diag}\,(\mathbf{Q}_j[n])\mathbf{H}_{i,j}^H[n]$. Similarly the self interence parts $\mathbf{\Theta}_{i,i}[n], \Phi_{i,i}[n]$ are also defined. The covariance matrix of the effective noise part at the output of the RF chains, $\mathbf{R}_{\bar{i}}[n]$ can be approximated under $\alpha_i \ll 1, \beta_i \ll 1$ as follows [59]

$$\mathbf{R}_{\bar{i}}[n] = \mathbf{F}_{RF,i}(\alpha_j\Phi_{i,j}[n] + \alpha_i\Phi_{i,i}[n])\mathbf{F}_{RF,i}^H + \beta_i\,\mathrm{diag}\,(\mathbf{F}_{RF,i}(\mathbf{\Theta}_{i,j}[n] + \mathbf{\Theta}_{i,i}[n])\mathbf{F}_{RF,i}^H) \tag{3.9}$$

Also define, $\mathbf{R}_i[n] = \mathbf{R}_{\bar{i}}[n] + \mathbf{F}_{RF,i}\mathbf{\Theta}_{i,j}[n]\mathbf{F}_{RF,i}^H$,

where $\mathbf{R}_i[n]$ is the signal plus interference plus noise covariance matrix. Further after the receive combining, we obtain $\mathbf{\Sigma}_{\bar{i}}[n] = \mathbf{F}_{BB,i}[n]\mathbf{R}_{\bar{i}}[n]\mathbf{F}_{BB,i}^H[n]$ and $\mathbf{\Sigma}_i[n] = \mathbf{F}_{BB,i}[n]\mathbf{R}_i[n]\mathbf{F}_{BB,i}^H[n]$. Direct maximization of (3.8), however, requires a joint optimization over the four matrix variables $(\mathbf{V},\mathbf{G},\mathbf{F}_{RF},\mathbf{F}_{BB})$. Unfortunately, finding a global optimum solution for similarly constrained optimization is found to be intractable. So we decouple the joint transmitter-receiver optimization and focus on the design of the Rx combiners first. We assume that the node i applies the hybrid combiner $\mathbf{F}_i[n] = \mathbf{F}_{BB,i}[n]\mathbf{F}_{RF,i}$ to estimate the signal transmitted from node j. The analog combiner $\mathbf{F}_{RF,i}$ serves to reduce the SI component from the received signal, while the digital combiner $\mathbf{F}_{BB,i}$ decouples the streams $(\mathbf{d}_j)$ intended for user i from j. The estimated signal $\widehat{\mathbf{d}}_j[n]$ can be written as

$$\widehat{\mathbf{d}}_j[n] = \mathbf{F}_i[n]\mathbf{H}_{i,j}[n]\mathbf{x}_j[n] + \mathbf{F}_{BB,i}[n]\mathbf{v}_i[n]. \tag{3.10}$$

At the Rx side, maximizing the WSR is equivalent to minimizing the weighted MSE with the MSE weights being chosen as $\mathbf{W}_i[n] = u_i\mathbf{R}_{\tilde{\mathbf{d}}_j\tilde{\mathbf{d}}_j}[n]^{-1}$ [9,54]. Further, we can obtain the error covariance matrix for the detection of $\mathbf{d}_j$ at node i as

$$\begin{aligned}\mathbf{R}_{\tilde{\mathbf{d}}_j\tilde{\mathbf{d}}_j}[n] &= \mathrm{E}\{(\widehat{\mathbf{d}}_j[n] - \mathbf{d}_j[n])(\widehat{\mathbf{d}}_j[n] - \mathbf{d}_j[n])^H\} \\ &= (\mathbf{F}_i[n]\mathbf{H}_{i,j}[n]\mathbf{V}^j\mathbf{G}_j[n] - \mathbf{I})(\mathbf{F}_i[n]\mathbf{H}_{i,j}[n]\mathbf{V}^j\mathbf{G}_j[n] - \mathbf{I})^H + \mathbf{F}_{BB,i}\mathbf{R}_{\bar{i}}[n]\mathbf{F}_{BB,i}[n]^H.\end{aligned} \tag{3.11}$$

The MMSE Rx combiner can be alternatively optimized as follows

$$\begin{aligned}[\mathbf{F}_{RF,i}, \mathbf{F}_{BB,i}[n], \forall n] &= \arg\min_{\mathbf{F}_{RF,i},\mathbf{F}_{BB,i}[n]} \sum_{n=1}^{N_s} \mathrm{tr}\{\mathbf{R}_{\tilde{\mathbf{d}}_j\tilde{\mathbf{d}}_j}[n]\}, \\ \mathbf{F}_{BB,i}[n] &= \mathbf{G}_j^H[n]\mathbf{V}^{j\,H}\mathbf{H}_{i,j}^H[n]\mathbf{F}_{RF,i}^H\mathbf{R}_i[n]^{-1}\end{aligned} \tag{3.12}$$

Optimization of the digital BF in (3.12) can be done independently across different subcarriers, as it is evident. We define $\mathbf{F}_{BB,i}^H[n]\mathbf{F}_{BB,i}[n] = \mathbf{P}_{B,i}[n]$, $\sum_{n=1}^{N_s}[(\mathbf{\Theta}_{i,j}[n])^T \otimes \mathbf{P}_{B,i}[n] + ((\alpha_j\Phi_{i,j}[n] + \alpha_i\Phi_{i,i}[n])^T \otimes \mathbf{P}_{B,i}[n]) + (\beta_i(\mathbf{\Theta}_{i,j}[n] + \mathbf{\Theta}_{i,i}[n])^T \otimes \mathrm{diag}(\mathbf{P}_{B,i}[n])) = \mathbf{B}_i$. To derive the unconstrained analog BF matrix, we take the gradient of (3.12) w.r.t $\mathbf{F}_{RF,i}^*$

$$\begin{aligned}&\sum_{n=1}^{N_s}\mathbf{P}_{B,i}[n]\mathbf{F}_{RF,i}\mathbf{\Theta}_{i,j}[n] - \mathbf{F}_{BB,i}^H[n]\mathbf{G}_j^H[n]\mathbf{V}^{jH}\mathbf{H}_{i,j}^H[n] + \\ &\mathbf{P}_{B,i}[n]\mathbf{F}_{RF,i}(\alpha_j\Phi_{i,j}[n] + \alpha_i\Phi_{i,i}[n]) + \beta_i\,\mathrm{diag}(\mathbf{P}_{B,i}[n])\mathbf{F}_{RF,i}(\mathbf{\Theta}_{i,j}[n] + \mathbf{\Theta}_{i,i}[n]) = 0,\\ &\mathbf{B}_i\,vec(\mathbf{F}_{RF,i}) \overset{(a)}{=} vec(\sum_{n=1}^{N_s}\mathbf{F}_{BB,i}^H[n]\mathbf{G}_j^H[n]\mathbf{V}^{jH}\mathbf{H}_{i,j}^H[n]).\end{aligned} \tag{3.13}$$

Note that the gradient calculation is done through Wirtinger Calculus [60]. In (a), we use the result $vec(\mathbf{AXB}) = (\mathbf{B}^T \otimes \mathbf{A})\,vec(\mathbf{X})$ [31]. Further, we obtain the expression for the analog combiner as

$$vec(\mathbf{F}_{RF,i}) = \mathbf{B}_i^H\,vec(\sum_{n=1}^{N_s}\mathbf{F}_{BB,i}^H[n]\mathbf{G}_j^H[n]\mathbf{V}^{jH}\mathbf{H}_{i,j}^H[n]), \tag{3.14}$$

where $(\cdot)^H$ represents the pseudoinverse.

3.3.1 Two-stage transmit BF design

In this section, we consider the design of two-stage Tx BFs $\mathbf{V}^j, \mathbf{G}_j[n]$ under a sum power constraint at the Tx. To facilitate the gradients, we use the result $\frac{\partial \mathrm{tr}\{\mathbf{A}\,\mathrm{diag}(\mathbf{CXD})\mathbf{B}\}}{\partial \mathbf{X}} = [\mathbf{D}\,\mathrm{diag}(\mathbf{BA})\mathbf{C}]^T$. The derivations for this gradient result are provided in Appendix D. We propose to design the Tx BFs using weighted sum MSE and can be formulated as follows

$$\begin{aligned}&\min_{\substack{\mathbf{V}^i,\mathbf{G}_i[n],\\ \mathbf{V}^j,\mathbf{G}_j[n]}}\ \sum_{n=1}^{N_s}\mathrm{tr}\{\mathbf{W}_i[n]\,\mathrm{E}(\widehat{\mathbf{d}}_j[n] - \mathbf{d}_j[n])(\widehat{\mathbf{d}}_j[n] - \mathbf{d}_j[n])^H\} + \\ &\mathrm{tr}\{\mathbf{W}_j[n]\,\mathrm{E}\{(\widehat{\mathbf{d}}_i[n] - \mathbf{d}_i[n])(\widehat{\mathbf{d}}_i[n] - \mathbf{d}_i[n])^H\},\\ &\text{s.t.}\ \sum_{n=1}^{N_s}\mathrm{tr}\{\mathbf{Q}_i[n]\} <= P_i, \forall i.\end{aligned} \tag{3.15}$$

Here $\mathbf{W}_i[n]$ represents the weight matrix of size $d_i \times d_i$. Augmenting the power constraints, the Lagrangian function can be written as

$$\begin{aligned}\mathscr{L} = &\sum_{n=1}^{N_s}\sum_{i=1}^{2}\sum_{j=1,j\neq i}^{2}\mathrm{tr}\{\mathbf{W}_i[n](\mathbf{I} - \mathbf{G}_j^H[n]\mathbf{V}^{jH}\mathbf{H}_{i,j}^H[n]\mathbf{F}_i^H[n]\\ &- \mathbf{F}_i[n]\mathbf{H}_{i,j}[n]\mathbf{V}^j\mathbf{G}_j[n] + \mathbf{F}_i[n]\mathbf{H}_{i,j}[n]\mathbf{Q}_j\mathbf{H}_{i,j}^H[n]\mathbf{F}_i^H[n]\\ &+ \mathbf{F}_{BB,i}[n]\mathbf{R}_{\bar{i}}[n]\mathbf{F}_{BB,i}^H[n])\} + (\sum_{i=1}^{2}\lambda_i(\sum_{n=1}^{N_s}\mathrm{tr}\{\mathbf{Q}_i[n]\}) - P_i),\end{aligned} \tag{3.16}$$

For convenience of the analysis, we define

$$\begin{aligned}\mathbf{A}_j[n] &= \mathbf{F}_j^H[n]\mathbf{W}_j[n]\mathbf{F}_j[n],\\ \widehat{\mathbf{A}}_j[n] &= \mathbf{F}_{BB,j}^H[n]\mathbf{W}_j[n]\mathbf{F}_{BB,j}[n].\end{aligned} \tag{3.17}$$

Taking the partial derivative of (3.16) with respect to the inner BF $\mathbf{G}_j[n]$, we obtain

$$\begin{aligned}&-\mathbf{V}^{j\,H}\mathbf{H}_{i,j}^H[n]\mathbf{F}_i^H[n]\mathbf{W}_i[n]+\mathbf{V}^{j\,H}\mathbf{H}_{i,j}^H[n]\mathbf{A}_i[n]\mathbf{H}_{i,j}[n]\mathbf{V}^j\mathbf{G}_j[n]\\&+\frac{\mathrm{tr}\{\mathbf{F}_{BB,i}[n]^H\mathbf{F}_{BB,i}[n]\partial\mathbf{R}_{\bar{i}}[n]\}}{\partial\mathbf{G}_j[n]}+\frac{\mathrm{tr}\{\mathbf{F}_{BB,j}[n]^H\mathbf{F}_{BB,j}[n]\partial\mathbf{R}_{\bar{j}}[n]\}}{\partial\mathbf{G}_j[n]}+\lambda_j\mathbf{V}^{j\,H}\mathbf{V}^j\mathbf{G}_j[n]=\mathbf{0},\text{ where, } i\neq j.\end{aligned}\tag{3.18}$$

Using the expression for $\mathbf{R}_{\bar{i}}[n]$ in (3.9), we can write

$$\begin{aligned}\frac{\mathrm{tr}\{\mathbf{F}_{BB,i}[n]^H\mathbf{F}_{BB,i}[n]\partial\mathbf{R}_{\bar{i}}[n]\}}{\partial\mathbf{G}_j[n]}&=\alpha_j\mathbf{V}^{j\,H}\,\mathrm{diag}\,(\mathbf{H}_{i,j}^H[n]\mathbf{A}_i[n]\mathbf{H}_{i,j}[n])\mathbf{V}^j\mathbf{G}_j[n]\\&\quad+\beta_i\mathbf{V}^{j\,H}\mathbf{H}_{i,j}^H[n]\mathbf{F}_{RF,i}^H\,\mathrm{diag}\,(\widehat{\mathbf{A}}_i[n])\mathbf{F}_{RF,i}\mathbf{H}_{i,j}[n]\mathbf{V}^j\mathbf{G}_j[n],\\\frac{\mathrm{tr}\{\mathbf{F}_{BB,j}[n]^H\mathbf{F}_{BB,j}[n]\partial\mathbf{R}_{\bar{j}}[n]\}}{\partial\mathbf{G}_j[n]}&=\alpha_j\mathbf{V}^{j\,H}\,\mathrm{diag}(\mathbf{H}_{j,j}^H[n]\mathbf{A}_j[n]\mathbf{H}_{j,j})\mathbf{V}^j\mathbf{G}_j[n]\\&\quad+\beta_j\mathbf{V}^{j\,H}\mathbf{H}_{j,j}^H[n]\mathbf{F}_{RF,i}^H\,\mathrm{diag}(\widehat{\mathbf{A}}_j[n])\mathbf{F}_{RF,i}\mathbf{H}_{j,j}[n]\mathbf{V}^j\mathbf{G}_j[n],\end{aligned}\tag{3.19}$$

By substituting (3.19) in (3.18), we obtain the optimal $\mathbf{G}_j[n]$ as

$$\mathbf{G}_j[n]=(\mathbf{S}_j[n]+\lambda_j\mathbf{V}^{j\,H}\mathbf{V}^j)^{-1}\mathbf{V}^{j\,H}\mathbf{H}_{i,j}^H[n]\mathbf{F}_i^H[n]\mathbf{W}_i[n],\tag{3.20}$$

where $\mathbf{S}_j[n]$ can be interpreted as the signal plus interference power seen by the digital BF at the Tx side and is expressed as

$$\begin{aligned}\mathbf{S}_j[n]&=\mathbf{V}^{j\,H}\mathbf{H}_{i,j}^H[n]\mathbf{A}_i[n]\mathbf{H}_{i,j}[n]\mathbf{V}^j+\alpha_j\mathbf{V}^{j\,H}\,\mathrm{diag}\,(\mathbf{H}_{i,j}^H[n]\mathbf{A}_i[n]\mathbf{H}_{i,j}[n])\mathbf{V}^j\\&\quad+\beta_i\mathbf{V}^{j\,H}\mathbf{H}_{i,j}^H[n]\mathbf{F}_{RF,i}^H\,\mathrm{diag}\,(\widehat{\mathbf{A}}_i[n])\mathbf{F}_{RF,i}\mathbf{H}_{i,j}[n]\mathbf{V}^j+\alpha_j\mathbf{V}^{j\,H}\,\mathrm{diag}\,(\mathbf{H}_{j,j}^H[n]\mathbf{A}_j[n]\mathbf{H}_{j,j}[n])\mathbf{V}^j\\&\quad+\beta_j\mathbf{V}^{j\,H}\mathbf{H}_{j,j}^H[n]\mathbf{F}_{RF,j}^H\,\mathrm{diag}\,(\widehat{\mathbf{A}}_j[n])\mathbf{F}_{RF,j}\mathbf{H}_{j,j}[n]\mathbf{V}^j\end{aligned}\tag{3.21}$$

The values of the Lagrangian multipliers $\lambda_j\geq 0,\forall\, j$ are chosen such that the respective power constraint is satisfied (3.15). To compute this, we follow a similar approach as in [28] but extended to two-stage BF here. Considering the eigen decomposition of $\mathbf{S}_j[n]=\mathbf{U}_j\mathbf{\Lambda}_j[n]\mathbf{U}_j^H$, $\mathbf{V}^{j\,H}\mathbf{V}^j=\mathbf{U}_j\Delta_j\mathbf{U}_j^H$ and let

$$\Phi[n]=\mathbf{U}_j^H\mathbf{V}^{j\,H}\mathbf{H}_{i,j}^H[n]\mathbf{F}_i^H[n]\mathbf{W}_i[n]\mathbf{W}_i[n]^H\mathbf{F}_i[n]\mathbf{H}_{i,j}[n]\mathbf{V}^j\mathbf{U}_j\tag{3.22}$$

and expanding the power constraint

$$\sum_{n=1}^{N_s}\mathrm{tr}\{\mathbf{V}^j\mathbf{G}_j[n](\lambda_j)\mathbf{G}_j^H[n](\lambda_j)\mathbf{V}^{j\,H}\}=P_j,\tag{3.23}$$

we get the simplified expression

$$\sum_{n=1}^{N_s}\sum_{k=1}^{M_t^j}\frac{\Phi[n]_{k,k}(\Delta_j)_{k,k}}{((\mathbf{\Lambda}_j[n])_{k,k}+\lambda_j(\Delta_j)_{k,k})^2}=P_j.\tag{3.24}$$

Here $\mathbf{X}_{k,k}$ represents the k^{th} diagonal element of the matrix $\mathbf{X}$. Note that the $\lambda_j\geq 0$ and the left-hand side of (3.24) is a decreasing function of λ_j for $\lambda_j>0$. Hence we can compute the values of λ_j using one-dimensional linear search techniques such as bisection. Further, we consider the optimization of the outer BF at the Tx side, $\mathbf{V}^j$. Given the inner BFs, we update the outer beamformers $\mathbf{V}^j$. Taking the partial derivative of (3.16) with respect to the inner BF $\mathbf{V}^j$, we obtain

$$\begin{aligned}&-\mathbf{H}_{i,j}^H[n]\mathbf{F}_i^H[n]\mathbf{W}_i[n]\mathbf{G}_j^H[n]+\mathbf{H}_{i,j}^H[n]\mathbf{F}_i^H[n]\mathbf{W}_i[n]\mathbf{F}_i[n]\mathbf{H}_{i,j}[n]\mathbf{V}^j\mathbf{G}_j[n]\mathbf{G}_j^H[n]\\&+\frac{\mathrm{tr}\{\mathbf{F}_{BB,i}[n]^H\mathbf{F}_{BB,i}[n]\partial\mathbf{R}_{\bar{i}}[n]\}}{\partial\mathbf{V}^j[n]}+\frac{\mathrm{tr}\{\mathbf{F}_{BB,j}[n]^H\mathbf{F}_{BB,j}[n]\partial\mathbf{R}_{\bar{j}}[n]\}}{\partial\mathbf{V}^j[n]}+\lambda_j\mathbf{V}^j\mathbf{G}_j[n]\mathbf{G}_j^H[n]=0,\text{where, } i\neq j.\end{aligned}\tag{3.25}$$

For notational convenience, we define $\mathbf{P}_{G,j}[n] = \mathbf{G}_j[n]\mathbf{G}_j^H[n]$. Using the expression for $\mathbf{R}_{\bar{i}}[n]$ in (3.9), we can write

$$
\begin{aligned}
\frac{\mathrm{tr}\{\mathbf{F}_{BB,i}[n]^H\mathbf{F}_{BB,i}[n]\partial\mathbf{R}_{\bar{i}}[n]\}}{\partial\mathbf{V}^j[n]} &= \alpha_j \,\mathrm{diag}\,(\mathbf{H}_{i,j}^H[n]\mathbf{A}_i[n]\mathbf{H}_{i,j}[n])\mathbf{V}^j\mathbf{P}_{G,j}[n] \\
&\quad + \beta_i\mathbf{H}_{i,j}^H[n]\mathbf{F}_{RF,i}^H\,\mathrm{diag}\,(\widehat{\mathbf{A}}_i[n])\mathbf{F}_{RF,i}\mathbf{H}_{i,j}[n]\mathbf{V}^j\mathbf{P}_{G,j}[n], \\
\frac{\mathrm{tr}\{\mathbf{F}_{BB,j}[n]^H\mathbf{F}_{BB,j}[n]\partial\mathbf{R}_{\bar{j}}[n]\}}{\partial\mathbf{V}^j[n]} &= \alpha_j \,\mathrm{diag}\,(\mathbf{H}_{j,j}^H[n]\mathbf{A}_j[n]\mathbf{H}_{j,j}[n])\mathbf{V}^j\mathbf{P}_{G,j}[n] \\
&\quad + \beta_j\mathbf{H}_{j,j}^H[n]\mathbf{F}_{RF,j}^H\,\mathrm{diag}\,(\widehat{\mathbf{A}}_j[n])\mathbf{F}_{RF,j}\mathbf{H}_{j,j}[n]\mathbf{V}^j\mathbf{P}_{G,j}[n]],
\end{aligned}
\tag{3.26}
$$

By substituting (3.26) in (3.25) and using the result $vec(\mathbf{AXB}) = (\mathbf{B}^T \otimes \mathbf{A})vec(\mathbf{X})$, we obtain the optimal $\mathbf{V}^j$ as

(3.27)

$$
\begin{aligned}
vec(\mathbf{V}^j) &= \mathbf{B}_j^H \sum_{n=1}^{N_s} \mathbf{H}_{i,j}^H[n]\mathbf{F}_i^H[n]\mathbf{W}_i[n]\mathbf{G}_j^H[n], \text{ where} \\
\mathbf{B}_j &= \sum_{n=1}^{N_s} (\mathbf{P}_{G,j}[n] \otimes \mathbf{H}_{i,j}^H[n]\mathbf{A}_i[n]\mathbf{H}_{i,j}[n]) + \alpha_j\mathbf{P}_{G,j}[n] \otimes \mathrm{diag}\,(\mathbf{H}_{i,j}^H[n]\mathbf{A}_i[n]\mathbf{H}_{i,j}[n]) \\
&\quad + \beta_i\mathbf{P}_{G,j}[n] \otimes \mathbf{H}_{i,j}^H[n]\mathbf{F}_{RF,i}^H\,\mathrm{diag}\,(\widehat{\mathbf{A}}_i[n])\mathbf{F}_{RF,i}\mathbf{H}_{i,j}[n] + \alpha_j\mathbf{P}_{G,j}[n] \otimes \mathrm{diag}\,(\mathbf{H}_{j,j}^H\mathbf{A}_j\mathbf{H}_{j,j}[n]) \\
&\quad + \beta_j(\mathbf{P}_{G,j}[n] \otimes (\mathbf{H}_{j,j}^H[n]\mathbf{F}_{RF,j}^H\,\mathrm{diag}\,(\widehat{\mathbf{A}}_j[n])\mathbf{F}_{RF,j}\mathbf{H}_{j,j}[n])).
\end{aligned}
$$

Alternating WSR maximization between digital and analog BF or the two-stage BFs at Tx/Rx now leads to Algorithm 7. We remark that we propose to either use a two-stage BF at Tx or hybrid combiner at the Rx to null the SI power and both stages are not required if the antenna or BF/combiner dimensions are sufficient as discussed in Section 3.3.2.

Algorithm 7: LDR Multi Stage BF Design via WSMSE

Given: $P_i, \mathbf{H}_{i,j}, \mathbf{H}_{i,i}, u_i \,\forall\, i, j$.
Initialization: $\mathbf{F}_{RF,i} = e^{j\angle\mathbf{V}_{1:M^i}(\mathbf{H}_{t,i,j})}$, The $\mathbf{G}_i$ are taken as the ZF precoders for the effective channels $\mathbf{V}^i\mathbf{H}_{j,i}$ with uniform powers.
Iteration (t) :

1. Update the Rx side HBF, i.e $\mathbf{F}_{BB,i}^{(t)}[n], \mathbf{F}_{RF,i}^{(t)} \forall\, i$ using (3.12), (3.14) respectively.
2. Update $\mathbf{G}_i'^{(t)}[n]$, $\forall\, i$, from (3.20).
3. Update $\mathbf{V}^{i\,(t)}, \forall\, i$ from (3.27) and λ_i using bisection method from (3.24).
4. Check for convergence of the WSR: if not go to step 1.

3.3.2 Hybrid Combiner/Two-Stage BF Capabilities for SI Power Reduction

In this section, we analyze to what extent a hybrid combiner can achieve the same performance as a fully digital BF and reduce the LDR noise originating from both the direct and SI channels.

Directly optimizing the phasor values of the analog combiner alternatively using the WSR cost function, which is a non-convex function, results in a lot of local optima depending on the initialization [10]. So we utilize here one approach called deterministic annealing (DA) to avoid the problem of local optima and it is discussed in detail in our papers [44, Algorithm 3], [46].

In particular, we shall see that this is possible for a sufficient number of RF chains and with the arbitrary antenna array responses. Consider a specular or pathwise channel model with say L_d multi-paths per link for the direct channel and L_I for the SI channel. For notational simplicity, we shall consider a uniform L_d, L_I and $N_k = N_t^i = N_r^i, \forall i$.

Theorem 3. *For a bidirectional full-duplex MIMO system with the number of Rx RF chains $M_r^i \geq L_d$ or the number of Tx RF chains $M_t^i >= L_d$ and arbitrary antenna responses for the direct channel, to achieve optimal all-digital precoding performance at high SNR and mitigation of LDR noise, the unconstrained analog combiner or the time domain Tx BF can be chosen as matched filtered to the direct link channel projected on the orthogonal complement of the low rank SI channel.*

Proof: From [9] or [11, eq. (13)], the optimal all-digital beamformer is of the form

$$\begin{aligned}\mathbf{F}_i[n] &= \mathbf{G}_j^H[n]\mathbf{V}^{jH}\mathbf{H}_{i,j}^H[n](\mathbf{H}_{i,j}[n]\mathbf{Q}_j[n]\mathbf{H}_{i,j}^H[n] + \mathbf{R}_{\tilde{i}}[n])^{-1} \\ &= \mathbf{G}_j^H[n]\mathbf{V}^{jH}\mathbf{H}_{t,i,j}\sum_{d=1}^{D}\mathbf{A}_{d,i,j}[n]^H\mathbf{H}_{r,i,j}^H(\mathbf{H}_{i,j}[n]\mathbf{Q}_j[n]\mathbf{H}_{i,j}^H[n] + \mathbf{R}_{\tilde{i}}[n])^{-1},\end{aligned} \tag{3.28}$$

where $\mathbf{R}_{\tilde{i}}[n]$ is the interference plus noise power received. For the mmWave channel model (3.6), when $N_k \to \infty$, the terms of the form $\mathbf{H}_{i,j}[n]\mathbf{Q}_j[n]\mathbf{H}_{i,j}^H[n]$ can be simplified as

$$\begin{aligned}\mathbf{H}_{i,j}[n]\mathbf{Q}_j[n]\mathbf{H}_{i,j}^H[n] &= \mathbf{H}_{r,i,j}(\sum_{d=1}^{D}\mathbf{A}_{d,i,j}[n])\mathbf{H}_{t,i,j}^H\mathbf{Q}_j[n]\mathbf{H}_{t,i,j}(\sum_{d=1}^{D}\mathbf{A}_{d,i,j}^H[n])\mathbf{H}_{r,i,j}^H \\ &\overset{(a)}{=} \frac{1}{N_t^j}\mathbf{H}_{r,i,j}(\sum_{d=1}^{D}\mathbf{A}_{d,i,j}^2)\mathrm{tr}\{\mathbf{Q}_j[n]\}\mathbf{H}_{r,i,j}^H,\end{aligned} \tag{3.29}$$

where $\sum_{d=1}^{D}\mathbf{A}_{d,i,j}^2 = \sum_{d=1}^{D}\mathbf{A}_{d,i,j}[n]\mathbf{A}_{d,i,j}^H[n]$ is independent of the subcarrier index. In (a), we made the assumption that the Tx array response becomes asymptotically orthogonal. Further assuming that at high SNR the power transmitted across each subcarrier becomes same, then $\mathrm{tr}\{\mathbf{Q}_j[n]\} = \frac{P_j}{N_s}$ and thus $\mathbf{H}_{i,j}[n]\mathbf{Q}_j[n]\mathbf{H}_{i,j}^H[n]$ becomes independent of the frequency. Similarly $\mathbf{R}_{\tilde{i}}[n]$ also becomes independent of the frequency since the terms in $\mathbf{R}_{\tilde{i}}[n]$ are also of similar form as $\mathbf{H}_{i,j}[n]\mathbf{Q}_j[n]\mathbf{H}_{i,j}^H[n]$. We denote $\mathbf{R}_i = \mathbf{H}_{i,j}[n]\mathbf{Q}_j[n]\mathbf{H}_{i,j}^H[n] + \mathbf{R}_{\tilde{i}}[n]$. Thus we can separate the BFs as

$$\begin{aligned}\mathbf{F}_{RF,i} &= \mathbf{H}_{r,i,j}^H\mathbf{R}_i^{-1}, \\ \mathbf{F}_{BB,i}[n] &= \mathbf{G}_j^H[n]\mathbf{V}^{jH}\mathbf{H}_{t,i,j}\sum_{d=1}^{D}\mathbf{A}_{d,i,j}[n].\end{aligned} \tag{3.30}$$

Similarly, considering the Tx side BF design, the optimal fully digital BF can be written as (3.20)

$$\mathbf{G}_j[n] = (\mathbf{S}_j[n] + \lambda_j\mathbf{I})^{-1}\mathbf{H}_{i,j}^H[n]\mathbf{F}_i^H[n]\mathbf{W}_i[n], \tag{3.31}$$

As $N_r^i \to \infty$ and substituting the pathwise model for the direct channels similar to the discussions above, we can observe that the quadratic term $(\sum_{d=1}^{D}\mathbf{A}_{d,i,j}[n]^H)\mathbf{H}_{r,i,j}^H\mathbf{A}_i\mathbf{H}_{r,i,j}(\sum_{d=1}^{D}\mathbf{A}_d[n]) = \mathbf{P}_r^i[n]$, where $\mathbf{P}_r^i[n]$ can be interpreted as the effective received power in subcarrier n, $\mathbf{P}_r^i[n] = \frac{1}{N_r^i}(\sum_{d=1}^{D}\mathbf{A}_{d,i,j}[n]^2)\mathrm{tr}\mathbf{A}_i$. $(\sum_{d=1}^{D}\mathbf{A}_{d,i,j}[n]^2)$ is independent of the subcarrier index (3.6) and hence effective received power in all the subcarrier becomes the same in the large antenna limit. Further,

substituting the $(\sum_{d=1}^{D} \mathbf{A}_{d,i,j}[n]^H)\mathbf{H}_{r,i,j}^H \mathbf{A}_i \mathbf{H}_{r,i,j}(\sum_{d=1}^{D} \mathbf{A}_d[n])$ in (3.21), we can see that $\mathbf{S}_j[n]$ is independent of the subcarrier index. Further defining $\widehat{\mathbf{S}}_j = \mathbf{S}_j[n] + \lambda_j \mathbf{I}$, we obtain, $\mathbf{V}^j = \widehat{\mathbf{S}}_j^{-1}\mathbf{H}_{t,i,j}$ and $\mathbf{G}_j = (\sum_{d=1}^{D} \mathbf{A}_{d,i,j}[n]^H)\mathbf{H}_{r,i,j}^H[n]\mathbf{F}_i^H[n]\mathbf{W}_i[n]$. Hence we can conclude that $\mathbf{V}^j, \mathbf{F}_{RF,j}$ depends only on the Tx/Rx antenna array responses. □

Note that whereas the digital BF or combiner $\mathbf{G}, \mathbf{F}_{BB}$ in (3.30) is a function of the instantaneous CSIT, the analog combiner $\mathbf{F}_{RF}$ or the outer precoder $\mathbf{V}$ is only a function of antenna array responses of the direct and SI channels, hence only of the slow fading channel components. Hence analog BF can be the same across all the subcarriers. We also remark that at high SNR, $\widehat{\mathbf{R}}_i$ or $\widehat{\mathbf{S}}_j$ converges to the projection matrix for the null space of the SI channel's Rx or Tx antenna array response matrix respectively. We remark that the main advantage of adding an analog BF stage is to suppress the SI before reaching the ADC and still preserving the signal dimensions by choosing a sufficient number of RF chains. Also, note that the analytical analog BF solution discussed here is unconstrained and further it requires the DA method to reach a phasor BF solution.

3.3.3 Simulation Results

Extensive Monte-Carlo simulations are conducted to validate the performance of the proposed hybrid BF algorithms that are presented for a bidirectional FD system under the LDR noise model. We follow the pathwise channel model $\mathbf{H}_{i,j}$ as in Section II.A, where the complex path gains are assumed to be Gaussian with variance distributed according to an exponential profile. For the SI channel, we ignore the near field effect of amplitude variation with distance and the near field effects in the phase variation. In the Uniform Linear Array (ULA), the AoD or AoA ϕ, θ are assumed to be uniformly distributed in the interval $[0^o, 30^o]$.

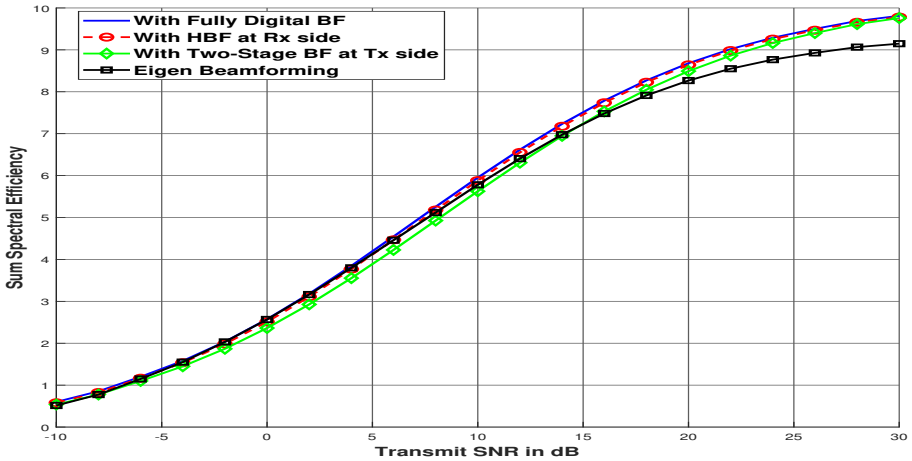


Figure 3.2: Sum Rate comparisons for, Single Carrier, $N_t^i = N_r^i = 8, M_t^i = M_r^i = 4, d_i = 1, \forall i, L = 4$ paths.

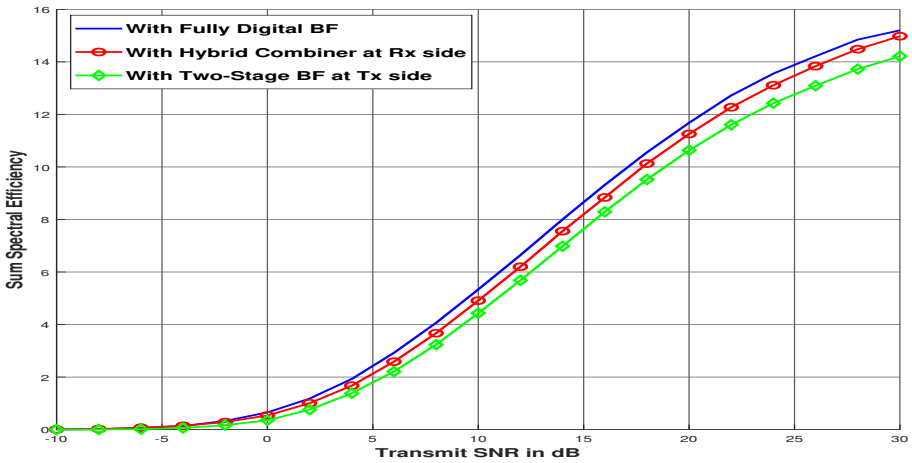


Figure 3.3: Sum Rate comparisons (per-subcarrier) for, OFDM, $N_s = 4, N_t^i = N_r^i = 8, M_t^i = M_r^i = 4, d_i = 1, \forall i, L = 4$ paths.

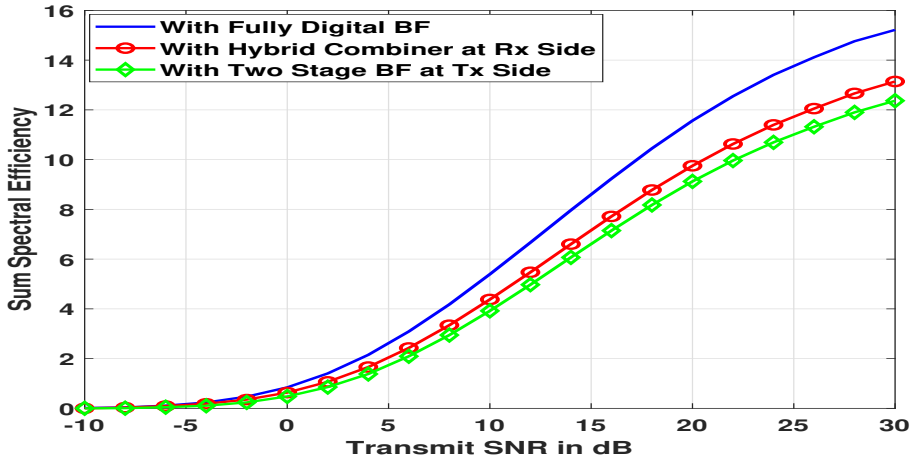


Figure 3.4: Sum Rate comparisons (per-subcarrier) for, OFDM, $N_s = 256, N_t^i = N_r^i = 8, M_t^i = M_r^i = 4, d_i = 1, \forall i, L = 6$ paths.

The dimensions of the two-stage BF and hybrid BF are such that the zero forcing capabilities at both sides are comparable. However, the number of LDR noises is the number of antennas at the Tx side, whereas, for the analog Rx stage, the number of LDR noises is the number of analog BF outputs, which is less. We conjecture that the analog BF reduces the LDR noise to a significant level and this would explain the better performance of the analog stage at Rx (in both figures) compared to the two-stage architecture at Tx. In Figure 3.2, we compare against the eigen beamforming (where the left and right singular vectors of the corresponding channels are used as the Combiner/BF and fully digital) and shows that its performance is inferior compared to our proposed design. However, one issue which remains to be investigated is shown by the performance in Figure 3.4, where we extend the number of subcarriers to 256. For higher number of subcarriers, the performance decrease due to the usage of common hybrid or time domain BF across all the subcarriers. It has to be mentioned that most of the state of the art works on hybrid beamforming solution (for half duplex systems) assumes a common hybrid beamforming stage for all subcarriers and hence it remains as an open problem for wideband OFDM systems to

design an efficient HBF solutions for large number of subcarriers.

3.3.4 Conclusion

In this chapter, we looked at beamforming solutions to null the SI power under a more practical noise model called as limited dynamic range. We proposed a multi-stage beamforming design (whose performance is validated through simulations), with a frequency flat analog or time domain combiner/BF stage and a frequency dependent baseband precoder/combiner. We decoupled the beamforming design for the Tx and Rx side. An iterative algorithm is obtained which jointly optimizes both analog/time domain and digital beamformers at the Tx/Rx side. We also discussed the dimensions of the BFs or combiners designed (for example, the minimum number of RF chains required) such that the SI power can be mitigated fully at high SNR.

3.4 Robust Beamforming Design under Partial CSIT

Note that $\widehat{\mathbf{H}}_{i,i}$ is the estimated SI channel at the baseband and since $\mathbf{x}_i$ is already known to node i, we can rewrite the received signal at the baseband as

$$\begin{aligned}\mathbf{y}_i'[n] &= \mathbf{y}_i[n] - \mathbf{F}_{RF,i}\widehat{\mathbf{H}}_{i,i}[n]\mathbf{x}_i[n] \\ &= \mathbf{F}_{RF,i}\mathbf{H}_{i,j}[n]\mathbf{x}_j[n] + \mathbf{v}_i[n],\end{aligned} \tag{3.32}$$

where $\mathbf{v}_i[n] = \mathbf{F}_{RF,i}(\mathbf{H}_{i,j}[n]\mathbf{c}_j[n] + \mathbf{H}_{i,i}[n]\mathbf{c}_i[n]) + \mathbf{e}_i[n] + \mathbf{F}_{RF,i}\widetilde{\mathbf{H}}_{i,i}[n]\mathbf{x}_i[n] + \mathbf{F}_{RF,i}\mathbf{n}_i[n]$ is the unknown interference plus noise component after SI cancellation. Note that our BF design under partial CSIT proposed here is applicable only for flat fading Kronecker channel models (details of the channel model follows later). Considering the SI channel, as the distance between the transmit and receive arrays does not satisfy the far-field range condition, we need to employ the near-field model which has spherical wavefront, see example [57]. Further, we assume that at both nodes, we have available a deterministic least squares (LS) channel estimate, which can be parametrized as follows

$$\begin{aligned}\widehat{\mathbf{H}}_{LS} &= \mathbf{H} + \widetilde{\mathbf{H}}_{LS}, \\ \mathbf{H} &= \mathbf{C}_r^{1/2}\mathbf{H}_v\mathbf{C}_t^{1/2}.\end{aligned} \tag{3.33}$$

where each element of the estimation error matrix, $\widetilde{\mathbf{H}}_{LS}$ is distributed as circularly symmetric complex Gaussian random variable, $\widetilde{\mathbf{H}}_{LS} \sim \mathscr{CN}(\mathbf{0}, \tilde{\sigma}^2\mathbf{I})$ and also each element of $\mathbf{H}_v$ is distributed as $\sim \mathscr{CN}(0,1)$. Also, $\widetilde{\mathbf{H}}_{LS}$ is independent of $\mathbf{H}$. Note that throughout the thesis, wherever we consider partial CSIT, we start from a deterministic LS estimate. How can we arrive at this? This can be arrived at after an LS estimate of the DL channel from the received signals in the pilot transmission phase. Assuming TDD reciprocity, the DL channel can be estimated using UL pilots. During the pilot transmission, assuming that all the UEs use orthogonal pilot sequences (with $\mathbf{S}\mathbf{S}^H = \mathbf{I}_{N_r}$ under the condition that $\tau_p \geq N_r$)

$$\mathbf{Y}_p = \mathbf{H}^H\mathbf{S} + \mathbf{N}. \tag{3.34}$$

After doing an LS estimate

$$\mathbf{Y}_p\mathbf{S}^H = \mathbf{H}^H + \mathbf{N}\mathbf{S}^H, \tag{3.35}$$

where $\mathbf{N}\mathbf{S}^H$ has the same statistical distribution as $\mathbf{N}$ and $\mathbf{Y}_p\mathbf{S}^H = \widehat{\mathbf{H}}_{LS}^H$. The positive semidefinite matrices $\mathbf{C}_r, \mathbf{C}_t$ represent the Rx and Tx side covariance matrices respectively. Assuming that

the full covariance information is known at both the nodes, we can construct an MMSE channel estimate for $vec(\mathbf{H}) = (\mathbf{C}_t^{1/2} \otimes \mathbf{C}_r^{1/2}) vec(\mathbf{H}_v)$ as follows ($\widehat{\mathbf{H}}$ representing the MMSE estimate)

$$(\mathbf{C}_t \otimes \mathbf{C}_r)(\mathbf{C}_t \otimes \mathbf{C}_r + \widetilde{\sigma}^2 \mathbf{I})^{-1} vec(\widehat{\mathbf{H}}_{LS}) = vec(\widehat{\mathbf{H}}). \tag{3.36}$$

More detailed derivation of the LMMSE estimate under Kronecker channel model described above (3.36) appears in the Appendix A. However, one question here is that how would doing an MMSE estimate from the LS channel estimate perform compared to doing an MMSE estimate directly from the received signals in the pilot transmission phase (equation (3.34)). In fact, it can be shown that LS estimate is a sufficient estimate for MMSE estimate of the channel from (3.34). The proof for this appears in [61, Appendix C.2.1]. To simplify further, we consider the eigen decomposition of $\mathbf{C}_t = \mathbf{U}_t \mathbf{\Lambda}_t \mathbf{U}_t^H, \mathbf{C}_r = \mathbf{U}_r \mathbf{\Lambda}_r \mathbf{U}_r^H$. It is straightforward to show that $(\mathbf{C}_t \otimes \mathbf{C}_r + \widetilde{\sigma}^2 \mathbf{I})^{-1} = (\mathbf{U}_t \otimes \mathbf{U}_r)[\mathbf{\Lambda}_t \otimes \mathbf{\Lambda}_r + \widetilde{\sigma}^2 \mathbf{I}_{N_t} \otimes \mathbf{I}_{N_r}]^{-1} (\mathbf{U}_t^H \otimes \mathbf{U}_r^H)$. It follows from using the identity $(\mathbf{A} \otimes \mathbf{B})^{-1} = \mathbf{A}^{-1} \otimes \mathbf{B}^{-1}$, if $\mathbf{A}^{-1}, \mathbf{B}^{-1}$ exists. Further, we can simplify $(\mathbf{C}_t \otimes \mathbf{C}_r)(\mathbf{C}_t \otimes \mathbf{C}_r + \widetilde{\sigma}^2 \mathbf{I})^{-1} = (\mathbf{U}_t \otimes \mathbf{U}_r)(\mathbf{\Lambda}_{tr})(\mathbf{U}_t^H \otimes \mathbf{U}_r^H)$, where $\mathbf{\Lambda}_{tr} = (\mathbf{\Lambda}_t \otimes \mathbf{\Lambda}_r)[(\mathbf{\Lambda}_t \otimes \mathbf{\Lambda}_r) + \widetilde{\sigma}^2 \mathbf{I}_{N_r N_t}]^{-1}$. We define $\mathbf{\Lambda}'_{tr} = [(\mathbf{\Lambda}_t \otimes \mathbf{\Lambda}_r) + \widetilde{\sigma}^2 \mathbf{I}_{N_r N_t}]^{-1}$. $\mathbf{\Lambda}_{tr} = \sum_{i=1}^{N_t} (\mathbf{\Lambda}_t)_{i,i} (\mathbf{e}_i \mathbf{e}_i^H) \otimes (\mathbf{\Lambda}_r \mathbf{\Lambda}'_{tr,i})$. We denote $\mathbf{\Lambda}'_{tr,i}$ or $\mathbf{\Lambda}_{tr,i}$ as the diagonal matrix which forms i^{th} $N_r \times N_r$ block of $\mathbf{\Lambda}'_{tr}$ or $\mathbf{\Lambda}_{tr}$. Here $(\mathbf{A})_{i,i}$ represents the i^{th} diagonal element of any matrix $\mathbf{A}$. Further we can write

$$\begin{aligned}
\widehat{\mathbf{H}} &= \sum_{i=1}^{N_t} \widehat{\mathbf{C}}_{r,i} \mathbf{H}_{LS} \widehat{\mathbf{C}}_{t,i}, \\
\widehat{\mathbf{C}}_{t,i} &= \mathbf{U}_t \widehat{\mathbf{\Lambda}}_{t,i} \mathbf{U}_t^H, \\
\widehat{\mathbf{C}}_{r,i} &= \mathbf{U}_r \widehat{\mathbf{\Lambda}}_{r,i} \mathbf{U}_r^H, \\
\widehat{\mathbf{\Lambda}}_{t,i} &= (\mathbf{\Lambda}_t)_{i,i} (\mathbf{e}_i \mathbf{e}_i^H), \\
\widehat{\mathbf{\Lambda}}_{r,i} &= \mathbf{\Lambda}_r \mathbf{\Lambda}'_{tr,i}.
\end{aligned} \tag{3.37}$$

The estimation error covariance matrix can be obtained as

$$(\mathbf{C}_t \otimes \mathbf{C}_r) - (\mathbf{C}_t \otimes \mathbf{C}_r)(\mathbf{C}_t \otimes \mathbf{C}_r + \widetilde{\sigma}^2 \mathbf{I})^{-1} (\mathbf{C}_t \otimes \mathbf{C}_r), \tag{3.38}$$

which gets simplified as $\sum_{i=1}^{N_t} \widetilde{\mathbf{C}}_{t,i} \otimes \widetilde{\mathbf{C}}_{r,i}$, where $\widetilde{\mathbf{C}}_{t,i} = (\mathbf{\Lambda}_t)_{i,i} \mathbf{U}_t (\mathbf{e}_i \mathbf{e}_i^H) \mathbf{U}_t^H, \widetilde{\mathbf{C}}_r = \mathbf{U}_r (\mathbf{\Lambda}_r (\mathbf{I}_{N_r} - \mathbf{\Lambda}_{tr,i})) \mathbf{U}_r^H$. Thus we finally obtain the estimation error as, $\widetilde{\mathbf{H}} = \sum_{i=1}^{N_t} \widetilde{\mathbf{C}}_{r,i} \widetilde{\mathbf{H}}_v \widetilde{\mathbf{C}}_{t,i}$ and $\mathbf{H} = \widehat{\mathbf{H}} + \widetilde{\mathbf{H}}$. In the massive MIMO limit, where $N_r, N_t \to \infty$, we get convergence for any terms of the form $\mathbf{HQH}^H$ as below [62]. This result gets used extensively in the following sections.

$$\mathbf{HQH}^H \xrightarrow[a.s]{M \to \infty} \mathrm{E}_{\mathbf{H}|\widehat{\mathbf{H}}} \mathbf{HQH}^H = \widehat{\mathbf{H}} \mathbf{Q} \widehat{\mathbf{H}}^H + \mathrm{tr}\{\mathbf{Q} \widetilde{\mathbf{C}}_t\} \widetilde{\mathbf{C}}_r. \tag{3.39}$$

3.4.1 EWSR maximization through alternating minorization

In this section, consider the optimization of the two-stage BF/hybrid combiner design using WSR maximization of the Multi-cell MU-MIMO system. Since the CSIT is imperfect, we consider here the optimization of the ergodic capacity. First, the WSR is averaged over the channels given a particular channel estimate, which leads to a cost function in the MaMIMO limit and it is denoted as Expected Signal and Interference Power WSR (ESIP-WSR). ESIP-WSR is optimized to

compute the BFs and then it is again averaged over the channel estimates, to evaluate the final ergodic WSR. The results in this section are discussed in our paper [63].

$$
\begin{aligned}
[\mathbf{V}\,\mathbf{G}\,\mathbf{F}_{RF}\,\mathbf{F}_{BB}] &= \arg \max_{\substack{\mathbf{V},\mathbf{G},\\ \mathbf{F}_{RF},\mathbf{F}_{BB}}} EWSR(\mathbf{G},\mathbf{V},\mathbf{F}_{RF},\mathbf{F}_{BB}) \\
&= \arg\max_{\mathbf{V},\mathbf{G}} \sum_{i=1}^{2}\sum_{n=1}^{N_s} \mathrm{E}_{\mathbf{H}|\widehat{\mathbf{H}}}(u_i \ln\det\left(\mathbf{R}_{\bar{i}}^{-1}[n]\mathbf{R}_i[n]\right)) \\
&= \arg\max_{\mathbf{V},\mathbf{G}} \sum_{i=1}^{2}\sum_{n=1}^{N_s} (u_i[\ln\det\left(\mathrm{E}_{\mathbf{H}|\widehat{\mathbf{H}}}\mathbf{R}_i[n]\right) - \ln\det\left(\mathrm{E}_{\mathbf{H}|\widehat{\mathbf{H}}}\mathbf{R}_{\bar{i}}[n]\right))] \\
&= ESIP - WSR\,(\mathbf{G},\mathbf{V},\mathbf{F}_{RF},\mathbf{F}_{BB}),
\end{aligned} \tag{3.40}
$$

where the u_i are the rate weights (used to denote priorities assigned to users, refer Section 2.1.1 for more details), $\mathbf{G}$ represents the collection of digital BFs $\mathbf{G}_i[n]$, $\mathbf{V}$ the collection of analog BFs $\mathbf{V}^i$. We remark that in the massive MIMO limit, the ESIP-WSR represents an upper bound as is shown in [64] (also in Chapter 7), where the channels are MISO. However, to extend the same for the MIMO case is straightforward and involve the same argument that the interference power converge to its expectation and further using the Jensen's inequality. At the receiver, we apply a hybrid combiner with analog BF denoted by $\mathbf{F}_{RF,i}$ of size $M_r^i \times N_r^i$, where M_r^i represents the number of RF chains at the Rx side. $\mathbf{F}_{BB,i}$ represent the baseband digital combiner of size $d_j \times M_r^i$. The covariance matrix of $\mathbf{v}_i[n]$, $\mathbf{R}_{\bar{i}}[n]$ can be approximated under $k_i \ll 1, l_i \ll 1$ as follows [59]

$$
\begin{aligned}
\mathbf{R}_{\bar{i}}[n] =\ & k_j\mathbf{F}_{RF,i}\mathbf{H}_{i,j}[n]\,\mathrm{diag}\,(\mathbf{Q}_j[n])\mathbf{H}_{i,j}^H[n]\mathbf{F}_{RF,i}^H + k_i\mathbf{F}_{RF,i}\mathbf{H}_{i,i}[n]\,\mathrm{diag}\,(\mathbf{Q}_i[n])\mathbf{H}_{i,i}^H[n]\mathbf{F}_{RF,i}^H \\
&+ l_i\,\mathrm{diag}\,(\mathbf{F}_{RF,i}\mathbf{H}_{i,j}[n]\mathbf{Q}_j[n]\mathbf{H}_{i,j}^H[n]\mathbf{F}_{RF,i}^H) + l_i\,\mathrm{diag}\,(\mathbf{F}_{RF,i}\mathbf{H}_{i,i}[n]\mathbf{Q}_i[n]\mathbf{H}_{i,i}^H[n]\mathbf{F}_{RF,i}^H) \\
&+ \mathbf{F}_{RF,i}\widetilde{\mathbf{H}}_{i,i}[n]\mathbf{Q}_i[n]\widetilde{\mathbf{H}}_{i,i}[n]^H\mathbf{F}_{RF,i}^H + \mathbf{F}_{RF,i}\mathbf{F}_{RF,i}^H, \\
\text{Also, } \mathbf{R}_i[n] =\ & \mathbf{R}_{\bar{i}}[n] + \mathbf{F}_{RF,i}\mathbf{H}_{i,j}[n]\mathbf{Q}_j[n]\mathbf{H}_{i,j}^H[n]\mathbf{F}_{RF,i}^H,
\end{aligned} \tag{3.41}
$$

where $\mathbf{R}_i[n]$ is the signal plus interference plus noise covariance matrix. For notational simplicity, we define $\widehat{\mathbf{H}}_{i,j}[n]\mathbf{Q}_j[n]\widehat{\mathbf{H}}_{i,j}^H[n] = \widehat{\mathbf{\Theta}}_{i,j}[n]$, which can be interpreted as the effective Rx signal covariance matrix before the analog combiner given a particular channel estimate. Also,

$$
\begin{aligned}
\widehat{\mathbf{H}}_{i,j}[n]\,\mathrm{diag}\,(\mathbf{Q}_j[n])\widehat{\mathbf{H}}_{i,j}^H[n] &= \widehat{\mathbf{\Psi}}_{i,j}[n], \\
\widehat{\mathbf{\Theta}}_{i,j}[n] + \mathrm{tr}\{\mathbf{Q}_j[n]\widetilde{\mathbf{C}}_{t,i,j}\}\widetilde{\mathbf{C}}_{r,i,j} &= \mathbf{\Theta}_{i,j}[n], \\
\widehat{\mathbf{\Psi}}_{i,j}[n] + \mathrm{tr}\{\mathrm{diag}\,(\mathbf{Q}_j)\widetilde{\mathbf{C}}_{t,i,j}\}\widetilde{\mathbf{C}}_{r,i,j} &= \mathbf{\Psi}_{i,j}[n].
\end{aligned} \tag{3.42}
$$

Further, we obtain the expected signal and interference plus noise power ($\overline{\mathbf{R}}_i[n]$) and expected interference plus noise power ($\overline{\mathbf{R}}_{\bar{i}}[n]$) as

$$
\begin{aligned}
\overline{\mathbf{R}}_{\bar{i}}[n] =\ & k_j\mathbf{F}_{RF,i}\mathbf{\Psi}_{i,j}[n]\mathbf{F}_{RF,i}^H + k_i\mathbf{F}_{RF,i}\mathbf{\Psi}_{i,i}[n]\mathbf{F}_{RF,i}^H \\
&+ l_i\,\mathrm{diag}\,(\mathbf{F}_{RF,i}\mathbf{\Theta}_{i,j}[n]\mathbf{F}_{RF,i}^H) + l_i\,\mathrm{diag}\,(\mathbf{F}_{RF,i}\mathbf{\Theta}_{i,i}[n]\mathbf{F}_{RF,i}^H) \\
&+ \mathrm{tr}\{\mathbf{Q}_i[n]\widetilde{\mathbf{C}}_{t,i,i}\}\mathbf{F}_{RF,i}\widetilde{\mathbf{C}}_{r,i,i}\mathbf{F}_{RF,i}^H + \mathbf{F}_{RF,i}\mathbf{F}_{RF,i}^H, \\
\text{Also, } \overline{\mathbf{R}}_i[n] =\ & \overline{\mathbf{R}}_{\bar{i}}[n] + \mathbf{F}_{RF,i}\mathbf{\Theta}_{i,j}[n]\mathbf{F}_{RF,i}^H.
\end{aligned} \tag{3.43}
$$

Direct maximization of (3.40), however, requires a joint optimization over the four matrix variables $(\mathbf{V},\mathbf{G},\mathbf{F}_{RF},\mathbf{F}_{BB})$. Unfortunately, finding a global optimum solution for similarly constrained optimization is found to be intractable. So we decouple the joint transmitter-receiver optimization and focus on the design of the Rx combiners first. We assume that the node i applies the

frequency selective hybrid combiner $\mathbf{F}_{BB,i}[n]$ at the output of the Rx RF chains and after the IFFT, to estimate the signal transmitted from node j. The analog combiner $\mathbf{F}_{RF,i}$ serves to reduce the SI component from the received signal, while the digital combiner $\mathbf{F}_{BB,i}$ decouples the streams ($\mathbf{d}_j$) intended for user i from j.

$$\widehat{\mathbf{d}}_j[n] = \mathbf{F}_{BB,i}[n]\mathbf{y}_i[n] + \mathbf{F}_{BB,i}[n]\mathbf{v}_i[n]. \tag{3.44}$$

At the Rx side, maximizing the WSR is equivalent to minimizing the weighted MSE with the MSE weights being chosen as $\mathbf{W}_i[n] = \frac{u_i}{\ln 2}\mathbf{R}_{\tilde{\mathbf{d}}_j\tilde{\mathbf{d}}_j}[n]^{-1}$ [9, 54]. However, with partial CSIT, we chose to minimize the expected weighted MSE (EWSMSE) for the Rx side digital combiner. We can write the error covariance matrix for the detection of $\mathbf{d}_j$ at node i as

$$\begin{aligned}\mathbf{R}_{\tilde{\mathbf{d}}_j\tilde{\mathbf{d}}_j}[n] &= \mathrm{E}_{\mathbf{H}|\widehat{\mathbf{H}}}\{(\widehat{\mathbf{d}}_j[n]-\mathbf{d}_j[n])(\widehat{\mathbf{d}}_j[n]-\mathbf{d}_j[n])^H\}\\ &= (\mathbf{F}_i[n]\widehat{\mathbf{H}}_{i,j}[n]\mathbf{Q}_j[n]\widehat{\mathbf{H}}_{i,j}[n]^H\mathbf{F}_i[n]^H + \mathrm{tr}\{\mathbf{Q}_j[n]\widetilde{\mathbf{C}}_{t,i,j}\}\mathbf{F}_i[n]\widetilde{\mathbf{C}}_{r,i,j}\mathbf{F}_i[n]^H\\ &\quad - \mathbf{F}_i[n]\widehat{\mathbf{H}}_{i,j}[n]\mathbf{V}^j\mathbf{G}_j[n] - \mathbf{G}_j[n]^H\mathbf{V}^j\widehat{\mathbf{H}}_{i,j}[n]^H\mathbf{F}_i[n]^H + \overline{\boldsymbol{\Sigma}}_{\tilde{i}}[n].\end{aligned} \tag{3.45}$$

The MMSE receive combiner at the baseband side can be alternatively optimized, $\forall n$, as follows

$$\begin{aligned}\mathbf{F}_{BB,i}[n] &= \arg\min_{\mathbf{F}_{BB,i}[n]} \mathrm{tr}\{\mathbf{R}_{\tilde{\mathbf{d}}_i\tilde{\mathbf{d}}_i}[n]\},\\ &= \mathbf{G}_j^H[n]\mathbf{V}^{jH}\widehat{\mathbf{H}}_{i,j}^H[n]\mathbf{F}_{RF,i}^H\overline{\mathbf{R}}_i[n]^{-1}.\end{aligned} \tag{3.46}$$

Optimizing the digital BF in (3.46) above can be done independently across different subcarriers, obviously. Further, to optimize the analog combiner, we directly optimize the ESIP-WSR. We make use of certain results on matrix differentiation. It was shown in [65] that $\frac{\partial \ln\det(\mathbf{A}+\mathbf{BXC})}{\partial \mathbf{X}} = \left[\mathbf{C}(\mathbf{A}+\mathbf{BXC})^{-1}\mathbf{B}\right]^T$. Taking the gradient of (3.40) w.r.t. $\mathbf{F}_{RF,i}$

$$\begin{aligned}\sum_{n=1}^{N_s}\mathbf{R}_i^{-1}[n]\mathbf{F}_{RF,i}(\boldsymbol{\Theta}_{i,j}[n]) &= \sum_{n=1}^{N_s}(\mathbf{R}_{\tilde{i}}^{-1}[n]-\mathbf{R}_i^{-1}[n])\mathbf{F}_{RF,i}\Big(k_j\boldsymbol{\Psi}_{i,j}[n]+k_i\boldsymbol{\Psi}_{i,i}[n]\Big)\\ &\quad + l_i\,\mathrm{diag}(\mathbf{R}_{\tilde{i}}^{-1}[n]-\mathbf{R}_i^{-1}[n])\mathbf{F}_{RF,i}\Big(\boldsymbol{\Theta}_{i,j}[n]+\boldsymbol{\Theta}_{i,i}[n])\\ &\quad + \mathrm{tr}\{\mathbf{Q}_i[n]\widetilde{\mathbf{C}}_{t,i,i}\}(\mathbf{R}_{\tilde{i}}^{-1}[n]-\mathbf{R}_i^{-1}[n])\mathbf{F}_{RF,i}(\widetilde{\mathbf{C}}_{r,i,i}) + (\mathbf{R}_{\tilde{i}}^{-1}[n]-\mathbf{R}_i^{-1}[n])\mathbf{F}_{RF,i}\Big),\end{aligned} \tag{3.47}$$

Vectorizing both sides, we obtain

$$\begin{aligned}\sum_{n=1}^{N_s}\Big((\boldsymbol{\Theta}_{i,j}[n])^T\otimes\mathbf{R}_i^{-1}[n]\Big)vec(\mathbf{F}_{RF,i}) &\overset{(a)}{=} \sum_{n=1}^{N_s}\Big[\Big(k_j\boldsymbol{\Psi}_{i,j}[n]+k_i\boldsymbol{\Psi}_{i,i}[n]\Big)^T\otimes(\mathbf{R}_{\tilde{i}}^{-1}[n]-\mathbf{R}_i^{-1}[n])\\ &\quad + l_i(\boldsymbol{\Theta}_{i,j}[n]+\boldsymbol{\Theta}_{i,i}[n])^T\otimes\mathrm{diag}(\mathbf{R}_{\tilde{i}}^{-1}[n]-\mathbf{R}_i^{-1}[n])\\ &\quad + \mathrm{tr}\{\mathbf{Q}_i[n]\widetilde{\mathbf{C}}_{t,i,i}\}[\widetilde{\mathbf{C}}_{r,i,i}\otimes(\mathbf{R}_{\tilde{i}}^{-1}[n]-\mathbf{R}_i^{-1}[n])]\\ &\quad + \mathbf{I}_{N_r^i}\otimes(\mathbf{R}_{\tilde{i}}^{-1}[n]-\mathbf{R}_i^{-1}[n])\Big]vec(\mathbf{F}_{RF,i})\end{aligned} \tag{3.48}$$

In (a), we use the result $vec(\mathbf{AXB}) = (\mathbf{B}^T \otimes \mathbf{A})\, vec(\mathbf{X})$ from [31]. Further, this leads to a generalized eigenvector solution for the analog combiner

$$
\begin{aligned}
vec(\mathbf{F}_{RF,i}) &= \mathbf{V}_{max}(\widehat{\mathbf{B}}_i, \mathbf{A}_i), \\
\widehat{\mathbf{B}}_i &= \sum_{n=1}^{N_s} (\mathbf{\Theta}_{i,j}[n])^T \otimes \mathbf{R}_i^{-1}[n], \\
\widehat{\mathbf{A}}_i &= \sum_{n=1}^{N_s} \Big[\Big(k_j \mathbf{\Psi}_{i,j}[n] + k_i \mathbf{\Psi}_{i,i}[n] \Big)^T \otimes (\mathbf{R}_{\bar{i}}^{-1}[n] - \mathbf{R}_i^{-1}[n]) \\
&\quad + l_i \Big(\mathbf{\Theta}_{i,j}[n] + \mathbf{\Theta}_{i,i}[n] \Big)^T \otimes \operatorname{diag}(\mathbf{R}_{\bar{i}}^{-1}[n] - \mathbf{R}_i^{-1}[n] \\
&\quad + \operatorname{tr}\{\mathbf{Q}_i[n] \widetilde{\mathbf{C}}_{t,i,i}\} [\widetilde{\mathbf{C}}_{r,i,i} \otimes (\mathbf{R}_{\bar{i}}^{-1}[n] - \mathbf{R}_i^{-1}[n])] + \mathbf{I}_{N_r^i} \otimes (\mathbf{R}_{\bar{i}}^{-1}[n] - \mathbf{R}_i^{-1}[n])) \Big]
\end{aligned} \tag{3.49}
$$

3.4.2 Two-stage transmit BF design

We define the following Lemma below which proves the concavity of a part of the EWSR (3.40).

Lemma 1. *For each* $i \in 1,2, n \in 1,..,N_s$, $f_i\left(\mathbf{Q}_j[n], \mathbf{Q}_{\bar{j}}[n]\right) = \ln\det\left(\overline{\mathbf{R}}_{\bar{i}}^{-1}[n]\overline{\mathbf{R}}_i[n]\right)$ *is concave w.r.t* $\mathbf{Q}_j[n]$, *where* $\mathbf{Q}_j[n]$ *is a positive semidefinite matrix.*

Proof: Using the technique from [65, Th. 2], the concavity of $f_i\left(\mathbf{Q}_j[n], \mathbf{Q}_{\bar{j}}[n]\right)$ w.r.t $\mathbf{Q}_j[n]$ can be proved by showing that $\widetilde{f}_i(t) = f_i\left(\mathbf{X}_j + t\mathbf{Y}_j, \mathbf{Q}_{\bar{j}}[n]\right)$ is concave w.r.t $t \in [0,1]$, where $\mathbf{X}_i$ is positive semidefinite and $\mathbf{Y}_i$ being Hermitian. The derivative of $\widetilde{f}_i(t)$ w.r.t t can be written as

$$
\begin{aligned}
\frac{\partial}{\partial t}\widetilde{f}_i(t) = \operatorname{tr}\{\overline{\mathbf{R}}_i^{-1}[n]\Big(&\frac{\partial \overline{\mathbf{R}}_{\bar{i}}[n]}{\partial t} + \mathbf{F}_{RF,i}\widehat{\mathbf{H}}_{i,j}[n]\mathbf{Y}_j\widehat{\mathbf{H}}_{i,j}^H[n]\mathbf{F}_{RF,i}^H\Big) \\
&+ \operatorname{tr}\{\mathbf{Y}_j\widetilde{\mathbf{C}}_{t,i,j}\}\mathbf{F}_{RF,i}\widetilde{\mathbf{C}}_{r,i,j}\mathbf{F}_{RF,i}^H\Big) - \overline{\mathbf{R}}_{\bar{i}}^{-1}[n]\frac{\partial \overline{\mathbf{R}}_{\bar{i}}}{\partial t}\}
\end{aligned} \tag{3.50}
$$

where

$$
\begin{aligned}
\frac{\partial \mathbf{R}_{\bar{i}}[n]}{\partial t} &= k_j\mathbf{F}_{RF,i}\widehat{\mathbf{H}}_{i,j}[n]\operatorname{diag}(\mathbf{Y}_j[n])\widehat{\mathbf{H}}_{i,j}^H[n]\mathbf{F}_{RF,i}^H \\
&+ k_j\operatorname{tr}\{\operatorname{diag}(\mathbf{Y}_j[n])\widetilde{\mathbf{C}}_{t,i,j}\}\mathbf{F}_{RF,i}\widetilde{\mathbf{C}}_{r,i,j}\mathbf{F}_{RF,i}^H + l_i\operatorname{diag}(\mathbf{F}_{RF,i}\widehat{\mathbf{H}}_{i,j}[n]\mathbf{Y}_j[n]\widehat{\mathbf{H}}_{i,j}^H[n]\mathbf{F}_{RF,i}^H) \\
&+ l_i\operatorname{tr}\{\mathbf{Y}_j[n]\widetilde{\mathbf{C}}_{t,i,j}\}\operatorname{diag}(\mathbf{F}_{RF,i}\widetilde{\mathbf{C}}_{i,j}\mathbf{F}_{RF,i}^H) \text{ does not depend on } t.
\end{aligned} \tag{3.51}
$$

Further

$$
\frac{\partial^2}{\partial t^2}\widetilde{f}_i(t) = \operatorname{tr}\{-\overline{\mathbf{R}}_i^{-1}[n]\Big(\frac{\partial \overline{\mathbf{R}}_{\bar{i}}[n]}{\partial t} + \mathbf{N}_i\Big)\overline{\mathbf{R}}_i^{-1}[n]\Big(\frac{\partial \overline{\mathbf{R}}_{\bar{i}}[n]}{\partial t} + \mathbf{N}_i\Big) + \overline{\mathbf{R}}_{\bar{i}}^{-1}[n]\frac{\partial \overline{\mathbf{R}}_{\bar{i}}[n]}{\partial t}\overline{\mathbf{R}}_{\bar{i}}^{-1}[n]\frac{\partial \overline{\mathbf{R}}_{\bar{i}}[n]}{\partial t}\} \tag{3.52}
$$

where $\mathbf{N}_i = \mathbf{F}_{RF,i}\widehat{\mathbf{H}}_{i,j}\mathbf{Y}_j\widehat{\mathbf{H}}_{i,j}^H[n]\mathbf{F}_{RF,i}^H + \operatorname{tr}\{\mathbf{Y}_j\widetilde{\mathbf{C}}_{t,i,j}\}\mathbf{F}_{RF,i}\widetilde{\mathbf{C}}_{r,i,j}\mathbf{F}_{RF,i}^H$. Since we assume that $k_i, l_i \ll 1$, the second term inside the trace in (3.52) will contain quadratic terms in k_i or l_i and thus becomes negligible. Further, we can show similar as in [65, Th. 2] that the first term in (3.52) is negative and thus we can conclude that $\widetilde{f}_i(t)$ is concave. Here ends the proof.

Consider the dependence of EWSR on $\mathbf{Q}_j[n]$ alone.

$$
\begin{aligned}
EWSR &= u_i \ln\det\left(\overline{\mathbf{R}}_{\bar{i}}^{-1}[n]\overline{\mathbf{R}}_i[n]\right) + EWSR_{\bar{i}}[n] + \sum_{m=1, m\neq n}^{N_s} EWSR_i[m], \text{ where} \\
EWSR_{\bar{i}}[n] &= u_j \ln\det\left(\overline{\mathbf{R}}_{\bar{j}}^{-1}[n]\overline{\mathbf{R}}_j[n]\right),\ j \neq i
\end{aligned} \tag{3.53}
$$

From Lemma 1, we can see that the first term in the above summation is a concave function in $\mathbf{Q}_j[n]$. However, the rest of the terms are convex due to the dependency of $\mathbf{Q}_j[n]$ through the interference terms. To solve this non-convex problem, we further consider a difference of convex (DC) function approach [30]. DC approach linearizes the convex part through a first order Taylor series expansion (around $\widehat{\mathbf{Q}}_j[n]$ and $\widehat{\mathbf{R}}_i[n]$ represents the corresponding $\mathbf{R}_i[n]$) as below

$$
\begin{aligned}
\underline{EWSR}_{\bar{i}}(\mathbf{Q}_j[n],\widehat{\mathbf{Q}}[n]) &= EWSR_{\bar{i}}(\widehat{\mathbf{Q}}_j[n],\widehat{\mathbf{Q}}[n]) - \mathrm{tr}\{(\mathbf{Q}_j[n]-\widehat{\mathbf{Q}}_j[n])\widehat{\mathbf{A}}_j[n]\},\\
\widehat{\mathbf{A}}_j[n] &= -\left.\frac{\partial EWSR_{\bar{i}}\left(\mathbf{Q}_j[n],\widehat{\mathbf{Q}}[n]\right)}{\partial \mathbf{Q}_j[n]}\right|_{\widehat{\mathbf{Q}}_j[n]}\\
&\overset{(a)}{=} u_j k_j \,\mathrm{diag}\,(\widehat{\mathbf{H}}_{j,j}^H[n]\mathbf{F}_{RF,j}^H(\widehat{\mathbf{R}}_{\bar{j}}^{-1}[n]-\widehat{\mathbf{R}}_j^{-1}[n])\mathbf{F}_{RF,j}\widehat{\mathbf{H}}_{j,j}[n])\\
&+ l_j u_j \widehat{\mathbf{H}}_{j,j}^H[n]\mathbf{F}_{RF,j}^H \mathrm{diag}\,(\widehat{\mathbf{R}}_{\bar{j}}^{-1}[n]-\widehat{\mathbf{R}}_j^{-1}[n])\mathbf{F}_{RF,j}\widehat{\mathbf{H}}_{j,j}[n]\\
&+ u_j l_j \mathrm{tr}\{\mathrm{diag}\,(\mathbf{F}_{RF,j}\widetilde{\mathbf{C}}_{r,j,j}\mathbf{F}_{RF,j}^H)(\widehat{\mathbf{R}}_{\bar{j}}^{-1}[n]-\widehat{\mathbf{R}}_j^{-1}[n]\}\widetilde{\mathbf{C}}_{t,j,j}\\
&+ u_j k_j \mathrm{tr}\{(\mathbf{F}_{RF,j}\widetilde{\mathbf{C}}_{r,j,j}\mathbf{F}_{RF,j}^H)(\widehat{\mathbf{R}}_{\bar{j}}^{-1}[n]-\widehat{\mathbf{R}}_j^{-1}[n])\}\,\mathrm{diag}\,(\widetilde{\mathbf{C}}_{t,j,j})\\
&+ u_j \mathrm{tr}\{\mathbf{F}_{RF,j}\widetilde{\mathbf{C}}_{r,j,j}\mathbf{F}_{RF,j}^H(\widehat{\mathbf{R}}_{\bar{j}}^{-1}[n]-\widehat{\mathbf{R}}_j^{-1}[n])\}\widetilde{\mathbf{C}}_{t,j,j}.
\end{aligned}
\tag{3.54}
$$

In the above equation, for the trace term, we made use of the gradient result derived in the Appendix, $\frac{\partial \ln\det\mathbf{Y}}{\partial\mathbf{X}} = [\mathbf{D}^T\mathrm{tr}\{\mathbf{B}^T\mathbf{Y}^{-1}\}]$, where, $\mathbf{Y}=\mathrm{tr}\{\mathbf{XD}\}\mathbf{B}$. The Taylor series expansion is done around the point $\widehat{\mathbf{Q}}_j[n]$ (which represent the computed previous iteration values) and the corresponding $\overline{\mathbf{R}}_i[n]$ is $\widehat{\mathbf{R}}_i[n]$. Then, dropping constant terms, reparameterizing the $\mathbf{Q}_j[n]$ as in (3.43), performing this linearization for all users, and augmenting the EWSR cost function with the Tx power constraints, we get the Lagrangian (7.17) which gets maximized alternatingly [27] between digital and analog BF.

$$
\mathscr{L}(\mathbf{V},\mathbf{G},\mathbf{\Lambda}) = \sum_{i=1}^{2}\lambda_i P_i + \sum_{i=1}^{2}\sum_{n=1}^{N_s} u_i \ln\det\left(\overline{\mathbf{R}}_{\bar{i}}^{-1}[n]\overline{\mathbf{R}}_i[n]\right) - \mathrm{tr}\{\mathbf{G}_i^H[n](\mathbf{V}^{iH}(\widehat{\mathbf{A}}_i[n]+\lambda_{b_k}\mathbf{I})\mathbf{V}^i)\mathbf{G}_i[n]\}.
\tag{3.55}
$$

In Appendix E, we derive the gradient expressions when there are terms of the form $\ln\det(\mathbf{Y}+\mathbf{F}(\mathbf{X}))$ where $\mathbf{Y}=\mathbf{A}\,\mathrm{diag}\,(\mathbf{CXD})\mathbf{B}+\mathbf{F}(\mathbf{X})$. Using this result, we take the derivative of (7.17) w.r.t the digital BF $\mathbf{G}_j$ which leads to

$$
\begin{aligned}
&\mathbf{V}^{jH}\widehat{\mathbf{H}}_{i,j}[n]^H\mathbf{F}_{RF,i}^H(\widehat{\mathbf{R}}_i^{-1}[n]+l_i\,\mathrm{diag}\,(\widehat{\mathbf{R}}_i^{-1}[n]-\widehat{\mathbf{R}}_{\bar{i}}^{-1}[n]))\mathbf{F}_{RF,i}\widehat{\mathbf{H}}_{i,j}[n]\mathbf{V}^j\mathbf{G}_j[n]\\
&+k_j\mathbf{V}^{jH}\,\mathrm{diag}\,(\widehat{\mathbf{H}}_{i,j}[n]^H\mathbf{F}_{RF,i}^H(\widehat{\mathbf{R}}_i^{-1}[n]-\widehat{\mathbf{R}}_{\bar{i}}^{-1}[n])\mathbf{F}_{RF,i}\widehat{\mathbf{H}}_{i,j}[n])\mathbf{V}^j\mathbf{G}_j[n]\\
&+\mathbf{V}^{jH}(\mathrm{tr}\{\mathbf{F}_{RF,i}^H\widehat{\mathbf{R}}_i^{-1}[n]\widetilde{\mathbf{C}}_{r,i,j}\}\widetilde{\mathbf{C}}_{t,i,j}+l_i\mathrm{tr}\{\mathrm{diag}\,(\mathbf{F}_{RF,i}\widetilde{\mathbf{C}}_{r,i,j}\mathbf{F}_{RF,i}^H)(\widehat{\mathbf{R}}_{\bar{i}}^{-1}[n]-\widehat{\mathbf{R}}_i^{-1}[n])\}\widetilde{\mathbf{C}}_{t,i,j}\\
&+k_j\mathrm{tr}\{(\mathbf{F}_{RF,i}\widetilde{\mathbf{C}}_{r,i,j}\mathbf{F}_{RF,i}^H)(\widehat{\mathbf{R}}_{\bar{i}}^{-1}[n]-\widehat{\mathbf{R}}_i^{-1}[n])\}\,\mathrm{diag}\,(\widetilde{\mathbf{C}}_{t,i,j}))\mathbf{V}^{jH}\mathbf{G}_j[n]\\
&=\mathbf{V}^{jH}\widehat{\mathbf{A}}_j[n]\mathbf{V}^j\mathbf{G}_j[n]
\end{aligned}
\tag{3.56}
$$

This can be interpreted as the dominant generalized eigenvectors solution for the digital BF

$$
\mathbf{G}_j[n] = \mathbf{V}_{1:d_j}(\mathbf{V}^{jH}\widehat{\mathbf{B}}_j[n]\mathbf{V}^j, \mathbf{V}^{jH}(\widehat{\mathbf{A}}_j[n]+\widehat{\mathbf{C}}_j[n]+\lambda_j\mathbf{I})\mathbf{V}^j),
\tag{3.57}
$$

where

$$
\begin{aligned}
\widehat{\mathbf{B}}_j[n] &= \widehat{\mathbf{H}}_{i,j}[n]^H \mathbf{F}_{RF,i}^H \widehat{\mathbf{R}}_i^{-1}[n] \mathbf{F}_{RF,i} \widehat{\mathbf{H}}_{i,j}[n] + \mathrm{tr}\{\mathbf{F}_{RF,i}^H \widehat{\mathbf{R}}_i^{-1}[n] \widetilde{\mathbf{C}}_{r,i,j}\} \widetilde{\mathbf{C}}_{t,i,j}. \\
\widehat{\mathbf{C}}_j[n] &= -\widehat{\mathbf{H}}_{i,j}[n]^H \mathbf{F}_{RF,i}^H (l_i \,\mathrm{diag}\,(\widehat{\mathbf{R}}_i^{-1} - \widehat{\mathbf{R}}_{\bar{i}}^{-1})) \mathbf{F}_{RF,i} \widehat{\mathbf{H}}_{i,j} \\
&\quad + k_j \,\mathrm{diag}\,(\widehat{\mathbf{H}}_{i,j}[n]^H \mathbf{F}_{RF,i}^H (\widehat{\mathbf{R}}_i^{-1}[n] - \widehat{\mathbf{R}}_{\bar{i}}^{-1}[n]) \mathbf{F}_{RF,i} \widehat{\mathbf{H}}_{i,j}[n]) \\
&\quad + l_i \mathrm{tr}\{\mathrm{diag}\,(\mathbf{F}_{RF,i} \widetilde{\mathbf{C}}_{r,i,j} \mathbf{F}_{RF,i}^H)(\widehat{\mathbf{R}}_{\bar{i}}^{-1}[n] - \widehat{\mathbf{R}}_i^{-1}[n])\} \widetilde{\mathbf{C}}_{t,i,j} \\
&\quad + k_j \mathrm{tr}\{(\mathbf{F}_{RF,i} \widetilde{\mathbf{C}}_{r,i,j} \mathbf{F}_{RF,i}^H)(\widehat{\mathbf{R}}_{\bar{i}}^{-1}[n] - \widehat{\mathbf{R}}_i^{-1}[n])\} \,\mathrm{diag}\,(\widetilde{\mathbf{C}}_{t,i,j}).
\end{aligned} \tag{3.58}
$$

Further considering the derivative of (7.17) w.r.t the analog BF $\mathbf{V}^j$, we get

$$
(\widehat{\mathbf{B}}_j[n] - \widehat{\mathbf{C}}_j[n]) \mathbf{V}^j \mathbf{G}_j[n] \mathbf{G}_j^H[n] = (\widehat{\mathbf{A}}_j[n] + \lambda_j \mathbf{I}) \mathbf{V}^j \mathbf{G}_j[n] \mathbf{G}_j^H[n]. \tag{3.59}
$$

Further utilizing the result $vec(\mathbf{AXB}) = (\mathbf{B}^T \otimes \mathbf{A})\, vec(\mathbf{X})$ [31], we get

$$
\sum_{n=1}^{N_s} ((\mathbf{G}_j[n] \mathbf{G}_j^H[n])^T \otimes \widehat{\mathbf{B}}_j[n])\, vec(\mathbf{V}^j) = \sum_{n=1}^{N_s} ((\mathbf{G}_j[n] \mathbf{G}_j^H[n])^T \otimes \widehat{\mathbf{E}}_j[n])\, vec(\mathbf{V}^j). \tag{3.60}
$$

where we define $\widehat{\mathbf{E}}_j[n] = \widehat{\mathbf{A}}_j[n] + \widehat{\mathbf{C}}_j[n] + \lambda_j \mathbf{I}$. This leads to the generalized eigenvector solution and can be written as $vec(\mathbf{V}^j) = \mathbf{V}_{max}((\sum_{n=1}^{N_s} (\mathbf{G}_j[n] \mathbf{G}_j^H[n])^T \otimes \widehat{\mathbf{B}}_j[n]), \sum_{n=1}^{N_s} ((\mathbf{G}_j[n] \mathbf{G}_j^H[n])^T \otimes \widehat{\mathbf{E}}_j[n]))$.

3.4.3 Optimization of stream powers

One advantage of the Lagrangian formulation (7.17) is that it allows introducing stream powers for each BS, so $\mathbf{G}_j[n] = \mathbf{G}'_j[n] \mathbf{P}_j^{1/2}[n]$, where the diagonal matrix $\mathbf{P}_j[n]$ represents the power allocated to an unknown number of supportable streams for BS j. To render a feasible solution for the stream powers, we approximate the concave part of the EWSR by a first order local minorizer function.

$$
\begin{aligned}
&\ln\det(\mathbf{I} + \mathbf{G}_j^H[n] \mathbf{V}^{j\,H} \widehat{\mathbf{B}}_j[n] \mathbf{V}^j \mathbf{G}_j[n]) = \ln\det(\mathbf{I} + \mathbf{P}_j[n] \widehat{\mathbf{S}}_j[n]) + \mathrm{tr}\{(\mathbf{P}_j[n] - \widehat{\mathbf{P}}_j[n]) \widehat{\mathbf{T}}_j\}, \text{where,} \\
&\qquad \widehat{\mathbf{T}}_j[n] = \mathbf{G}_j^H[n] \mathbf{V}^{j\,H} \widehat{\mathbf{E}}_j[n] \mathbf{V}^j \mathbf{G}_j[n], \\
&\qquad \widehat{\mathbf{S}}_j[n] = \mathbf{G}_j^H[n] \mathbf{V}^{j\,H} \widehat{\mathbf{B}}_j[n] \mathbf{V}^j \mathbf{G}_j[n]
\end{aligned} \tag{3.61}
$$

For the concave local minorization considered above, this works well as long as the next optimum is within the minorization range. Note that $\widehat{\mathbf{S}}_j[n], \widehat{\mathbf{T}}_j[n]$ are diagonal since $\mathbf{G}_j[n]$ diagonalizes the matrices $\mathbf{V}^{j\,H} \widehat{\mathbf{B}}_j[n] \mathbf{V}^j$ and $\mathbf{V}^{j\,H} \widehat{\mathbf{E}}_j[n] \mathbf{V}^j$. Further optimizing w.r.t $\mathbf{P}_j[n]$ leads to the self-interference and LDR aware water-filling (SILA-WF) solution for the stream powers

$$
\mathbf{P}_j[n] = \left(u_j \widehat{\mathbf{T}}_j^{-1}[n] - \widehat{\mathbf{S}}_j^{-1}[n] \right)^+ \tag{3.62}
$$

where $(x)^+ = \max(0, x)$ is applied to all diagonal elements and the Lagrange multipliers are adjusted to satisfy the power constraints. This can be done by bisection and gets executed per BS.

3.4.3.1 Analog Phase Shifter Design

For the constrained analog BF case, where the BF coefficients are chosen to be phasors, we utilize the DA based approach proposed earlier in our work [44, 46]. We refer the reader for a more detailed discussion on this to our paper [44, Algorithm 3]. We remark here that in this chapter, we consider only a case of two backhaul nodes for simplicity. The extension to the multi-user case with multiple FD or half-duplex nodes, example [55] (which is fully digital), is quite straightforward and left as future work.

Algorithm 8: Minorization based multi-stage/HBF design

Given: $P_j, \mathbf{H}_{i,j}[n], u_i, \mathbf{H}_{i,i}[n]\ \forall i, j, n$.
Initialization: $\mathbf{V}^j$ is selected as the eigenvectors of the direct channel covariance matrix, The $\mathbf{G}_j^{(0)}[n]$ are initialized to be ZF precoders for the effective channels $\mathbf{H}_{i,j}[n]\mathbf{V}^j$, with uniform power distribution across the streams. **Iteration** (t) :

1. Compute the Rx side digital combiner $\mathbf{F}_{BB,i}^{(t)}$ from (3.46).
2. Update the Rx side analog combiner $\mathbf{F}_{RF,i}^{(t)}$ using (3.49).
3. Compute $\widehat{\mathbf{B}}_j[n], \widehat{\mathbf{A}}_j[n]$,from (3.54) and $\widehat{\mathbf{C}}_j[n]\ \forall j, n$.
4. Update $\mathbf{G}_j^{'(t)}[n]$ from (3.57), and $\mathbf{P}_j[n]$ from (3.62), $\forall k, n$. Compute λ_j using bisection.
5. Update $(\mathbf{V}^j)^{(t)}, \forall j$, using DA (phasor constrained) or from (3.60) (unconstrained).
6. If the algorithm is converged, exit the loop, otherwise go to step 1.

3.5 Simulation Results

Simulations to validate the performance of the proposed hybrid BF algorithms are presented for a bidirectional FD system under the LDR noise model. We follow the partial CSIT model $\mathbf{H}_{i,j}$ as in Section II.A. For the SI channel, we ignore the near field effect of amplitude variation with distance and the near field effects in the phase variation. In the Uniform Linear Array (ULA), the AoD or AoA ϕ, θ are assumed to be uniformly distributed in the interval $[0^o, 30^o]$.

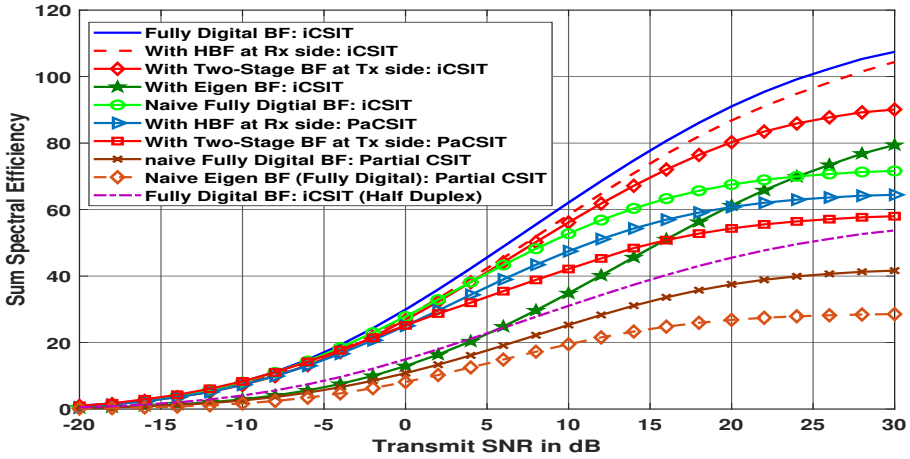


Figure 3.5: Ergodic Capacity Analysis: Sum Rate comparisons for, OFDM, $N_s = 8, N_t^i = N_r^i = 8, M_t^i = M_r^i = 4, d_i = 1, \forall i, L = 4$ paths.

In Figure 3.5, Eigen BF corresponds to the sub-optimal BF (fully digital) with the BFs at both sides selected as the right or left singular vectors of the direct channel which is projected onto the orthogonal complement of the SI channel, respectively. The superior performance of our proposed approach is due to the fact that our BFs are optimized to take into account the LDR noise on both sides. In Figure 3.5, we look at ergodic capacity analysis with the proposed ESIP-

WSR based BF design here. Notations: “paCSIT” corresponds to partial CSIT and “iCSIT” corresponds to perfect or instantaneous CSIT. Naive BFs in the case of partial CSIT corresponds to the case when we treat the estimated channel as true channel and the BFs being optimized using the WSR. So error covariance information is not exploited for the naive BFs. So Figure 3.5 clearly shows the advantage in exploiting the error covariance information which the proposed ESIP-WSR does. Also, the curve “Naive Fully Digital BF:iCSIT” is the scenario where we ignore the presence of LDR noise in the design of BFs. Ignoring the LDR noises results in a significant reduction in the sum rate. The dimensions of the two-stage BF and hybrid BF are such that the zero forcing capabilities at both sides are comparable. However, the number of LDR noises is the number of antennas at the Tx side, whereas, for the analog Rx stage, the number of LDR noises is the number of analog BF outputs, which is less. We conjecture that this would explain the better performance of the analog stage at Rx (in both figures) compared to the two-stage architecture at Tx for SI nulling.

3.6 Conclusion

In this chapter, we looked at BF solutions to null the SI power under a more practical noise model termed limited dynamic range. We proposed a multi-stage BF design (whose performance is validated through simulations), with a frequency flat analog or time domain combiner/BF stage and a frequency dependent baseband precoder/combiner. We optimized the EWSR using an alternating minorization approach which converges to a local optimum. We considered a Massive MIMO limit approximation of the EWSR termed as ESIP-WSR which has significant performance gains compared to an Expected Weighted Sum MSE (EWSMSE) based BF design (which represents a lower bound to the ergodic capacity) [64].

Chapter 4

NONCOHERENT MULTI-USER MIMO COMMUNICATIONS USING COVARIANCE CSIT

4.1 Introduction

Till now, we looked at beamforming design for a hybrid MaMIMO system under perfect channel knowledge. However, the practical importance of fully digital systems cannot be fully overshadowed by HBF systems as detailed in the introductory chapter. Moreover, it becomes quite simplistic to analyze the spectral efficiency behavior for a fully digital system as we show later in the thesis. Moreover, perfect channel knowledge is very impractical and from here on we start looking at imperfections in the channel estimate. As a starting point, we look at robust BF designs under partial CSIT. Note that, the BF designs outlined in this chapter can easily be extended to HBF system also, for partial CSIT case.

The Multi-User downlink, particularly in a Multi-Cell Massive MIMO setting, requires enormous amounts of instantaneous CSIT (Channel State Information at the Transmitter(s)), iCSIT. Here we focus on exploiting channel covariance CSIT (coCSIT) only. In particular multipath induced structured low rank covariances are considered that arise in Massive MIMO and mmWave settings, which we call pathwise CSIT (pwCSIT). The resulting non-Kronecker MIMO channel covariance structures lead to a split between the roles of transmitters and receivers in MIMO systems. For the beamforming optimization, we consider a minorization approach applied to the Massive MIMO limit of the Expected Weighted Sum Rate. Simulations indicate that the pwCSIT based designs may lead to limited spectral efficiency loss compared to iCSIT based designs, while trading fast fading CSIT for slow fading CSIT. We also point out that the pathwise approach may lead to distributed designs with only local pwCSIT, and analyze the sum rates for iCSIT and pwCSIT in the low and high SNR limits.

Interference is the main limiting factor in wireless transmission. Base stations (BSs) disposing of multiple antennas are able to serve multiple Mobile Terminals (MTs) simultaneously, which is called Spatial Division Multiple Access (SDMA) or Multi-User (MU) MIMO. However, MU systems have precise requirements for Channel State Information at the Tx (CSIT) which is more difficult to acquire than CSI at the Rx (CSIR). Hence we focus here on the more challenging downlink (DL).

The recent development of Massive MIMO (MaMIMO) [66] opens new possibilities for increased system capacity while at the same time simplifying system design. We refer to [67] for a further discussion of the state of the art, in which MIMO Interference Alignment (IA) requires global MIMO channel CSIT. Recent works focus on intercell exchange of only scalar quantities, at fast fading rate, as also on two-stage approaches in which the intercell interference gets zero-

forced (ZF). Also, massive MIMO in most works refers actually to MU MISO.

Whereas the exploitation of covariance CSIT (coCSIT) may be beneficial, in a MaMIMO context it may quickly lead to high computational complexity and estimation accuracy issues. Computational complexity may be reduced (and the benefit of coCSIT enhanced) in the case of low rank or related covariance structure, but the use and tracking of subspaces may still be cumbersome. In the pathwise approach, these subspaces are very parsimoniously parameterized. In a FDD setting, these parameters may even be estimated from the uplink (UL). As opposed to the instantaneous channel CSIT (iCSIT), the pathwise CSIT (pwCSIT) is not affected by fast fading.

Massive MIMO makes the pathwise approach viable. Indeed, with enough antennas, pwCSIT by itself may allow zero forcing (ZF) [68], which is of interest at high SNR. However, we are particularly concerned here with maximum Weighted Sum Rate (WSR) designs accounting for finite SNR. ZF of all interfering links leads to significant reduction of useful signal strength. We briefly allude to the general case of Gaussian partial CSIT (paCSIT), in which the combined availability of channel estimates (mean CSIT) and coCSIT can be exploited. Such general paCSIT scenario can example be particularized as in [69] to the case of perfect iCSIT for intracell channels and pwCSIT for intercell channels. This leads to 2-stage BF expressions, similar to hybrid beamforming. The slow stage handles intercell interference, and is frequency-flat. It can be exploited also to separate the cells for channel estimation purposes. In what follows we consider in more detail pwCSIT for all channels (both intercell and intracell). Also, in this (as any) case of paCSIT, the WSR criterion needs to be modified. We shall consider the Expected WSR (EWSR). Furthermore, we shall take advantage of a Massive MIMO setting to exploit a simple Massive EWSR limit that results from the law of large numbers. This MaEWSR limit leads to a loss of all (narrowband) frequency-selectivity in the channel and also leaves no utility for space-time coding, though this can be expected to bring some benefits.

4.2 Streamwise IBC Signal Model

We start with a per stream approach (which in the perfect CSI case would be equivalent to per user). In an IBC formulation, one stream per user can be expected to be the usual scenario. In the development below, in the case of more than one stream per user, treat each stream as an individual user. So, consider again an IBC with C cells with a total of K users. We shall consider a system-wide numbering of the users. User k is served by BS b_k. The $N_k \times 1$ received signal at user k in cell b_k is

$$\mathbf{y}_k = \underbrace{\mathbf{H}_{k,b_k}\,\mathbf{g}_k\,x_k}_{\text{signal}} + \underbrace{\sum_{\substack{i\neq k\\ b_i=b_k}} \mathbf{H}_{k,b_k}\,\mathbf{g}_i\,x_i}_{\text{intracell interf.}} + \underbrace{\sum_{j\neq b_k}\sum_{i:b_i=j} \mathbf{H}_{k,j}\,\mathbf{g}_i\,x_i}_{\text{intercell interf.}} + \mathbf{v}_k \tag{4.1}$$

where x_k is the intended (white, unit variance) scalar signal stream, $\mathbf{H}_{k,b_k}$ is the $N_k \times M_{b_k}$ channel from BS b_k to user k. BS b_k serves $K_{b_k} = \sum_{i:b_i=b_k} 1$ users. We considering a noise whitened signal representation so that we get for the noise $\mathbf{v}_k \sim \mathcal{CN}(0, I_{N_k})$. The $M_{b_k} \times 1$ spatial Tx filter or beamformer (BF) is $\mathbf{g}_k$. Treating interference as noise, user k will apply a linear Rx filter $\mathbf{f}_k$ to maximize the signal power (diversity) while reducing any residual interference that would not have been (sufficiently) suppressed by the BS Tx. The Rx filter output is $\widehat{x}_k = \mathbf{f}_k^H \mathbf{y}_k$.

4.3 Max WSR with Perfect CSIT

Consider as a starting point for the optimization the weighted sum rate (WSR)

$$WSR = WSR(\mathbf{g}) = \sum_{k=1}^{K} u_k \ln \frac{1}{e_k} \tag{4.2}$$

where $\mathbf{g}$ represents the collection of BFs $\mathbf{g}_k$, the u_k are rate weights, the $e_k = e_k(\mathbf{g})$ are the Minimum Mean Squared Errors (MMSEs) for estimating the x_k:

$$\begin{aligned} \frac{1}{e_k} &= 1 + \mathbf{g}_k^H \mathbf{H}_{k,b_k}^H \mathbf{R}_{\overline{k}}^{-1} \mathbf{H}_{k,b_k} \mathbf{g}_k \\ &= (1 - \mathbf{g}_k^H \mathbf{H}_{k,b_k}^H \mathbf{R}_k^{-1} \mathbf{H}_{k,b_k} \mathbf{g}_k)^{-1} \\ \mathbf{R}_k &= \mathbf{H}_{k,b_k} \mathbf{Q}_k \mathbf{H}_{k,b_k}^H + \mathbf{R}_{\overline{k}},\ \mathbf{Q}_i = \mathbf{g}_i \mathbf{g}_i^H, \\ \mathbf{R}_{\overline{k}} &= \sum_{i \neq k} \mathbf{H}_{k,b_i} \mathbf{Q}_i \mathbf{H}_{k,b_i}^H + \mathbf{I}_{N_k}. \end{aligned} \tag{4.3}$$

$\mathbf{R}_k$, $\mathbf{R}_{\overline{k}}$ are the total and interference plus noise Rx covariance matrices resp. and e_k is the MMSE obtained at the output $\widehat{x}_k = \mathbf{f}_k^H \mathbf{y}_k$ of the optimal (MMSE) linear Rx $\mathbf{f}_k$,

$$\mathbf{f}_k = \mathbf{R}_k^{-1} \mathbf{H}_{k,b_k} \mathbf{g}_k = \mathbf{R}_k^{-1} \mathbf{h}_{k,k}. \tag{4.4}$$

The WSR cost function needs to be augmented with the power constraints

$$\sum_{k:b_k=j} \mathrm{tr}\{\mathbf{Q}_k\} \leq P_j. \tag{4.5}$$

4.3.1 From Max WSR to Min WSMSE

For a general Rx filter $\mathbf{f}_k$ we have the MSE

$$\begin{aligned} e_k(\mathbf{f}_k, \mathbf{g}) &= (1 - \mathbf{f}_k^H \mathbf{H}_{k,b_k} \mathbf{g}_k)(1 - \mathbf{g}_k^H \mathbf{H}_{k,b_k}^H \mathbf{f}_k) + \sum_{i \neq k} \mathbf{f}_k^H \mathbf{H}_{k,b_i} \mathbf{g}_i \mathbf{g}_i^H \mathbf{H}_{k,b_i}^H \mathbf{f}_k + ||\mathbf{f}_k||^2 \\ &= 1 - \mathbf{f}_k^H \mathbf{H}_{k,b_k} \mathbf{g}_k - \mathbf{g}_k^H \mathbf{H}_{k,b_k}^H \mathbf{f}_k + \sum_{i} \mathbf{f}_k^H \mathbf{H}_{k,b_i} \mathbf{g}_i \mathbf{g}_i^H \mathbf{H}_{k,b_i}^H \mathbf{f}_k + ||\mathbf{f}_k||^2. \end{aligned} \tag{4.6}$$

The $WSR(\mathbf{g})$ is a non-convex and complicated function of $\mathbf{g}$. Inspired by [9], we introduced [70], [11] an augmented cost function, the Weighted Sum MSE, $WSMSE(\mathbf{g}, \mathbf{f}, w)$

$$= \sum_{k=1}^{K} u_k (w_k\, e_k(\mathbf{f}_k, \mathbf{g}) - \ln w_k) + \sum_{i=1}^{C} \lambda_i \Big(\sum_{k:b_k=i} ||\mathbf{g}_k||^2 - P_i \Big) \tag{4.7}$$

where λ_i = Lagrange multipliers. After optimizing over the aggregate auxiliary Rx filters $\mathbf{f}$ and weights w, we get the WSR back:

$$\min_{\mathbf{f}, w} WSMSE(\mathbf{g}, \mathbf{f}, w) = -WSR(\mathbf{g}) + \sum_{k=1}^{K} u_k \tag{4.8}$$

The advantage of the augmented cost function: alternating optimization leads to solving simple quadratic or convex functions:

$$\begin{aligned}
&\min_{w_k} WSMSE \Rightarrow\ w_k = 1/e_k \\
&\min_{\mathbf{f}_k} WSMSE \Rightarrow \mathbf{f}_k = (\sum_i \mathbf{H}_{k,b_i}\mathbf{g}_i\mathbf{g}_i^H\mathbf{H}_{k,b_i}^H + \mathbf{I}_{N_k})^{-1}\mathbf{H}_{k,b_k}\mathbf{g}_k \\
&\min_{\mathbf{g}_k} WSMSE \Rightarrow \\
&\qquad\qquad \mathbf{g}_k = (\sum_i u_i w_i \mathbf{H}_{i,b_k}^H \mathbf{f}_i\mathbf{f}_i^H\mathbf{H}_{i,b_k} + \lambda_{b_k}\mathbf{I}_M)^{-1}\mathbf{H}_{k,b_k}^H\mathbf{f}_k u_k w_k
\end{aligned} \tag{4.9}$$

UL/DL duality: the optimal Tx filter g_k is of the form of a MMSE linear Rx for the dual UL in which λ plays the role of Rx noise variance and $u_k w_k$ plays the role of stream variance.

4.3.2 Minorization (DC Programming)

In this section, we look at the BF design using alternating minorization concept proposed for the hybrid beamforming in Section 2.2. The digital BF derived here is a special case with the number of RF chains equal to the number of antennas. Hence, by substituting for $\mathbf{V}^c = \mathbf{I}_M$, we obtain the all digital BF expressions and hence we can skip the detailed derivation here. We define the transmit covariance matrix as $\mathbf{Q}_k = \mathbf{G}_k\mathbf{G}_k^H$ and the auxiliary quantities involved in the BF expressions as

$$\begin{aligned}
\mathbf{B}_k &= \mathbf{H}_{k,b_k}^H \mathbf{R}_{\bar{k}}^{-1}\mathbf{H}_{k,b_k}, \\
\mathbf{A}_k &= \sum_{i\neq k}^{K} u_i \mathbf{H}_{i,b_k}^H(\mathbf{R}_{\bar{i}}^{-1} - \mathbf{R}_i^{-1})\mathbf{H}_{i,b_k}
\end{aligned} \tag{4.10}$$

Further, the digital BF ($\mathbf{G}_k = \overline{\mathbf{G}}_k\mathbf{P}_k^{\frac{1}{2}}$)expression can be obtained as

$$\overline{\mathbf{G}}_k = V_{max}(\mathbf{B}_k, \mathbf{A}_k + \lambda_{b_k}\mathbf{I}) \tag{4.11}$$

are the (normalized) "max" generalized eigenvectors of the two indicated matrices, with eigenvalues $\mathbf{\Sigma}_k = \mathbf{\Sigma}_{max}(\mathbf{B}_k, \mathbf{A}_k + \lambda_{b_k} I)$. Let $\mathbf{\Sigma}_k^{(1)} = \overline{\mathbf{G}}_k^H\mathbf{B}_k\overline{\mathbf{G}}_k$ and $\mathbf{\Sigma}_k^{(2)} = \overline{\mathbf{G}}_k^H\mathbf{A}_k\mathbf{G}_k$. The optimization of power leads to the following interference leakage aware water filling (WF) (jointly for the P_k and λ_c)

$$\mathbf{P}_k = \left(u_k(\mathbf{\Sigma}_k^{(2)} + \lambda_{b_k}\mathbf{I})^{-1} - \mathbf{\Sigma}_k^{-(1)}\right)^+, \ \sum_{k:b_k=c} \mathrm{tr}\{\mathbf{P}_k\} = P_c \tag{4.12}$$

where the Lagrange multipliers are adjusted to satisfy the power constraints. This can be done by bisection and gets executed per BS. Note that some Lagrange multipliers could be zero. Note also that as with any alternating optimization procedure, there are many updating schedules possible, with different impact on convergence speed. The quantities to be updated are the $\overline{\mathbf{g}}_k$, the $\mathbf{P}_k$ and the λ_c. Note that the minorization approach, which avoids introducing Rxs, can at every BF update allow to introduce an arbitrary number of streams per user by determining multiple dominant generalized eigenvectors, and then let the WF operation decide how many streams can actually be sustained.

In contrast, in [30], for given λ, the $\mathbf{G}$ get iterated till convergence and the λ are found by duality (line search):

$$\min_{\lambda\geq 0}\max_{\mathbf{G}}[\sum_{j}^{C}\lambda_j P_j+\sum_k\{u_k\ln\det(\mathbf{R}_k^{-1}\mathbf{R}_k)-\lambda_{b_k}\mathrm{tr}\{\mathbf{P}_k\}\}] = \min_{\lambda\geq 0} WSR(\lambda). \tag{4.13}$$

This typically leads to higher computational complexity for a given convergence precision.

4.3.3 Pathwise Wireless MIMO Channel Model

In this section we drop the user index k for simplicity. The MIMO channel transfer matrix at any particular subcarrier n of a given OFDM symbol can be written as [71], [72]

$$\begin{aligned}\mathbf{H}[n] &= \sum_{i=1}^{L} A_i e^{j\psi_i[n]}\mathbf{h}_r(\phi_i)\mathbf{h}_t^T(\theta_i)\\ &= \mathbf{H}_r\,\mathbf{\Psi}[n]\,\mathbf{D}\,\mathbf{H}_t^H,\end{aligned}$$

$$\mathbf{H}_r = \left[\mathbf{h}_r(\phi_1)\,\mathbf{h}_r(\phi_1)\cdots\right],$$

$$\mathbf{\Psi}[n] = \begin{bmatrix} e^{j\psi_1[n]} & & \\ & e^{j\psi_2[n]} & \\ & & \ddots \end{bmatrix}, \tag{4.14}$$

$$\mathbf{D} = \begin{bmatrix} A_1 & & \\ & A_2 & \\ & & \ddots \end{bmatrix},$$

$$\mathbf{H}_t^H = \begin{bmatrix} \mathbf{h}_t^T(\theta_1) \\ \mathbf{h}_t^T(\theta_2) \\ \vdots \end{bmatrix}$$

where there are L (specular) pathwise contributions with

- $A_i > 0$: path amplitude
- $\psi_i[n]$: path phase
- θ_i: angle of departure (AoD)
- ϕ_i: angle of arrival (AoA)
- $\mathbf{h}_t(.)/\mathbf{h}_r(.)$: $M/N\times 1$ Tx/Rx antenna array response

with $||\mathbf{h}_t(.)|| = 1$, $||\mathbf{h}_r(.)|| = N$. For wideband scenarios, all factors may become frequency-dependent. The antenna array responses are just functions of angles AoD, AoA in the case of standard antenna arrays with scatterers in the far field. The fast variation of the phases ψ_i (due to Doppler) corresponds to the fast fading. All the other parameters vary on a slower time scale and correspond to slow fading. In the pathwise CSIT (pwCSIT) model, we shall assume the ψ_i to be i.i.d. uniformly distributed and all slow parameters to be known. Note that the pathwise channel model, which leads here to a type of Tx covariance CSIT, does not lead to the usual separable

covariance case, which is discussed example in [67]. In previous work, we essentially modeled the whole of $\mathbf{H}_r\mathbf{\Psi}$ as i.i.d. random, which leads to a special case of the MIMO channel with separable correlation structure. Here the knowledge of $\mathbf{H}_r$ is exploited, leading to an appearance of (implicit) Rxs who contribute to the interference management.

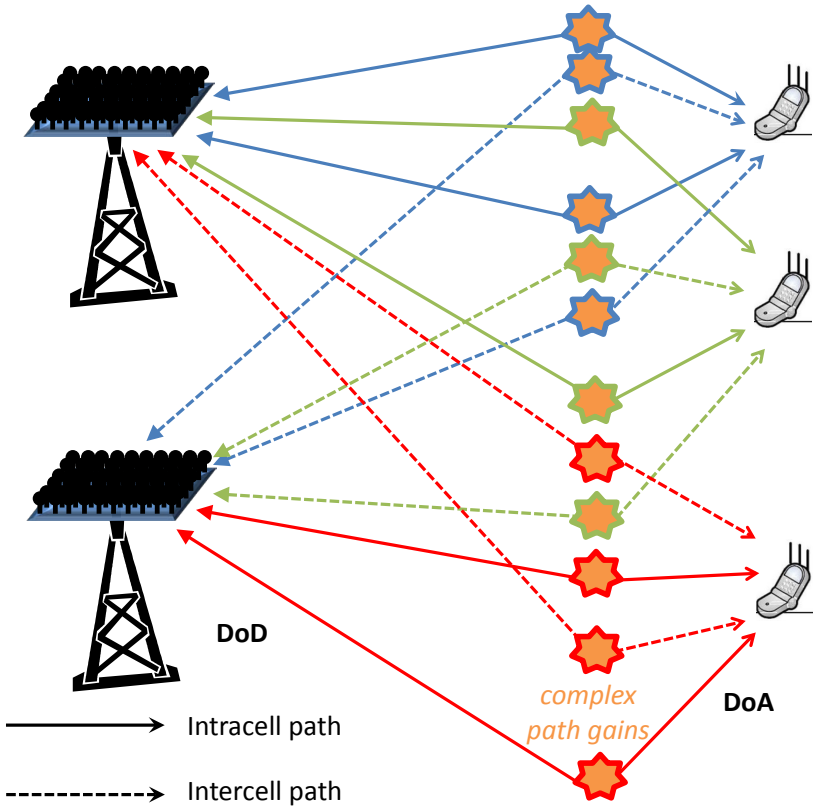


Figure 4.1: Pathwise Multi-User Multi-Cell scenario.

4.4 MIMO Interference Alignment (IA)

ZF (IA) feasibility for both the general reduced rank MIMO channels case and the pathwise MIMO case has been discussed in [68], in particular also when only based on Tx side covariance CSIT. It is shown how the IA responsability gets shared between Tx and Rx, requiring only local CSI. Also the role of Rx antennas is highlighted, leading to reduced (Tx covariance) rank channels.

4.5 Expected WSR (EWSR)

For the WSR criterion, we have assumed so far that the channel $\mathbf{H}$ is known. The scenario of interest however is that of partial CSIT. Once the CSIT is imperfect, various optimization criteria could be considered, such as outage capacity. Here we shall consider the expected weighted sum

rate

$$\begin{aligned} \mathrm{E}_{\mathbf{H}|\overline{\mathbf{H}}} WSR(\mathbf{g},\mathbf{H}) &= EWSR(\mathbf{g}) \\ &= \mathrm{E}_{\mathbf{H}|\overline{\mathbf{H}}} \sum_k u_k \ln(1+\mathbf{g}_k^H \mathbf{H}_{k,b_k}^H \mathbf{R}_{\overline{k}}^{-1} \mathbf{H}_{k,b_k} \mathbf{g}_k) \end{aligned} \tag{4.15}$$

where we now underlign the dependence of various quantities on $\mathbf{H}$ and $\overline{\mathbf{H}}$ is a channel estimate. The EWSR in (4.2) corresponds to perfect CSIR since the optimal Rx filters $\mathbf{f}_k$ as a function of the aggregate $\mathbf{H}$ have been substituted, namely $WSR(\mathbf{g},\mathbf{H}) = \max_{\mathbf{f}} \sum_k u_k(-\ln(e_k(\mathbf{f}_k,\mathbf{g})))$.

In the MaMIMO limit, we obtain the *Massive EWSR limit* in which

$$\mathrm{E}_{\mathbf{H}|\overline{\mathbf{H}}} \ln\det(\mathbf{I}+\mathbf{H}\mathbf{Q}\mathbf{H}^H) \to \ln\det(\mathbf{I}+\mathrm{E}_{\mathbf{H}|\overline{\mathbf{H}}}\{\mathbf{H}\mathbf{Q}\mathbf{H}^H\}) \tag{4.16}$$

when $M \to \infty$ for finite N. The gap between both sides in (4.16) can be analyzed and is bounded for any MIMO size by γ (Euler-Mascheroni) in the worst case of only a single Rayleigh fading entry. The RHS also corresponds to the Expected Weighted Sum Unbiased MSE (EWSUMSE) approach introduced in [73], which is a useful formulation by itself. The RHS also becomes the exact mutual information if we consider Gaussian channel outputs instead of Gaussian channel inputs.

For the case of mean (channel estimate) and covariance CSIT being jointly captured by the Gaussian CSIT, $\mathrm{vec}(\mathbf{H}^T) = \mathbf{h} \sim \mathscr{CN}(\overline{\mathbf{h}}, C_{\mathbf{hh}})$ where $\overline{\mathbf{h}} = \mathrm{vec}(\overline{\mathbf{H}}^T)$, we get

$$\mathrm{E}\{\mathbf{H}\mathbf{g}\mathbf{g}^H\mathbf{H}^H\} = \overline{\mathbf{H}}\mathbf{g}\mathbf{g}^H\overline{\mathbf{H}}^H + (\mathbf{I}_N \otimes \mathbf{g}^T) C_{\mathbf{hh}} (\mathbf{I}_N \otimes \mathbf{g}^*)\,. \tag{4.17}$$

This general paCSIT model, even with a pathwise channel model, could account for unmodeled paths, estimation errors on the path parameters, etc. Here we shall consider that all paths are modeled and perfectly known, except for the path phases.

4.5.1 Massive EWSR with pwCSIT

For the special case of pwCSIT (4.14) considered here, if the total number of paths (all users) becomes very large, the path phases average out and by the law of large numbers

$$\begin{aligned} \mathrm{E}_\Psi \ln\det(\mathbf{I}+\mathbf{R}_{\overline{k}}^{-1}\mathbf{H}_{k,b_k}\mathbf{Q}_k\mathbf{H}_{k,b_k}^H) &\approx \ln\det(\mathbf{I}+\mathbf{R}_{\overline{k}}^{-1}\,\mathrm{E}_\Psi \mathbf{H}_{k,b_k}\mathbf{Q}_k\mathbf{H}_{k,b_k}^H) \\ \mathbf{H}_{k,b_k}\mathbf{Q}_k\mathbf{H}_{k,b_k}^H &\longrightarrow \mathrm{E}_\Psi\, \mathbf{H}_{k,b_k}\mathbf{Q}_k\mathbf{H}_{k,b_k}^H \\ &= \mathbf{H}_{r,k,b_k}\,\mathbf{D}_{k,b_k}\, diag(\mathbf{H}_{t,k,b_k}^H \mathbf{Q}_k \mathbf{H}_{t,k,b_k})\,\mathbf{D}_{k,b_k}\,\mathbf{H}_{r,k,b_k}^H \end{aligned} \tag{4.18}$$

which is now frequency-independent, and where diag(.) denotes the diagonal matrix obtained by taking the diagonal part of the matrix argument. Hence we get the following MaMIMO limit matrices

$$\begin{aligned} \mathbf{R}_k[n] &= \mathbf{I}_{N_k} + \sum_{i=1}^{K} \mathbf{H}_{r,k,b_i}\mathbf{D}_{k,b_i}^2\, diag(\mathbf{H}_{t,k,b_i}^H\mathbf{Q}_i\mathbf{H}_{t,k,b_i})\mathbf{H}_{r,k,b_i}^H \\ \mathbf{R}_{\overline{k}}[n] &= \mathbf{I}_{N_k} + \sum_{i\neq k} \mathbf{H}_{r,k,b_i}\mathbf{D}_{k,b_i}^2\, diag(\mathbf{H}_{t,k,b_i}^H\mathbf{Q}_i\mathbf{H}_{t,k,b_i})\mathbf{H}_{r,k,b_i}^H \end{aligned} \tag{4.19}$$

This leads to example (with $\mathbf{Q}_i = \mathbf{G}_i\mathbf{G}_i^H$) :

$$\frac{\partial \ln\det(\mathbf{R}_k)}{\partial \mathbf{G}_i^*} = \mathbf{H}_{t,k,b_i}\, diag(\mathbf{H}_{r,k,b_i}^H \mathbf{R}_k^{-1}\mathbf{H}_{r,k,b_i})\mathbf{D}_{k,b_i}^2\mathbf{H}_{t,k,b_i}^H\mathbf{G}_i, \tag{4.20}$$

and we can introduce

$$\begin{aligned}\bar{\mathbf{B}}_k &= \mathbf{H}_{t,k,b_k}\, diag(\mathbf{H}_{r,k,b_k}^H \bar{\mathbf{R}}_k^{-1} \mathbf{H}_{r,k,b_k})\mathbf{D}_{k,b_k}^2 \mathbf{H}_{t,k,b_k}^H,\\ \bar{\mathbf{A}}_k &= \sum_{i\neq k}^{K} u_i \mathbf{H}_{t,i,b_k} diag(\mathbf{H}_{r,i,b_k}^H (\bar{\mathbf{R}}_i^{-1} - \mathbf{R}_i^{-1})\mathbf{H}_{r,i,b_k})\mathbf{D}_{i,b_k}^2 \mathbf{H}_{t,i,b_k}^H.\end{aligned} \tag{4.21}$$

It suffices now to replace the matrices $\mathbf{A}_k$, $\mathbf{B}_k$ in the minorization approach in Section 4.3.2 by the matrices $\bar{\mathbf{A}}_k$, $\bar{\mathbf{B}}_k$ above to get a maximum EWSR design:

$$\begin{aligned}\mathbf{G}_k^{'} &= V_{1:d_k}(\bar{\mathbf{B}}_k, \bar{\mathbf{A}}_k + \lambda_{b_k}\mathbf{I}).\\ \text{With } \mathbf{\Sigma}_k^{(1)} &= \mathbf{G}_k^{'H}\bar{\mathbf{B}}_k\mathbf{G}_k^{'},\\ \mathbf{\Sigma}_k^{(2)} &= \mathbf{G}_k^{'H}\bar{\mathbf{A}}_k\mathbf{G}_k^{'}, \mathbf{G}_k = \mathbf{G}_k^{'}\mathbf{P}_k^{\frac{1}{2}},\end{aligned} \tag{4.22}$$

where

$$\mathbf{P}_k = \left(u_k(\mathbf{\Sigma}_k^{(2)} + \lambda_{b_k}\mathbf{I})^{-1} - \mathbf{\Sigma}_k^{-(1)}\right)^+, \sum_{k:b_k=c} \mathrm{tr}\{\mathbf{P}_k\} = P_c\,. \tag{4.23}$$

Further in this paragraph, we provide some intuitive interpretation of the above BF expressions. For a multi-user MIMO case, given the channel matrices, the optimal beamforming expressions as in [9] or the alternating minorization based approach derived in Chapter II is clearly understood. However, when the channel matrix can be accurately captured by a physical (geometric) scattering model across multiple clusters/paths as is the case in mmWave or massive MIMO systems, the structure of the optimal BF expressions are not clear. Towards this direction, in this chapter, using the derived expressions in (4.22), we provide a physical interpretation for this optimal structure, i.e., beam steering across the different paths with appropriate power allocation. Basically, $\bar{\mathbf{B}}_k$ represents a linear combination of the transmit antenna responses of the direct channel to user k and the weighting coefficients are proportional the path powers. Similarly, $\bar{\mathbf{A}}_k$ represents a weighted combination of the Tx side antenna array responses of all the leakage channels (channels to which BF $\mathbf{G}_k$ cause interference). Hence, we can conclude that the BF expression above is an optimal compromise (in terms of maximizing the massive MIMO limit of the ergodic capacity) between maximizing the direct channel power part and minimizing the leakage power part. The weighting coefficients in the BF expressions are oblivion to the path phases, since they are treated as unknown here.

4.5.2 Interference management by Tx/Rx

In this section, we discuss about the interference management by either Tx/Rx. First looking at the expression for $\mathbf{R}_k[n]$

$$\mathbf{R}_k[n] = \mathbf{I}_{N_k} + \sum_{i=1}^{K} \mathbf{H}_{r,k,b_i}\mathbf{D}_{k,b_i}^2\, diag(\underbrace{\mathbf{H}_{t,k,b_i}^H \mathbf{G}_i}\,\underbrace{\mathbf{G}_i^H \mathbf{H}_{t,k,b_i}})\mathbf{H}_{r,k,b_i}^H \tag{4.24}$$

where the underbraced terms would be zero for $i \neq k$ in case of a ZF design (high SNR optimal) for the Tx BF $\mathbf{G}_k$. Also, one can identify implicit Rxs from the expression of $\bar{\mathbf{A}}_k$ as follows

(4.25)

$$\begin{aligned}
&\operatorname{diag}(\mathbf{H}_{r,i,b_k}^H(\mathbf{R}_{\bar{i}}^{-1}-\mathbf{R}_i^{-1})\mathbf{H}_{r,i,b_k}) \\
&\qquad \overset{(a)}{=} \operatorname{diag}(\mathbf{H}_{r,i,b_k}^H \underbrace{\mathbf{R}_i^{-1}\mathbf{H}_{r,i,b_i}}_{\mathbf{F}_i}(\mathbf{D}_i-\mathbf{H}_{r,i,b_i}^H\mathbf{R}_i^{-1}\mathbf{H}_{r,i,b_i})^{-1}\mathbf{H}_{r,i,b_i}^H\mathbf{R}_i^{-1}\mathbf{H}_{r,i,b_k}) \\
&\qquad = \operatorname{diag}(\underbrace{\mathbf{H}_{r,i,b_k}^H\mathbf{F}_i}\widetilde{\mathbf{D}}_i\underbrace{\mathbf{F}_i^H\mathbf{H}_{r,i,b_k}})
\end{aligned}$$

where the $\mathbf{F}_i$ are implicit Rxs and again the underbraced terms would be zero for $i \neq k$ in case of a ZF design. To reach (a), we defined $\mathbf{D}_i^{-1} = \mathbf{D}_{i,b_k}^2 \operatorname{diag}(\underbrace{\mathbf{H}_{t,i,b_k}^H\mathbf{G}_k}\underbrace{\mathbf{G}_k^H\mathbf{H}_{t,i,b_k}})$ and $\mathbf{R}_{\bar{i}} = \mathbf{R}_i - \mathbf{H}_{r,i,b_i}\mathbf{D}_i^{-1}\mathbf{H}_{r,i,b_i}^H$. Further, we used matrix inversion lemma $(\mathbf{A}+\mathbf{BCD})^{-1} = \mathbf{A}^{-1}-\mathbf{A}^{-1}\mathbf{B}(\mathbf{C}^{-1}+\mathbf{DA}^{-1}\mathbf{B})^{-1}\mathbf{DA}^{-1}$ to arrive at (4.25). This indeed tells us that we can either use the Tx BF $\mathbf{G}_k$ or the Rx filter $\mathbf{F}_i$ to cancel the interference to the leakage channels.

4.5.3 Comparison of instantaneous CSIT and pathwise CSIT WSR at low SNR

We have the original WSR expression

$$WSR = \sum_{k=1}^{K} u_k \ln\det(\mathbf{I}+\mathbf{G}_k^H\mathbf{H}_{k,b_k}^H\mathbf{F}_k(\mathbf{F}_k^H\mathbf{R}_{\bar{k}}\mathbf{F}_k)^{-1}\mathbf{F}_k^H\mathbf{H}_{k,b_k}\mathbf{G}_k) \tag{4.26}$$

where $\mathbf{R}_{\bar{k}} = I$ for low SNR (or high SNR below) for both iCSIT and pwCSIT cases. At low SNR, the optimal Tx/Rx are matched filters (right or left singular vectors of $\mathbf{H}_{k,b_k}$). We get WSR at low SNR for iCSIT as follows

$$WSR = \sum_{k=1}^{K} u_k \ln\det(\mathbf{I}+\Sigma^2(\mathbf{H}_{k,b_k})P_k), \tag{4.27}$$

where $\Sigma(\mathbf{H}_{k,b_k})$ represents the singular value matrix of $\mathbf{H}_{k,b_k}$ and WSR at low SNR for pwCSIT can be simplified as

$$WSR = \sum_{k=1}^{K} u_k \ln\det(\mathbf{I}+\mathbf{H}_{r,k,b_k}^H\mathbf{H}_{r,k,b_k}\mathbf{D}_{k,b_k}^2 diag(\mathbf{H}_{t,k,b_k}^H\mathbf{Q}_k\mathbf{H}_{t,k,b_k})) \tag{4.28}$$

Optimizing (4.28) w.r.t $\mathbf{G}_k$ leads to the following solution for the BF $\mathbf{G}_k$ under pwCSIT.

$$\mathbf{G}_k' = \mathbf{V}_{1:d_k}\left(\mathbf{H}_{t,k,b_k} diag(\mathbf{H}_{r,k,b_k}^H\mathbf{H}_{r,k,b_k})\,\mathbf{D}_{k,b_k}^2\,\mathbf{H}_{t,k,b_k}^H.\right) \tag{4.29}$$

4.5.4 Comparison of instantaneous CSIT and pathwise CSIT WSR at high SNR

Starting from the WSR in (4.26), at high SNR the $\mathbf{G}$, $\mathbf{F}$ satisfy $\mathbf{F}_k^H\mathbf{H}_{k,b_i}\mathbf{G}_i = 0$, $i \neq k$ which reflects joint Tx/Rx ZF. . On the other hand, WSR at high SNR for pwCSIT behaves differently.

$$\mathbf{H}_{r,k,b_i} = \left[\underbrace{\mathbf{H}_{r,k,b_i,r}}_{\text{by UE}}\ \underbrace{\mathbf{H}_{r,k,b_i,t}}_{\text{by BS}}\right],\ \mathbf{H}_{t,k,b_i} = \left[\underbrace{\mathbf{H}_{t,k,b_i,r}}_{\text{by UE}}\ \underbrace{\mathbf{H}_{t,k,b_i,t}}_{\text{by BS}}\right] \tag{4.30}$$

where the underbraces indicate which nodes handle the interference of the indicated channel portions, and

$$
\begin{aligned}
\mathbf{F}_k &= P^{\perp}_{\mathbf{H}_{r,k,r}} \mathbf{H}_{r,k,b_k} (\mathbf{H}^H_{r,k,b_k} P^{\perp}_{\mathbf{H}_{r,k,r}} \mathbf{H}_{r,k,b_k})^{-\frac{1}{2}} \\
\mathbf{G}^{'}_k &= P^{\perp}_{\mathbf{H}_{t,k,t}} \mathbf{H}_{t,k,b_k} (\mathbf{H}^H_{t,k,b_k} P^{\perp}_{\mathbf{H}_{t,k,t}} \mathbf{H}_{t,k,b_k})^{-\frac{1}{2}} \\
WSR &= \sum_{k=1}^{K} u_k \ln\det(\mathbf{I} + \Sigma(\mathbf{S}_k^{\frac{1}{2}} \mathbf{D}^2_{k,b_k} diag\{\mathbf{T}_k\} \mathbf{S}_k^{\frac{1}{2}}) P_k) \\
\mathbf{S}_k &= \mathbf{H}^H_{t,k,b_k} P^{\perp}_{\mathbf{H}_{t,k,t}} \mathbf{H}_{t,k,b_k}, \ \mathbf{T}_k = \mathbf{H}^H_{r,k,b_k} P^{\perp}_{\mathbf{H}_{r,k,r}} \mathbf{H}_{r,k,b_k}.
\end{aligned}
\tag{4.31}
$$

In the pathwise case, the ZF task of all paths gets split between Tx and Rx, in which each does zero forcing of paths from either Tx or Rx side.

4.6 Simulation Results

Simulations are provided for the case of $C = 2$ cells, 2 users/cell, $L = 3$ paths in all channels, and varying Tx/Rx antenna numbers M, N. The expected sum rate is compared between the cases of perfect instantaneous CSIT (iCSIT) and (global) pathwise CSIT (pwCSIT). The loss is limited as soon as pathwise ZF is possible.

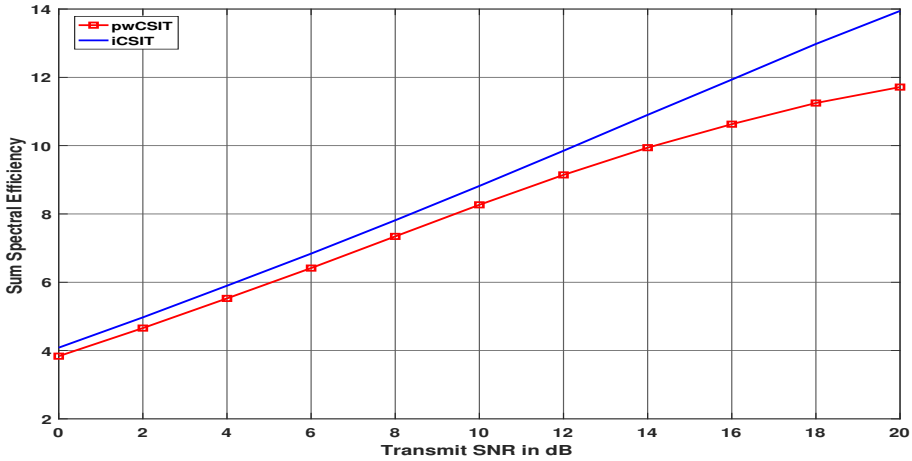


Figure 4.2: Expected sum rate comparison for $M = 3, N = 3$.

From, Figure 4.2 and Figure 4.3, we can make the following conclusions. The degraded performance of pwCSIT is due to the fact that the number of BS antennas are not enough to suppress or ZF the interference at high SNR. The interference subspace dimension sums up to 9 and the number of BS antennas used are less than 9.

In Figure 4.4, we have sufficient number of antennas (= 10) at the BS to do a ZF across the interfering subspace dimension which is 9. This accounts for the improved performance achieved by pwCSIT which overlaps with that of the iCSIT case. While in the other two cases above, there is sufficient gap between the performances of pwCSIT and iCSIT.

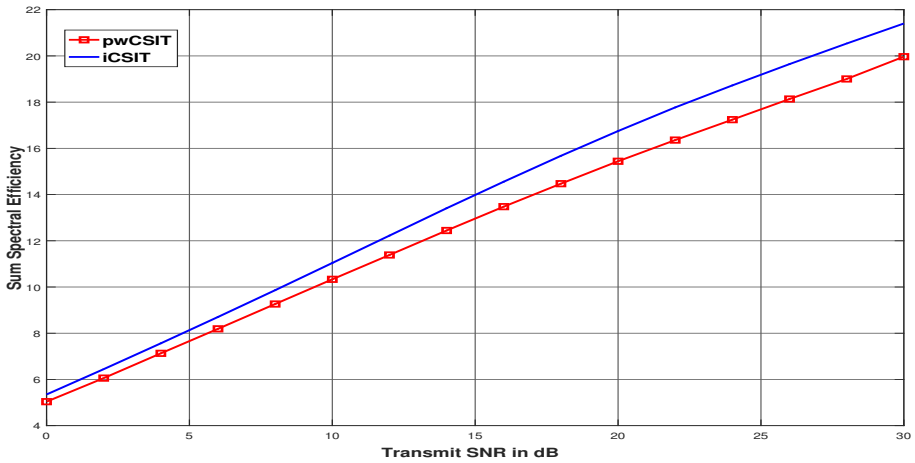


Figure 4.3: Expected sum rate comparison for $M = 4, N = 4$.

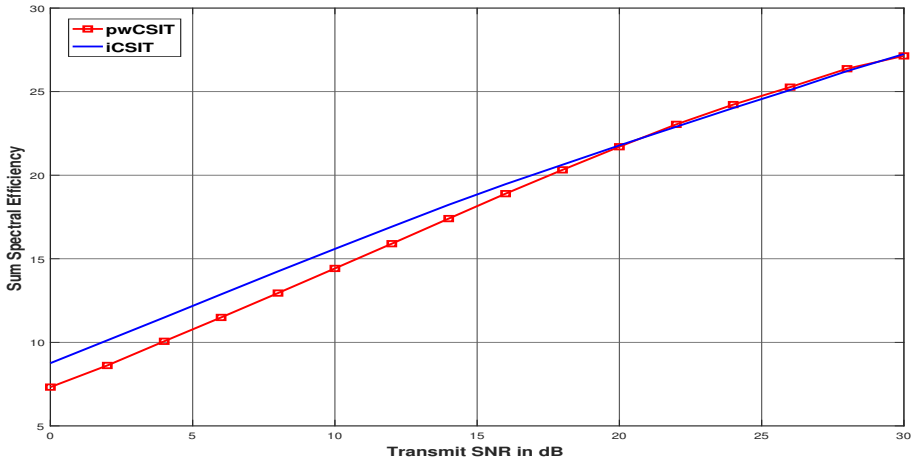


Figure 4.4: Expected sum rate comparison for $M = 10, N = 4$.

4.6.1 Conclusions and Perspectives

Conclusions and Perspectives 2

- In this chapter, we looked at robust BF design under partial CSIT for massive MIMO systems. In particular, we looked at a flat fading massive MIMO channel (a pathwise channel model parameterized by AoA, AoD and complex path coefficients). We assumed that the slow fading components (including AoA, AoD) and path amplitudes are perfectly known at the Tx side. Only the path phases are assumed to be unknown, which indeed correspond to the fast fading component. We denote the partial CSIT in this scenario as the pwCSIT.
- To optimize the BFs, we looked at an upper bound of the EWSR, which can be shown to be tight in the massive MIMO limit.

Conclusions and Perspectives 2 (cont.)

- Through Monte-Carlo simulations, we validated the performance of our pwCSIT based BF design. In the simulations, it became clear that when the number of antennas are high enough at the Tx and Rx side, the pwCSIT based design becomes close to optimal (optimal refers to the case of iCSIT based design). Intuitively, this means that the number of antennas are enough such that doing ZF across the various paths become optimal.
- An extension to the frequency selective case becomes straight forward, since in this case the pathwise channel model can be written as in the previous chapter, see for reference (3.6). The BF design gets separable across the subcarriers and at high SNR, pwCSIT based BF design converges to a ZF BF with ZF task gets split between Tx and Rx.
- We also remark that in order to extend the design here to the case when amplitude is also unknown becomes straightforward if we restrict to the case where complex path coefficients are i.i.d. $\mathscr{CN}(0,1)$ random variables.

Chapter 5

RATE SPLITTING FOR PILOT CONTAMINATION

5.1 Introduction

Massive MIMO (MaMIMO) is a wireless technology where the base stations (BSs) are equipped with a large number M of antennas to serve a multitude of single-antenna K user equipments (UEs) by spatial multiplexing [4]. The acquisition of channel state information (CSI) is the limiting factor in MaMIMO [4]. In a time-division duplex (TDD) mode, channel reciprocity allows to acquire all the necessary CSI for uplink (UL) and downlink (DL) transmissions from a finite number of UL pilot signals [4]. Thanks to the intense research performed over the last decade, MaMIMO is today a mature technology [74, 75], which has been adopted into the 5G NR standard [76].

One phenomenon that is tightly connected with MaMIMO is *pilot contamination*, which can be briefly explained as follows [4]. UEs that transmit the same pilot signal contaminate each others' channel estimates. This "pilot interference" not only reduces the CSI quality but also creates the so-called "coherent interference", which has been believed to fundamentally limit the spectral efficiency (SE) of MaMIMO, even when $M \to \infty$ [4, 74]. Recently, [77] showed that with optimal signal processing and spatially correlated channels, the SE increases without bound as $M \to \infty$ while K is fixed. The fact that there is no fundamental SE limit does not imply that the pilot contamination effect disappears; there is still an SE loss caused by estimation errors and interference rejection [78]. The aim of this chapter is to deal with this effect for a finite M.

Observe that, when the estimation error variance decays with the signal-to-noise-ratio (SNR) as $\mathcal{O}(\text{SNR}^{-\delta})$ for some $0 \leq \delta < 1$, conventional precoding techniques result in a sum degrees of freedom (DoF) of $K\delta$. This in turn reveals that as $\delta \to 0$ (implies constant channel estimation error), the system becomes interference limited. A possible solution to this issue is to take a rate splitting (RS) approach [79] that splits the UEs' messages into common and private parts, encode the commmon parts into a common stream, and private parts into private streams and superpose in a non-orthogonal manner the common stream on top of all private streams. The common stream is drawn from a codebook shared by all UEs and is intended to one only, but is decodable by all UEs. On the other hand, the private streams are to be decoded by the corresponding UEs only. The sum DoF achieved by RS in the DL is $1 + (K-1)\delta$ [80], which is higher than $K\delta$ and matches the upper bound obtained from the Aligned Image Sets in [81]. Interestingly, RS not only achieves the optimal sum-DoF but the entire DoF region of the K-UE channel with imperfect CSI [82].

Motivated by the above results, the design and optimization of RS at finite values of SNR has been investigated and was found to provide significant benefits in the DL with imperfect CSI,

compared to multi-user MIMO and NOMA [80, 83, 84], but also to Dirty Paper Coding [85]. The application of RS to an FDD MaMIMO system has been investigated in [86, 87]. Particularly, [86] shows that a two-layer RS architecture, so-called hierarchical RS (HRS), can bring significant benefits in MaMIMO.

5.1.1 Summary of this Chapter

- In this chapter, we focus on a TDD single-cell MaMIMO network and assume that all the UEs use the same pilot signal for channel estimation.
- Novel expressions for the SE achieved in the DL by a single-layer RS strategy are derived by applying the hardening bound to both common and private messages [74].
- A maximum ratio (MR) precoding scheme is used for private streams while a precoder based on a weighted combination of the channel estimates of all UEs is adopted for the common stream.
- A novel algorithm is proposed to allocate the power among the common and private streams.

5.2 System model

We consider a single-cell MaMIMO network where the BS is equipped with M antennas and serves K UEs. We denote $\mathbf{h}_i \in \mathbb{C}^M$ the channel from UE i to the BS, and consider a correlated Rayleigh block fading model $\mathbf{h}_i \sim \mathcal{CN}(\mathbf{0}, \mathbf{R}_i)$ where $\mathbf{R}_i \in \mathbb{C}^{M\times M}$ is the covariance matrix [75, Sec. 2.2]. The Gaussian distribution is used to model the small-scale fading variations, while $\mathbf{R}_i$ describes the macroscopic propagation characteristics. The normalized trace $\beta_i = \frac{1}{M}\text{tr}(\mathbf{R}_i)$ is the average channel gain from the BS to UE i.

The UEs are perfectly synchronized and operate according to a TDD protocol with a data transmission phase and a pilot phase for channel estimation [75]. We consider the standard block fading TDD protocol in which each coherence block consists of τ channel uses, whereof τ_p are used for UL pilots, τ_u for UL data, and τ_d for DL data, with $\tau = \tau_p + \tau_u + \tau_d$. Only the DL is considered in this chapter, i.e., $\tau_u = 0$.

5.2.1 Assumptions on the user channel

In this subsection, we describe certain assumptions on the channel between UE and BS.

- The channel covariance matrix $\mathbf{R}_i$ for any user i is assumed to be perfectly known at the BS.
- For the SINR expressions derived here to be valid, the matrix $\mathbf{R}_i$ should be invertible for all i.

5.2.2 Channel estimation

We assume that a single pilot sequence of length τ_p is used. For a total uplink pilot power of ρ_{tr} per UE, the BS obtains the MMSE estimate of $\mathbf{h}_i$ as

$$\widehat{\mathbf{h}}_i = \mathbf{R}_i\mathbf{Q}^{-1}\Big(\sum_{k=1}^{K}\mathbf{h}_k + \frac{1}{\sqrt{\rho_{\text{tr}}}}\mathbf{n}_i\Big) \sim \mathcal{CN}(\mathbf{0}, \mathbf{\Phi}_i) \tag{5.1}$$

$$\gamma_{k,c} = \frac{\rho_c |\mathbb{E}\{\mathbf{h}_k^{\mathrm{H}} \mathbf{w}_c\}|^2}{\sum_{i=1}^{K} \rho_i \mathbb{E}\{|\mathbf{h}_k^{\mathrm{H}} \mathbf{w}_i|^2\} + \rho_c \Big(\mathbb{E}\{|\mathbf{h}_k^{\mathrm{H}} \mathbf{w}_c|^2\} - |\mathbb{E}\{\mathbf{h}_k^{\mathrm{H}} \mathbf{w}_c\}|^2 \Big) + \sigma^2} \tag{11}$$

$$\gamma_{k} = \frac{\rho_k |\mathbb{E}\{\mathbf{h}_k^{\mathrm{H}} \mathbf{w}_k\}|^2}{\sum_{i=1}^{K} \rho_i \mathbb{E}\{|\mathbf{h}_k^{\mathrm{H}} \mathbf{w}_i|^2\} - \rho_k |\mathbb{E}\{\mathbf{h}_k^{\mathrm{H}} \mathbf{w}_k\}|^2 + \rho_c \Big(\mathbb{E}\{|\mathbf{h}_k^{\mathrm{H}} \mathbf{w}_c|^2\} - |\mathbb{E}\{\mathbf{h}_k^{\mathrm{H}} \mathbf{w}_c\}|^2 \Big) + \sigma^2} \tag{12}$$

where $\mathbf{n}_i \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_M)$ is noise, $\mathbf{\Phi}_i = \mathbf{R}_i \mathbf{Q}^{-1} \mathbf{R}_i$ and $\mathbf{Q} = \sum_{k=1}^{K} \mathbf{R}_k + \frac{1}{\rho_{\mathrm{tr}}} \mathbf{I}_M$. The estimation error $\widetilde{\mathbf{h}}_i = \mathbf{h}_i - \widehat{\mathbf{h}}_i \sim \mathcal{CN}(\mathbf{0}, \mathbf{R}_i - \mathbf{\Phi}_i)$ is independent of $\widehat{\mathbf{h}}_i$. The mutual interference generated by the pilot-sharing UEs is known as pilot contamination and has two main consequences in the channel estimation process. The first is the reduced estimation quality, whereas the second is that the estimates $\{\widehat{\mathbf{h}}_i\}$ become correlated. If $\mathbf{R}_k$ is invertible, we have that [75, Sec. 3.2]

$$\widehat{\mathbf{h}}_i = \mathbf{R}_i \mathbf{R}_k^{-1} \widehat{\mathbf{h}}_k \tag{5.2}$$

from which it follows that $\mathbb{E}\{\widehat{\mathbf{h}}_i \widehat{\mathbf{h}}_k^{\mathrm{H}}\} = \mathbf{R}_i \mathbf{Q}^{-1} \mathbf{R}_k$.

5.2.3 Rate Splitting in Downlink transmissions

The RS scheme is used in the DL for transmission. The message intended to UE k is split into two parts, $W_k = (W_{k0}, W_{k1})$. We assume that $W_{k0} \in \mathcal{W}_{k0}$ represents the common part and $W_{k1} \in \mathcal{W}_{k1}$ is the private part. All the common parts are packed into one common message, $W_c = (W_{k0}, ..., W_{K0}) \in \mathcal{W}_c$, which is encoded into a common stream ς_c using a common codebook. The private message W_{k1} is encoded in the conventional manner into the private stream ς_k. The resulting transmitted DL signal is:

$$\mathbf{x} = \underbrace{\mathbf{w}_c \varsigma_c}_{\text{Common message}} + \sum_{i=1}^{K} \underbrace{\mathbf{w}_i \varsigma_i}_{\text{Private messages}} \tag{5.3}$$

where $\varsigma_i \sim \mathcal{CN}(0, \rho_i)$ is assigned to a precoding vector $\mathbf{w}_i \in \mathbb{C}^M$ that determines the spatial directivity of the transmission and satisfies $\mathbb{E}\{\|\mathbf{w}_i\|^2\} = 1$ so that ρ_i represents the average transmit power of UE $\forall i$. Similarly, $\varsigma_c \sim \mathcal{CN}(0, \rho_c)$ denotes the common message, which is assigned to a precoding vector $\mathbf{w}_c \in \mathbb{C}^M$ with $\mathbb{E}\{\|\mathbf{w}_c\|^2\} = 1$ so that ρ_c represents its average transmit power. We assume that

$$\rho_c + \sum_{i=1}^{K} \rho_i \leq \rho_T \tag{5.4}$$

where ρ_T is the total transmit power in the DL. The received signal $y_k \in \mathbb{C}$ at UE k is given by

$$y_k = \mathbf{h}_k^{\mathrm{H}} \mathbf{w}_c \varsigma_c + \mathbf{h}_k^{\mathrm{H}} \mathbf{w}_k \varsigma_k + \sum_{i=1, i \neq k}^{K} \mathbf{h}_k^{\mathrm{H}} \mathbf{w}_i \varsigma_i + n_k \tag{5.5}$$

where $n_k \sim \mathcal{CN}(0, \sigma^2)$ is the receiver noise. At each UE k, the common stream is first decoded into $\widehat{W}_c$, by treating the interference from the private streams as noise. Then, successive interference cancellation (SIC) is performed, which removes the common message part from the received signal. Further, the private stream ς_k is decoded into $\widehat{W}_{k1}$ by treating the intra-cell interference as noise. UE k reconstructs the transmitted message by extracting $\widehat{W}_{k0}$ from $\widehat{W}_c$. Further, combining with the decoded private stream to form $\widehat{W}_k = (\widehat{W}_{k0}, \widehat{W}_{k1})$.

5.2.4 Spectral efficiency

Characterizing the SE in the DL is hard since it is unclear how UE k should best estimate the effective precoded channels $\mathbf{h}_k^{\mathrm{H}}\mathbf{w}_c$ and $\mathbf{h}_k^{\mathrm{H}}\mathbf{w}_k$ that are needed for decoding the common signal ς_c and the private signal ς_k. A common approach in classical MaMIMO is to resort to the *hardening bound* [75, Sec. 4.3]. This bound relies on the assumption that the deterministic average precoded channels $\mathbb{E}\{\mathbf{h}_k^{\mathrm{H}}\mathbf{w}_c\}$ and $\mathbb{E}\{\mathbf{h}_k^{\mathrm{H}}\mathbf{w}_k\}$ are known at UE k. The received signals for the common and private messages can then be expressed as

$$\begin{aligned} y_{k,c} = \mathbb{E}\{\mathbf{h}_k^{\mathrm{H}}\mathbf{w}_c\}\varsigma_c + \left(\mathbf{h}_k^{\mathrm{H}}\mathbf{w}_c - \mathbb{E}\{\mathbf{h}_k^{\mathrm{H}}\mathbf{w}_c\}\right)\varsigma_c \\ + \sum_{i=1}^{K} \mathbf{h}_k^{\mathrm{H}}\mathbf{w}_i\varsigma_i + n_k \end{aligned} \tag{5.6}$$

and (after SIC)

$$\begin{aligned} y_{k,p} = \mathbb{E}\{\mathbf{h}_k^{\mathrm{H}}\mathbf{w}_k\}\varsigma_k + \left(\mathbf{h}_k^{\mathrm{H}}\mathbf{w}_k - \mathbb{E}\{\mathbf{h}_k^{\mathrm{H}}\mathbf{w}_k\}\right)\varsigma_k \\ + \left(\mathbf{h}_k^{\mathrm{H}}\mathbf{w}_c - \mathbb{E}\{\mathbf{h}_k^{\mathrm{H}}\mathbf{w}_c\}\right)\varsigma_c + \sum_{i=1, i\neq k}^{K} \mathbf{h}_k^{\mathrm{H}}\mathbf{w}_i\varsigma_i + n_k. \end{aligned} \tag{5.7}$$

The following bounds can be computed.

Lemma 2. *Achievable rates for the common and private messages of UE k can be computed as*

$$\mathsf{SE}_{k,c} = \frac{\tau_d}{\tau}\log_2(1+\gamma_{k,c}) \tag{5.8}$$

and

$$\mathsf{SE}_k = \frac{\tau_d}{\tau}\log_2(1+\gamma_k) \tag{5.9}$$

with $\gamma_{k,c}$ and γ_k given by (11) *and* (12)*. The expectations are computed over channel realizations.*

It can be proved from (5.6) and (5.7) by using standard results in MaMIMO (example, [75, App. C.3.6]). Here ends the proof. The achievable rate of the common message is defined as

$$\mathsf{SE}_c = \frac{\tau_d}{\tau}\log(1+\gamma_{l_{\min},c}) \tag{5.13}$$

where

$$l_{\min} = \arg\min_k \gamma_{k,c} \tag{5.14}$$

Observe that the above achievable rates can be utilized along with any precoding scheme. Moreover, each of the expectations in $\gamma_{k,c}$ and γ_k can be computed separately by means of Monte Carlo simulations. Closed forms will be provided next for the proposed precoding schemes.

5.3 Power optimization and precoding design

A common and popular choice for $\mathbf{w}_k$ is MR precoding, defined as

$$\mathbf{w}_k^{\mathsf{MR}} = \frac{\widehat{\mathbf{h}}_k}{\sqrt{\mathrm{E}\{|\widehat{\mathbf{h}}_k|^2\}}} = \frac{\widehat{\mathbf{h}}_k}{\sqrt{\mathrm{tr}\{\Phi_k\}}} \tag{5.15}$$

which has low computational complexity and allows to compute some of the expectations in closed form. MR precoding is also called as conjugate beamforming (since the BF is proportional to the conjuagte of the estimated channel) [88]. For an operating point which yields low spectral efficiency and high energy efficiency, MR precoding outperforms ZF precoding. Particularly, we have that (example, [75, App. C.3.7])

$$|\mathbb{E}\{\mathbf{h}_k^{\mathrm{H}}\mathbf{w}_k^{\mathrm{MR}}\}|^2 = \mathrm{tr}\{\Phi_k\} \tag{5.16}$$

$$\mathbb{E}\{|\mathbf{h}_k^{\mathrm{H}}\mathbf{w}_i^{\mathrm{MR}}|^2\} = \frac{\mathrm{tr}\{\mathbf{R}_k\Phi_i\} + \left|\mathrm{tr}\{\mathbf{R}_k\mathbf{Q}^{-1}\mathbf{R}_i\}\right|^2}{\mathrm{tr}\{\Phi_i\}}. \tag{5.17}$$

In the remainder, we assume that MR precoding is used for private messages. Next, we look for the transmit powers that maximize the sum SE of the network and design the precoding vector for the common message.

5.3.1 Power optimization

From the above section, the sum SE, for any given precoding scheme, can be computed as:

$$\mathsf{SE} = \mathsf{SE}_c + \sum_{k=1}^{K} \mathsf{SE}_k \tag{5.18}$$

where SE_k and SE_c are given in (5.8) and (5.13), respectively. The power allocation problem can thus be formulated as:

$$\max_{\{\rho_c \geq 0, \boldsymbol{\rho} \geq \mathbf{0}\}} \quad \mathsf{SE}_c(\rho_c, \boldsymbol{\rho}) + \sum_{k=1}^{K} \mathsf{SE}_k(\rho_c, \boldsymbol{\rho}) \tag{5.19}$$

$$\text{s.t.} \qquad \rho_c + \sum_{i=1}^{K} \rho_i \leq \rho_T \tag{5.20}$$

with $\boldsymbol{\rho} = [\rho_1, \ldots, \rho_K]^{\mathrm{T}}$. Finding the solution to the above problem is a challenge since it is not in a convex form. A possible way out consists in using the method in [30], and linearize the sum SE in (5.18) using a first order Taylor series approximation. The optizimation is then carried out by adopting an iterative approach in which the variables ρ_c and $\{\rho_i : i = 1, \ldots, K\}$ are alternatively optimized. In Appendix F, it is shown that at iteration t the powers must be updated as follows

$$\rho_k^{(t)} = \left(\frac{1}{\mu^{(t)} + \sigma_k^{(2,t)}} - \frac{1}{\sigma_k^{(1,t)}} \right)^+ \tag{5.21}$$

and

$$\rho_c^{(t)} = \left(\frac{1}{\mu^{(t)} + \sigma_c^{(2,t)}} - \frac{1}{\sigma_c^{(1,t)}} \right)^+ \tag{5.22}$$

where $(x)^+ = \max(x, 0)$ and the quantities $\{\sigma_k^{(1,t)}, \sigma_c^{(1,t)}\}$ and $\{\sigma_k^{(2,t)}, \sigma_c^{(2,t)}\}$ are defined in Appendix F. The former represent the signal powers of private and common messages at iteration t, respectively, while the latter can be interpreted as the corresponding leakage powers. This is why (5.21) and (5.22) are called interference leakage-aware water-filling (ILA-WF) power allocations [46]. Note that the Lagrange multiplier $\mu^{(t)}$ needs to satisfy the power constraint in (5.20) and can be

Algorithm 9: ILA-WF power allocation

initialize $t=0$ and $\rho_c^{(0)}=0$ (no RS) and $\rho_k^{(0)}=\rho_T/K$. Also, $\mu^{(0)}=\frac{1}{2}(\mu_u^{(0)}+\mu_l^{(0)})$ with $\mu_u^{(0)}=10^5$ (or some very large value) and $\mu_l^{(0)}=0$.
repeat
  for $k=1$ to K **do**
    compute $\sigma_k^{(1,t)}$ and $\sigma_k^{(2,t)}$
    use μ^t to update p_k^t in (5.21)
  end for
  compute $\sigma_c^{(1,t)}$ and $\sigma_c^{(2,t)}$
  use μ^t to update p_c^t in (5.22)
  if $\rho_c^{(t)}+\sum_k \rho_k^{(t)} > \rho_T$ **then**
    $\mu_l^{(t+1)}=\mu^{(t)}$, $\mu_u^{(t+1)}=\mu_u^{(t)}$
  else
    $\mu_u^{(t+1)}=\mu^t$, $\mu_l^{(t+1)}=\mu_l^{(t)}$
  end if
  update $\mu^{(t+1)}=\frac{\mu_u^{(t+1)}+\mu_l^{(t+1)}}{2}$
  update $t=t+1$
until convergence

computed by a bisection method [39]. The entire procedure is summarized through Algorithm 1.

As done for $\gamma_{k,c}$ and γ_k, we observe that all the quantities involved in the computation of $\{\sigma_k^{(1)},\sigma_c^{(1)}\}$ and $\{\sigma_k^{(2)},\sigma_c^{(2)}\}$ are deterministic and can be computed by means of Monte Carlo simulations for any choice of the precoding scheme for the common message. Closed form expressions are provided below for a MR-inspired precoding scheme.

5.3.2 Precoding design for common message

The optimal design of the precoding vector $\mathbf{w}_c$ for the common message requires to solve a multi-objective problem involving $\gamma_{l_{\min},c}$ and $\{\gamma_i : \forall i\}$. To overcome this issue, we assume that the difference $\mathbb{E}\{|\mathbf{h}_k^{\mathrm{H}}\mathbf{w}_c|^2\}-|\mathbb{E}\{\mathbf{h}_k^{\mathrm{H}}\mathbf{w}_c\}|^2$ in (11) is small so that it can be neglected. The precoding vector is then suboptimally selected as the solution to the following problem:

$$\max_{\mathbf{w}_c}\min_k \pi_k|\mathbb{E}\{\mathbf{h}_k^{\mathrm{H}}\mathbf{w}_c\}|^2 \quad \text{s.t.} \quad \mathbb{E}\{\|\mathbf{w}_c\|^2\}=1 \tag{5.23}$$

where

$$\pi_k=\frac{1}{\sum_{i=1}^{K}\rho_i\mathbb{E}\{|\mathbf{h}_k^{\mathrm{H}}\mathbf{w}_i|^2\}+\sigma^2}. \tag{5.24}$$

Following [86], we heuristically select $\mathbf{w}_c$ as a linear combination of the estimated channel vectors $\{\widehat{\mathbf{h}}_i : \forall i\}$:

$$\mathbf{w}_c=\alpha\sum_{i=1}^{K} a_i\widehat{\mathbf{h}}_i \tag{5.25}$$

where α is needed to satisfy the constraint $\mathbb{E}\{\|\mathbf{w}_c\|^2\} = 1$. Plugging (5.25) into $\mathbb{E}\{\mathbf{h}_k^{\mathrm{H}}\mathbf{w}_c\}$, we may rewrite (5.23) as:

$$\max_{\{a_i\}} \min_k \pi_k \left| \sum_{i=1}^{K} a_i \mathrm{tr}\{\mathbf{R}_i \mathbf{Q}^{-1} \mathbf{R}_k\} \right|^2 \tag{5.26}$$

where we have neglected the scaling factor α^2. We now observe that (5.26) can be reformulated as a geometric programming problem [39]:

$$\max_{t>0} \ t, \ \text{s.t.} \ \mathbf{a}^{\mathrm{T}}\mathbf{u}_i \geq t, \ \forall \ i = 1, .., K \tag{5.27}$$

where we have defined $\mathbf{a} = [a_1, ..., a_K]^{\mathrm{T}}$ and $\mathbf{u}_i = [u_i(1), \ldots, u_i(K)]^{\mathrm{T}}$ with entries $u_i(k) = \mathrm{tr}\{\mathbf{R}_i \mathbf{Q}^{-1} \mathbf{R}_k\}$. Once the solution $\mathbf{a}^\star$ to (5.27) is computed, the optimal $\mathbf{w}_c^\star$ is obtained as:

$$\mathbf{w}_c^\star = \frac{\sum_{i=1}^{K} a_i^\star \widehat{\mathbf{h}}_i}{\sqrt{\sum_{i=1}^{K}\sum_{j=1}^{K} a_i^\star a_j^\star \mathrm{tr}\{\mathbf{R}_i \mathbf{Q}^{-1} \mathbf{R}_j\}}}. \tag{5.28}$$

The expectations that depend on $\mathbf{w}_c^\star$ can be computed in closed form as follows. By using (5.2) into (5.28) yields

$$\mathbb{E}\{\mathbf{h}_k^{\mathrm{H}}\mathbf{w}_c^\star\} = \frac{\sum_{i=1}^{K} a_i^\star \mathrm{tr}\{\mathbf{R}_i \mathbf{Q}^{-1} \mathbf{R}_k\}}{\sqrt{\sum_{i=1}^{K}\sum_{j=1}^{K} a_i^\star a_j^\star \mathrm{tr}\{\mathbf{R}_i \mathbf{Q}^{-1} \mathbf{R}_j\}}}. \tag{5.29}$$

To compute $\mathbb{E}\{|\mathbf{h}_k^{\mathrm{H}}\mathbf{w}_c^\star|^2\}$, observe that it can be rewritten as

$$\mathbb{E}\{|\mathbf{h}_k^{\mathrm{H}}\mathbf{w}_c^\star|^2\} = \frac{1}{\sum_{i=1}^{K}\sum_{j=1}^{K} a_i^\star a_j^\star \mathrm{tr}\{\mathbf{R}_i \mathbf{Q}^{-1} \mathbf{R}_j\}} \times$$

$$\left(\sum_{i=1}^{K} (a_i^\star)^2 \mathbb{E}\{|\widehat{\mathbf{h}}_i^{\mathrm{H}}\mathbf{h}_k|^2\} + \sum_{i=1}^{K}\sum_{j=1, j\neq i}^{K} a_i^\star a_j^\star \mathbb{E}\{\mathbf{h}_k^{\mathrm{H}}\widehat{\mathbf{h}}_i\widehat{\mathbf{h}}_j\mathbf{h}_k\} \right). \tag{5.30}$$

The first term in (5.30) becomes (example, [75, Eq. (C.65)])

$$\mathbb{E}\{|\widehat{\mathbf{h}}_i^{\mathrm{H}}\mathbf{h}_k|^2\} = \mathrm{tr}\{\mathbf{R}_k \Phi_i\} + \left|\mathrm{tr}\{\mathbf{R}_k \mathbf{Q}^{-1} \mathbf{R}_i\}\right|^2 \tag{5.31}$$

while the second one in (5.30) reduces to

$$\mathbb{E}\{\mathbf{h}_k^{\mathrm{H}}\widehat{\mathbf{h}}_i\widehat{\mathbf{h}}_j^{\mathrm{H}}\mathbf{h}_k\} \overset{(a)}{=} \mathbb{E}\{\mathbf{h}_k^{\mathrm{H}}\widehat{\mathbf{h}}_i\widehat{\mathbf{h}}_i^{\mathrm{H}}\mathbf{R}_i^{-1}\mathbf{R}_j\mathbf{h}_k\} \tag{5.32}$$

$$\overset{(b)}{=} \mathrm{tr}\left\{\mathbf{R}_i^{-1}\mathbf{R}_j \mathbb{E}\{\widehat{\mathbf{h}}_k\widehat{\mathbf{h}}_k^{\mathrm{H}}\widehat{\mathbf{h}}_i\widehat{\mathbf{h}}_i^{\mathrm{H}}\}\right\} + \mathrm{tr}\left\{\mathbf{R}_i^{-1}\mathbf{R}_j \mathbb{E}\{\widetilde{\mathbf{h}}_k\widetilde{\mathbf{h}}_k^{\mathrm{H}}\}\mathbb{E}\{\widehat{\mathbf{h}}_i\widehat{\mathbf{h}}_i^{\mathrm{H}}\}\right\} \tag{5.33}$$

$$\overset{(c)}{=} \mathrm{tr}\left\{\mathbf{R}_i^{-1}\mathbf{R}_j \mathbb{E}\{\widehat{\mathbf{h}}_k\widehat{\mathbf{h}}_k^{\mathrm{H}}\widehat{\mathbf{h}}_i\widehat{\mathbf{h}}_i^{\mathrm{H}}\}\right\} + \mathrm{tr}\left\{\mathbf{R}_i^{-1}\mathbf{R}_j (\mathbf{R}_k - \Phi_k)\Phi_i\right\} \tag{5.34}$$

where (a) uses $\widehat{\mathbf{h}}_j = \mathbf{R}_j \mathbf{R}_i^{-1} \widehat{\mathbf{h}}_i$ (as it follows from (5.2)), (b) follows from $\mathbf{h}_k = \widetilde{\mathbf{h}}_k + \widehat{\mathbf{h}}_k$ and the independence between the estimate $\widehat{\mathbf{h}}_k$ and estimation error $\widetilde{\mathbf{h}}_k$, whereas (c) uses $\mathbb{E}\{\widetilde{\mathbf{h}}_k \widetilde{\mathbf{h}}_k^{\mathrm{H}}\} \mathbb{E}\{\widehat{\mathbf{h}}_i \widehat{\mathbf{h}}_i^{\mathrm{H}}\} = (\mathbf{R}_k - \Phi_k)\Phi_i$. In Appendix G, it is shown that

$$\begin{aligned} \mathbb{E}\{\widehat{\mathbf{h}}_k \widehat{\mathbf{h}}_k^{\mathrm{H}} \widehat{\mathbf{h}}_i \widehat{\mathbf{h}}_i^{\mathrm{H}}\} &= \mathrm{tr}\{\mathbf{B}_{ik}\}\Phi_k \\ &\quad + \Phi_k^{1/2} \big(\mathrm{diag}(\mathbf{B}_{ik}) + \mathbf{B}_{ik} \big) (\Phi_k^{1/2})^{\mathrm{H}} \end{aligned} \tag{5.35}$$

where $\mathbf{B}_{ik} = (\Phi_k^{1/2})^{\mathrm{H}} \mathbf{R}_i \mathbf{R}_k^{-1} \Phi_k^{1/2}$ and $\mathrm{diag}(\cdot)$ indicates the main diagonal of the enclosed matrix.

Note that, by using the above expressions and those in (5.16) and (5.17), we can eventually compute in closed form all the expectations involved in (11) and (12).

5.4 Simulation Results

To quantify the SE that can be achieved in MaMIMO with RS, we consider a cell of size 250 m $\times$ 250 m. The UL pilot power is $\rho_{\mathrm{tr}} = 20$ dBm, whereas the noise power in UL and DL is $\sigma^2 = -94$ dBm. The samples per coherence block are $\tau = 200$ with $\tau_p = 10$. Each BS is equipped with a uniform linear array with half-wavelength antenna spacing. Each channel consists of $S = 6$ scattering clusters, which are modeled by the Gaussian local scattering model [75, Sec. 2.6]. Hence, the (m_1, m_2)th element of $\mathbf{R}_i$ is

$$\begin{aligned} &[\mathbf{R}_i]_{m_1,m_2} = \beta_i \times \\ &\frac{1}{S} \sum_{s=1}^{S} e^{\mathrm{j}\pi(m_1 - m_2)\sin(\varphi_{i,s})} e^{-\frac{\sigma_\varphi^2}{2}(\pi(m_1 - m_2)\cos(\varphi_{i,s}))^2} \end{aligned} \tag{5.36}$$

where β_i is the large-scale fading coefficient given by (in dB)

$$\beta_i|_{\mathrm{dB}} = -34.53 - 38 \log_{10}\left(\frac{d_i}{1\,\mathrm{km}}\right) + F_i \tag{5.37}$$

with UEs being placed uniformly at random and d_i (>= 35 m) represents the distance of UE i from the BS. $F_i \sim \mathcal{N}(0, 10)$ is the logarithm of the shadow fading between UE i and BS. Also, let φ_i be the geographical angle to UE i as seen from the BS. Cluster s is characterized by the randomly generated nominal angle-of-arrival $\varphi_{i,s} \sim \mathcal{U}[\varphi_i - 40^\circ, \varphi_i + 40^\circ]$ and the angles of the multipath components are Gaussian distributed around the nominal angle with standard deviation $\sigma_\varphi^2 = 10^\circ$.

Fig. 5.1 plots the sum SE as a function of the total transmit power defined as ρ_T (in dBm) with $M = 100$ and $K = 10$. Comparisons are made with a classical MaMIMO system with MR precoding and power allocated through Algorithm 1 with ρ_c fixed to 0. As seen, RS improves the sum SE significantly for values of ρ_T higher than 5 dBm. Moreover, the sum SE with RS does not saturate at high ρ_T values. This in contrast to what happens without RS, due to pilot contamination.

Fig. 5.2 illustrates the sum SE as a function of number of antennas, M, with $K = 10$ and transmit power $\rho_T = 20$ dBm. We observe that the RS scheme does help to mitigate the pilot contamination effect for a finite number of antennas.

Finally, in Fig. 5.3 we report the sum SE as a function of K with $M = 100$ and $\rho_T = 20$ dBm. As K increases, the gain provided by RS decreases. The larger K, the lower the common rate since the common message has to be decoded by all UEs. This issue can be solved by using HRS approach as in [86]; this is an interesting topic left for future work.

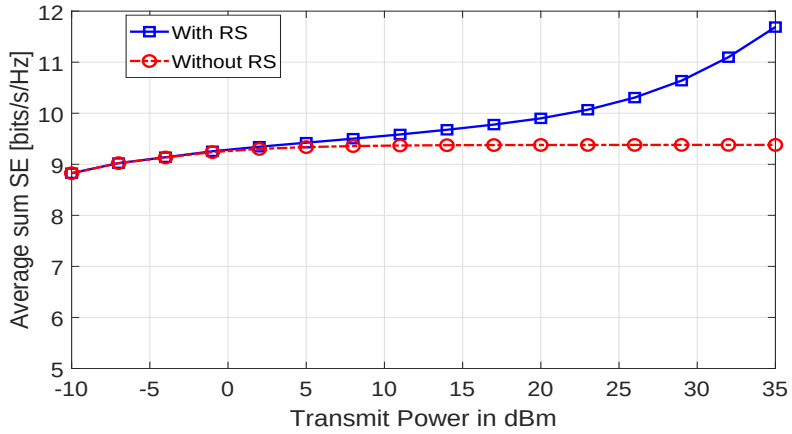


Figure 5.1: Sum SE versus transmit power, with $M = 100$ and $K = 10$.

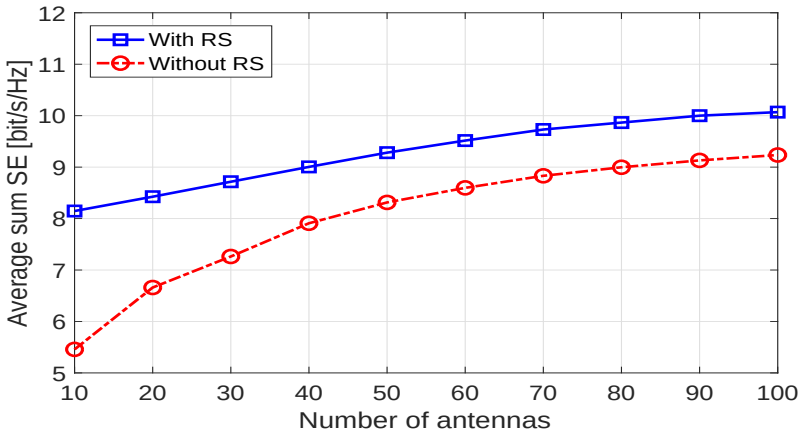


Figure 5.2: Sum SE versus number of antennas with $K = 10$ and $\rho_T = 20$ dBm.

5.5 Concluding Remarks

Remarks 3

- This chapter focused on a single-cell MaMIMO system in which all the UEs use the same pilot signal in the training phase. To deal with the reduced channel estimation quality, caused by pilot contamination, a single layer RS approach was proposed and shown to improve the SE at high SNR values. The results of this section appear in the paper [89].
- As the number of users (K) increases, the gain provided by RS decreases. The larger K, the lower the common rate since the common message has to be decoded by all UEs. We conjecture that a two layer RS would be a solution in this case and remains to be checked.
- However, we remark that much remains to be done, for example, extension of the current work to a multi-cell setting and the design of an efficient RS message scheme to mitigate the inter-cell and intra-cell interference.

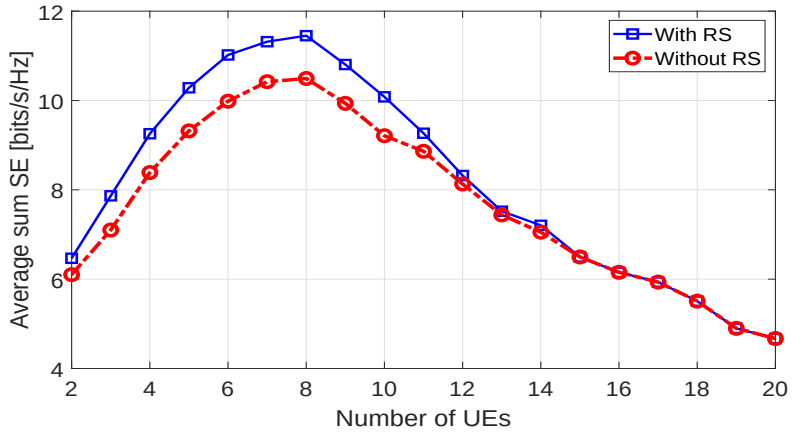


Figure 5.3: Sum SE versus number of UEs with $M = 100$ and $\rho_T = 20$ dBm.

Remarks 3 (cont.)

Part III

Stochastic Geometry based Large System Analysis

Chapter 6

ASYMPTOTIC ANALYSIS OF REDUCED ORDER ZERO FORCING BEAMFORMING

6.1 Introduction

Optimal linear transmitter beamformers in multi-antenna multi-user systems are of the Minimum Mean Squared Error (MMSE) type (dual uplink MMSE receivers). MMSE designs make an optimal compromise between noise enhancement and interference suppression and reduce to matched filters at low SNR and zero-forcing at high SNR. We consider a realistic scenario of user channels of varying attenuation and constrain the beamformers to either zero-force or ignore each interference term. This leads to a reduced-order zero-forcing (RO-ZF) design in which the number of interference sources being zero-forced increases with SNR. We apply a simple large systems analysis (applicable to Massive MIMO) to determine the asymptotic performance of RO-ZF designs, determine the optimal ZF orders, and compare to optimal and ZF linear and Dirty Paper Coding (DPC) designs. RO-ZF designs lead to variable reductions of computational complexity and channel state information (CSI) requirements (esp. in future multi-cell extensions), both important considerations in Massive MIMO systems.

Massive MIMO [90] which utilizes large number of antennas at the base station (BS) offers immense possibilities for increased system capacity. Multi-user MIMO (MU-MIMO) systems requires the global knowledge of the CSI at the Tx (CSIT) which is more difficult to acquire than CSI at the Rx. However, this leads to increased computational complexity owing to the large number of antennas. Recently, a number of research works have proposed to exploit the channel hardening in Massive MIMO (MaMIMO) to reduce global instantaneous CSIT requirements to local instantaneous CSIT plus global statistical CSIT [91]. Channel hardening occurs when the number of antennas at the BS are very high such that a fading channel behaves as if the effect of the randomness in the channel to spectral efficiency will be negligible. Extensive work on BF designs for BC (broadcast channel) or IBC (Interfering BC) with perfect or partial CSIT can be found in [11, 62, 70, 92, 93].

A significant contribution for large system analysis in MaMIMO systems appeared in [14]. It allows to compute deterministic (instead of fast fading channel dependent) expressions for various scalar quantities, facilitating the analysis and design of wireless systems. E.g. it may allow to conduct the performance analysis without computing explicit beamformers. Through large system analysis, [14] compute the optimal regularization factor in Regularized ZF (R-ZF) BF, both with perfect and partial CSIT. A little known extension appeared in [15] for weighted

Sum MSE (WSMSE) based optimal beamformers, but only for the perfect CSIT MISO (Multiple-Input Single-Output) BC case. Some other extensions appeared recently in [94] where MISO IBC is considered with perfect CSIT and weighted R-ZF BF, with two optimized weight levels, for intracell or intercell interference. [16] considers the large system analysis of the MIMO IBC with optimized BF under partial CSIT. [95] studied the energy consumption dynamics in a MISO BC with users moving around according to a random walk model.

6.1.1 Summary of this Chapter

In this chapter:

- We introduce the concept of reduced-order ZF BF and propose a greedy approach to optimize the reduced ZF orders.
- We propose a large system analysis for optimal BF and DPC with omnidirectional but differently attenuated user channels.
- We consider a novel simple large system analysis for ZF BF or DPC transmitters with omnidirectional channel covariances.
- We illustrate with numerical evaluations the complexity-performance tradeoff that RO-ZF permits.

6.2 Multi-User MIMO System Model

Consider a transmitter (BS) equipped with M antennas communicating with K single antenna users (MISO BC). Furthermore, under narrowband transmission, the received signal at user k can be written as,

$$y_k = \mathbf{h}_k^H \mathbf{x} + n_k, \ k = 1, 2, \ ..., \ K, \tag{6.1}$$

where $\mathbf{h}_k \in \mathscr{C}^M$ is the downlink channel between user k and BS, $\mathbf{x} \in \mathscr{C}^M$ is the transmit vector and the noise terms $n_k \in \mathscr{CN}(0, \sigma^2)$ are independent. The channel covariance matrix is defined as $\mathbf{\Theta}_k$ and thus correlated channel model can be written as, $\mathbf{h}_k = \sqrt{M}\mathbf{\Theta}_k^{1/2}\mathbf{z}_k$, where $\mathbf{z}_k$ has i.i.d. complex entries of zero mean and variance $1/M$ and $\mathbf{\Theta}_k^{1/2}$ is any Hermitian square root of $\mathbf{\Theta}_k$. The correlation matrix $\mathbf{\Theta}_k$ is non-negative Hermitian and of uniformly bounded spectral norm w.r.t. M. The transmit signal $\mathbf{x}$ can be written as, $\mathbf{x} = \sum_{i=1}^{K} \mathbf{g}_i s_i$, where $\mathbf{g}_k \in \mathscr{C}^M$ represents the transmit precoder matrix for user k and s_i is the i^{th} user symbol, with $s_i \sim \mathscr{CN}(0,1)$. The transmit power constraint can be written as, $\mathrm{E}(\mathbf{x}^H\mathbf{x}) = \mathrm{tr}(\sum_{i=1}^{K} \mathbf{g}_i \mathbf{g}_i^H) \leq P$. Under optimal single user decoding, the user rate can be defined as, $R_k = \log(1+\gamma_k)$, where the signal to interference plus noise ratio (SINR), γ_k is defined as,

$$\gamma_k = \frac{|\mathbf{h}_k^H \mathbf{g}_k|^2}{\sum_{i=1, i\neq k}^{K} |\mathbf{h}_k^H \mathbf{g}_i|^2 + \sigma^2}. \tag{6.2}$$

The transmit SNR is defined as $\rho = \frac{P}{\sigma^2}$ and $\beta = \frac{K}{M}$. In the large system limit, we assume that $M, K \to \infty$ at a fixed ratio $\beta < 1$. Further we assume that the channel covariance matrices are

represented by multiple of identity, $\mathbf{\Theta}_k = \frac{\theta_k}{M}\mathbf{I}$, with different user channel covariance matrices differentiated by the varying attenuation factor θ_k. Multiple of identity covariance structure reflects the fact that the user subspaces are randomly oriented even though we do not assume the knowledge of subspaces. Further it helps to analytically evaluate the RO-ZF BF and compare it to optimal BF. Moreover, we define the ordering of the multiple of identity for the covariance matrices as, $\theta_1 \geq \theta_2 \geq, ..., \geq \theta_K$, which means user 1 represents the strongest user and K is the weakest user.

6.3 Large System Analysis of Optimal BF-WSMSE

In this section, we refer to the iterative algorithm in [9] for the optimal linear transmit BF and superscript (j) refers to the iteration stage j. We simplify the large system analysis results of the optimal BF in [15] for the case of multiple of identity covariance matrices for the user channels and the result is stated below. Since a detailed derivation (for general covariance matrices) already appears in [15], we skip the details and simplify the results therein for our simplified channel model. In the following sections, we denote $(x^2)^{(j)} = (x^{(j)})^2$.

Theorem 4. *Let $\gamma_k^{(j)\,opt-WSMSE}$ be the SINR of user k (6.2) under optimal linear precoding, i.e., at the end of iteration j, $\mathbf{g}_k^{(j)} = \sqrt{\frac{P}{\psi^{(j)}}}\left(\mathbf{H}\mathbf{D}^{(j)}\mathbf{H}^H + \alpha^{(j)}\mathbf{I}\right)^{-1}\mathbf{h}_k a_k^{(j)} w_k^{(j)}$, a_k is the MMSE Rx filter, w_k is the MSE weight for user k, $\psi^{(j)}$ being the normalization constant and $\alpha^{(j)} = \frac{tr(\mathbf{D}^{(j)})}{\rho}$ with the $(k,k)^{th}$ element of the diagonal matrix $\mathbf{D}^{(j)}, d_k^{(j)} = (a_k^2)^{(j)} w_k^{(j)}$. $\mathbf{H}$ represents the channel matrix of all users, $\mathbf{H} = [\mathbf{h}_1, ..., \mathbf{h}_K]$. Then $\gamma_k^{(j)\,opt-WSMSE} - \overline{\gamma}_k^{(j)\,opt-WSMSE} \xrightarrow{M\to\infty} 0$, almost surely, where,*

$$\overline{\gamma}_k^{(j)\,opt-WSMSE} = \frac{\theta_k^2 \overline{w}_k^{(j)} (e^2)^{(j)}}{\Upsilon_k^{(j)} + \frac{\sigma^2 \overline{d}_k^{(j)} \overline{\psi}^{(j)}}{P}(1+\overline{d}_k^{(j)}\theta_k e^{(j)})^2}, \tag{6.3}$$

where $\overline{w}_k^{(j)}, \overline{d}_k^{(j)}, \overline{\psi}^{(j)}$ represent the deterministic equivalents for $w_k^{(j)}, d_k^{(j)}, \psi^{(j)}$ respectively, the expressions of which are given below. Further we can show that, since the logarithm is a continuous function, by applying the continuous mapping theorem [96], it follows from the almost sure convergence of $\gamma_k^{(j)\,opt-WSMSE}$ that, $R_k^{(j)} - \overline{R}_k^{(j)} \xrightarrow[a.s]{M\to\infty} 0$, where $R_k^{(j)}$ is the rate of user k, with $\overline{R}_k^{(j)} = \ln(1+\overline{\gamma}_k^{(j)\,opt-WSMSE})$.

Normalization term: A deterministic equivalent $\overline{\psi}^{(j)}$ such that $\psi^{(j)} - \overline{\psi}^{(j)} \xrightarrow{M\to\infty} 0$, almost surely, is given by

$$\overline{\psi}^{(j)} = \frac{1}{M}\sum_{k=1}^{K} \overline{w}_k^{(j)} \frac{\overline{d}_k^{(j)}\theta_k e^{(j)\prime}}{(1+\overline{d}_k^{(j)}\theta_k e^{(j)})^2}, \tag{6.4}$$

Using theorem 1 [14], $e^{(j)}$ is given as the unique positive solution of the following equation,

$$e^{(j)} = \left(\sum_{i=1}^{K} \frac{\overline{d}_i^{(j)}\theta_i}{1+\overline{d}_i^{(j)}\theta_i e^{(j)}} + \alpha^{(j)}\right)^{-1}. \tag{6.5}$$

$e^{(j)\prime}$, the derivative of $e^{(j)}$ w.r.t $-\alpha^{(j)}$, is obtained as,

$$e^{(j)\prime} = \frac{(e^2)^{(j)}}{1-(e^2)^{(j)}\sum_{i=1}^{K} \frac{(\overline{d}_i^2)^{(j)}\theta_i^2}{(1+\overline{d}_i^{(j)}\theta_i e^{(j)})^2}}. \tag{6.6}$$

Signal Power: A deterministic equivalent for the square root of the signal power, $\sqrt{\overline{P}_{S,k}^{(j)}}$ gets simplified as,

$$\sqrt{\overline{P}_{S,k}^{(j)}} = \sqrt{\frac{P}{\overline{\psi}^{(j)}}} \frac{\overline{d}_k^{(j)} \theta_k e^{(j)}}{\overline{a}_k^{(j)} (1+\overline{d}_k^{(j)} \theta_k e^{(j)})}. \tag{6.7}$$

Interference Power: Following [14, 15], the deterministic equivalent for the interference power can be obtained as,

$$\begin{aligned} \sum_{i=1,\neq k}^{K} \mathbf{h}_k^H \mathbf{g}_i^{(j)} \mathbf{g}_i^{(j)\,H} \mathbf{h}_k &= \frac{P}{\overline{d}_k^{(j)} \overline{\psi}^{(j)}} \frac{\Upsilon_k^{(j)}}{(1+\overline{d}_k^{(j)} \theta_k e^{(j)})^2}, \\ \text{where,} \quad \Upsilon_k^{(j)} &= \frac{1}{M} \sum_{i=1, i\neq k}^{K} \overline{w}_i^{(j)} \frac{\overline{d}_i^{(j)} \theta_i e^{(j)\prime}}{(1+\overline{d}_i^{(j)} \theta_i e^{(j)})^2}. \end{aligned} \tag{6.8}$$

Substituting the signal and interference powers, the deterministic equivalent of the SINR leads to (6.3).The deterministic equivalents for the $a_k^{(j)}, w_k^{(j)}, d_k^{(j)}$ are given by [15], $\overline{a}_k^{(j)} = \frac{\sigma}{\sqrt{\overline{P}_{S,k}^{(j-1)}}} \frac{\overline{\gamma}_k^{(j-1)}}{1+\overline{\gamma}_k^{(j-1)}}$, $\overline{w}_k^{(j)} = u_k(1+\overline{\gamma}_k^{(j-1)})$, and $\overline{d}_k^{(j)} = (\overline{a}_k^2)^{(j)} \overline{w}_k^{(j)}$.

6.4 Large System Analysis of Optimal DPC

The received signal at user k with DPC [6] (which achieves the capacity region of MIMO BC) at the BS is

$$y_k = \underbrace{\mathbf{h}_k^H \mathbf{g}_k s_k}_{\text{signal}} + \underbrace{\sum_{i=k+1}^{K} \mathbf{h}_k^H \mathbf{g}_i s_i}_{\text{interf. from weaker users}} + \mathbf{n}_k. \tag{6.9}$$

In optimal DPC, users are ordered in decreasing strength, as in RO-ZF. The interference that a user will cause to weaker users gets canceled non-linearly at the Tx (in other words, in the Rx SINR it does not need to be considered), and the BF handles only interference to stronger users. As usual, optimal BF does something in between ZF and matched filter (MF). So there will be residual interference at the stronger users.

Let γ_k^{DPC} be the SINR of user k under optimal DPC, i.e., at the end of iteration j, $\gamma_k^{(j)\,DPC} - \overline{\gamma}_k^{(j)\,DPC} \xrightarrow{M\to\infty} 0$, almost surely, where, the expression for $\overline{\gamma}_k^{(j)\,DPC}$ is same as (6.3). However, the expressions for each of the scalars got modified as,

$$\begin{aligned} e^{(j)} &= (\sum_{i=k}^{K} \frac{\overline{d}_i^{(j)} \theta_i}{1+\overline{d}_i^{(j)} \theta_i e^{(j)}} + \alpha^{(j)})^{-1}, \\ \Upsilon_k^{(j)} &= \frac{1}{M} \sum_{i=k+1}^{K} \overline{w}_i^{(j)} \frac{\overline{d}_i^{(j)} \theta_i e^{(j)\prime}}{(1+\overline{d}_i^{(j)} \theta_i e^{(j)})^2}. \end{aligned} \tag{6.10}$$

Note that the only change compared to the optimal WSMSE BF is that each summation term get replaced from k to K or $k+1$ to K.

6.5 Reduced Order ZF

In this section, we consider the BF to be a reduced order ZF (RO-ZF). This can be interpreted as the number of interfering channels to be zero-forced for a user k is much less than K. The RO-ZF BF $\mathbf{g}_k$ can be written as,

$$\mathbf{g}_k = \frac{\mathbf{P}^{\perp}_{\mathbf{H}_{I_k}} \mathbf{h}_k}{||\mathbf{P}^{\perp}_{\mathbf{H}_{I_k}} \mathbf{h}_k||}. \tag{6.11}$$

Here, $\mathbf{P}_H = \mathbf{H}(\mathbf{H}^H\mathbf{H})^{\#}\mathbf{H}^H$ represent the projection onto the column space of $\mathbf{H}$, $\mathbf{P}^{\perp}_{\mathbf{H}} = \mathbf{I} - \mathbf{P}_H$ is the projection onto its orthogonal complement (# represents the Moore-Penrose pseudo-inverse). For the convenience of analysis, we define the following: K_k represents the strongest interfering channel zero-forced by the BF of user k and I_k denotes the set of user indices for which the ZF is done. $\mathbf{H}_{I_k}$ represents the matrix of all the user channels in I_k. Complexity in the RO-ZF case will be about half of that of full ZF (multiplying the $M \times K$ $\mathbf{H}$ by a triangular $K \times K$ instead of a full $K \times K$, computation of the $K \times K$ inverse or triangular factor takes $O(K^3)$ operations, with a smaller factor if only a triangular factor is needed and not a full inverse).

6.6 Large System Analysis for RO-ZF, Full Order ZF and ZF-DPC

In this section we consider the large system analysis for the RO-ZF scheme proposed in this chapter and also the full order ZF (full order means $|I_k| = K-1, \forall k$). In this section, we split $\mathbf{g}_k = \sqrt{p_k}\mathbf{g}'_k$, where p_k is the power allocated to user k.

$$\begin{aligned}
\gamma_k^{RO-ZF} &= \frac{P_{S,k}}{P_{I,k} + \sigma_k^2} \\
&= \frac{p_k |\mathbf{h}_k^H \mathbf{g}'_k|^2}{\sum_{i=1, i\neq k}^{K} p_i |\mathbf{h}_k^H \mathbf{g}'_i|^2 + \sigma_k^2}, \\
\mathbf{g}'_k &= \frac{\mathbf{P}^{\perp}_{\mathbf{H}_{I_k}} \mathbf{h}_k}{||\mathbf{P}^{\perp}_{\mathbf{H}_{I_k}} \mathbf{h}_k||} \implies \\
\mathbf{h}_k^H \mathbf{g}'_k &= \left\| \mathbf{P}^{\perp}_{\mathbf{H}_{I_k}} \mathbf{h}_k \right\|.
\end{aligned} \tag{6.12}$$

Further, by the law of large numbers, $P_{S,k} - \overline{P}_{S,k} \xrightarrow[a.s]{M\to\infty} 0$, where,

$$\begin{aligned}
\overline{P}_{S,k} &= \mathrm{E}(|\mathbf{h}_k^H \mathbf{g}'_k|^2) \\
&= \mathrm{E}_{\mathbf{H}_{I_k}} \mathrm{E}_{\mathbf{h}_k} \mathrm{tr}(\mathbf{P}^{\perp}_{\mathbf{H}_{I_k}} \mathbf{h}_k \mathbf{h}_k^H) \\
&= \frac{\theta_k}{M} \mathrm{tr}(\mathbf{I}_M - \mathbf{H}_{I_k}(\mathbf{H}_{I_k}^H \mathbf{H}_{I_k})^{\#} \mathbf{H}_{I_k}^H) \\
&= \theta_k (1 - \frac{|I_k|}{M}),
\end{aligned} \tag{6.13}$$

where we use the property of the projection matrices that $\mathbf{P}^{\perp}_{\mathbf{H}_{I_k}} \mathbf{P}^{\perp}_{\mathbf{H}_{I_k}} = \mathbf{P}^{\perp}_{\mathbf{H}_{I_k}}$. Next, we consider the terms in $P_{I,k}$,

$$|\mathbf{h}_k^H \mathbf{g}'_i|^2 = \frac{|\mathbf{h}_k^H \mathbf{P}^{\perp}_{\mathbf{H}_{I_i}} \mathbf{h}_i|^2}{\left\| \mathbf{P}^{\perp}_{\mathbf{H}_{I_i}} \mathbf{h}_i \right\|^2}. \tag{6.14}$$

$$
\begin{aligned}
&\text{If } k \in I_i, \text{ then } |\mathbf{h}_k^H \mathbf{g}_i'|^2 = 0, \text{ else,} \\
&\mathrm{E}(|\mathbf{h}_k^H \mathbf{P}^{\perp}_{\mathbf{H}_{I_i}} \mathbf{h}_i|^2) = \mathrm{E}(\mathrm{tr}(\mathbf{P}^{\perp}_{\mathbf{H}_{I_i}} \mathbf{h}_i \mathbf{h}_i^H \mathbf{P}^{\perp}_{\mathbf{H}_{I_i}} \mathbf{h}_k \mathbf{h}_k^H)) \\
&\qquad = \frac{\theta_k \theta_i}{M^2} \mathrm{tr}(\mathbf{P}^{\perp}_{\mathbf{H}_{I_i}}) \\
&\qquad = \frac{\theta_k \theta_i}{M^2} \mathrm{tr}(\mathbf{I}_M - \mathbf{H}_{I_i} (\mathbf{H}_{I_i}^H \mathbf{H}_{I_i})^{\#} \mathbf{H}_{I_i}^H) \\
&\qquad = \frac{\theta_k \theta_i}{M} (1 - \frac{|I_i|}{M}).
\end{aligned} \tag{6.15}
$$

Finally we obtain $\mathrm{E}(|\mathbf{h}_k^H \mathbf{g}_i'|^2) = \frac{\frac{\theta_k \theta_i}{M}(1-\frac{|I_i|}{M})}{\theta_i(1-\frac{|I_i|}{M})} = \frac{\theta_k}{M}$. Further, we get the deterministic equivalent of the SINR in the large system limit as,

$$
\overline{\gamma}_k^{RO-ZF} = \frac{p_k \theta_k}{\frac{1}{M}\theta_k \sum\limits_{i=1, k \notin I_i}^{K} p_i + \sigma^2} \left(1 - \frac{|I_k|}{M}\right). \tag{6.16}
$$

For the full order ZF, the interference power vanishes from the SINR terms,

$$
\overline{\gamma}_k^{ZF} = \frac{p_k \theta_k}{\sigma^2} \left(1 - \frac{K-1}{M}\right). \tag{6.17}
$$

ZF-DPC combines zero-forcing and DPC technique. While DPC cancels the interference for users $i < k$, the interference of users $i > k$ are eliminated by designing the BF $\mathbf{g}_i$ such that $\mathbf{h}_k^H \mathbf{g}_i = 0$.The large system analysis for the ZF-DPC ($|I_k| = k-1$) is as follows : We define $J_k = \{1, 2, ..., k-1\}$.

$$
\gamma_k^{ZF-DPC} = \frac{P_{S,k}}{P_{I,k} + \sigma_k^2} = \frac{p_k |\mathbf{h}_k^H \mathbf{g}_k'|^2}{\sigma_k^2}, \text{ since, } P_{I,k} = 0,
$$

$$
\begin{aligned}
\mathbf{g}_k' &= \frac{\mathbf{P}^{\perp}_{\mathbf{H}_{J_k}} \mathbf{h}_k}{||\mathbf{P}^{\perp}_{\mathbf{H}_{J_k}} \mathbf{h}_k||}, \\
\implies \mathbf{h}_k^H \mathbf{g}_k' &= \left\| \mathbf{P}^{\perp}_{\mathbf{H}_{J_k}} \mathbf{h}_k \right\|, \\
\overline{P}_{S,k} &= \mathrm{E}(|\mathbf{h}_k^H \mathbf{g}_k'|^2) \\
&= \mathrm{E}_{\mathbf{H}_{J_k}} \mathrm{E}_{\mathbf{h}_k} \mathrm{tr}(\mathbf{P}^{\perp}_{\mathbf{H}_{J_k}} \mathbf{h}_k \mathbf{h}_k^H) \\
&= \frac{\theta_k}{M} \mathrm{tr}(\mathbf{I}_M - \mathbf{H}_{J_k} (\mathbf{H}_{J_k} \mathbf{H}_{J_k}^H)^{\#} \mathbf{H}_{J_k}) \\
&= \theta_k (1 - \frac{k-1}{M}).
\end{aligned} \tag{6.18}
$$

Therefore, the deterministic equivalent of the SINR becomes,

$$
\overline{\gamma}_k^{ZF-DPC} = \frac{p_k \theta_k}{\sigma^2} \left(1 - \frac{k-1}{M}\right). \tag{6.19}
$$

6.6.1 Optimization of user powers p_k

We consider here the approximation of the WSR according to the difference of convex (DC) functions approach as in [30]. Solving DC, we get the Lagrangian for the WSR,

$$
\begin{aligned}
&WSR(\mathbf{g},\lambda) = \lambda P + \sum_{k=1}^{K} u_k \ln\det(1+\mathbf{g}_k^H \mathbf{B}_k \mathbf{g}_k) - \mathbf{g}_k^H(\mathbf{A}_k+\lambda\mathbf{I})\mathbf{g}_k,\\
&\text{where, } \mathbf{B}_k = \mathbf{h}_k r_{\bar{k}}^{-1} \mathbf{h}_k^H,\\
&\mathbf{A}_k = \sum_{i\neq k, i\notin I_k}^{K} u_i \mathbf{h}_i (r_{\bar{i}}^{-1} - r_i^{-1})\mathbf{h}_i^H,\\
&r_{\bar{k}} = \sum_{i=1, i\neq k, k\notin I_i}^{K} |\mathbf{h}_k^H \mathbf{g}_i|^2 + \sigma^2,\\
&r_k = r_{\bar{k}} + |\mathbf{h}_k^H \mathbf{g}_k|^2.
\end{aligned}
\tag{6.20}
$$

Here $\mathbf{g}$ represents the set of BFs $\mathbf{g}_k$. Let $\sigma_k^{(1)} = \mathbf{g}_k'^H \mathbf{B}_k \mathbf{g}_k'$ and $\sigma_k^{(2)} = \mathbf{g}_k'^H \mathbf{A}_k \mathbf{g}_k'$. For full order ZF, $\mathbf{A}_k = 0$, thus $\sigma_k^{(2)} = 0$ and (6.20) reduces to standard waterfilling. The advantage of formulation (6.20) is that it allows straightforward power adaptation: introducing stream powers $p_k \geq 0$ and substituting $\mathbf{g}_k = \mathbf{g}_k' \sqrt{p_k}$ in (6.20) yields

$$
WSR(\mathbf{P},\lambda) = \lambda P + \sum_{k=1}^{K} [u_k \ln(1+p_k\sigma_k^{(1)}) - \mathrm{tr}(p_k(\sigma_k^{(2)}+\lambda))],
\tag{6.21}
$$

where $\mathbf{P}$ represents the set of powers p_k. Since this is a concave function w.r.t p_k, taking the derivative leads to the following interference leakage aware water filling (WF) (jointly for the p_k and λ)

$$
p_k = \left(u_k(\sigma_k^{(2)}+\lambda)^{-1} - \sigma_k^{-(1)}\right)^+, \sum_k p_k = P,
\tag{6.22}
$$

where the Lagrange multiplier is adjusted to satisfy the power constraints. This can be done by bisection.

6.7 Optimization of the ZF Order

In this section, we consider an alternating optimization algorithm (Algorithm 10) which computes the reduced ZF order for each user (I_k, K_k). In Algorithm 10, the text "if $u_k R_k + u_{K_k'} R_{K_k'}$ is increased" is meant to be understood "by adding the ZF from k to K_k'". In Algorithm 1, we consider the ordering for the case of, $I_k = \{K_k, K_k+1, ..., K_k+|I_k|-1\}$ if $k < K_k$, else $I_k = \{K_k, K_k+1, ..., K_k+|I_k|\}$. $|I_k|$ represents the cardinality of the set I_k. Also, $\mathbf{H}_{I_k} = [\mathbf{h}_{K_k}, ..., \mathbf{h}_{K_k+|I_k|-1}]$ or $\mathbf{H}_{I_k} = [\mathbf{h}_{K_k}, ..., \mathbf{h}_{K_k+|I_k|}]$. Note that at finite dimension MIMO, not only the channel strengths but also the relative orientation of the channel vectors count. However, in MaMIMO with multiple of identity covariances, there is no orientation issue, only the channel strengths count. So the user ordering is simple.

6.8 Simulation Results

In this section we illustrate the simulation results to validate our theoretical results. We compare the sum rate performance of RO-ZF BF scheme (which has the least complexity) to the optimal

Algorithm 10: Reduced Zero-Forcing Order Determination

Given: $K, M, \sigma^2, \theta_i, \forall i$, with ordering $\theta_1 \geq \theta_2 \geq ... \geq \theta_K$.
Initialization: Start with $K_k = k+1, \forall k = 1, ..., K-1$ and for user K, $|I_K| = 0$.

for $k = 1 : K$
  $K'_k = K_k$.
  while $(K'_k > 1)$
    $K'_k = K'_k - 1$.
    if$(K'_k \neq k)$
      if $u_k R_k + u_{K'_k} R_{K'_k}$ is increased
        $K_k = K'_k$, else exit while loop
    else end if
  end while
  $K'_k = K_k$.
  while $(K'_k < K)$
    $K'_k = K'_k + 1$.
    if$(K'_k \neq k)$
      if $u_k R_k + u_{K'_k} R_{K'_k}$ is increased
        $K_k = K'_k$, else exit while loop
    else end if
  end while
end for
Continue until convergence of $I_k, \forall k$.

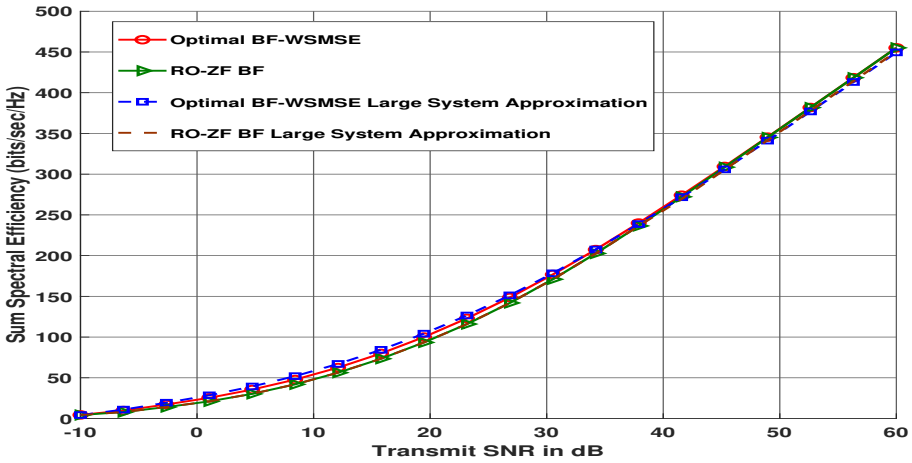


Figure 6.1: Accuracy of large system approximation $M = 64, K = 30$.

BF-WSMSE [9], optimal DPC and to the large system approximations of optimal BF-WSMSE, full order ZF and the ZF DPC. For the SNR ranges of interest, it can be seen that RO-ZF performs close to the optimal schemes with much lower complexity, see Figure 6.2. Figure 6.3 illustrates the sum rate difference of the various BF designs from the optimal DPC. In Figure 6.1, We first validate our large system approximation. It clearly shows that the rate expression resulting from large system approximation for RO-ZF BF matches exactly with that of the Monte-Carlo simulations.

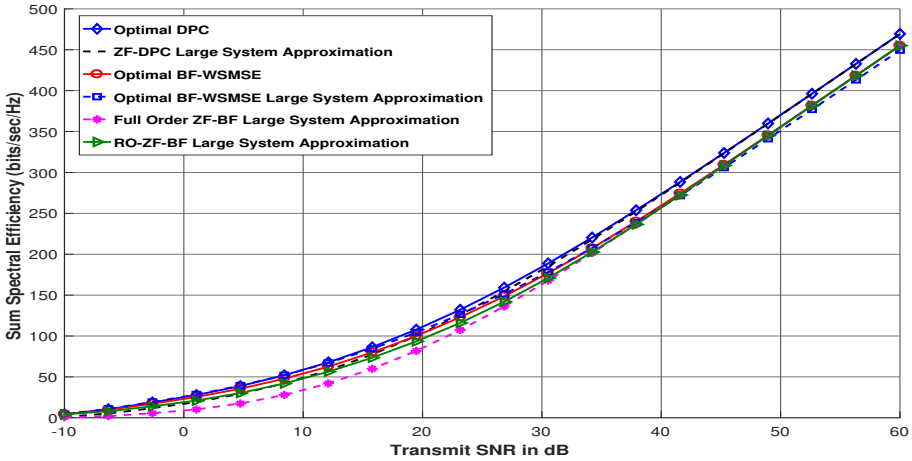


Figure 6.2: Sum rate comparison for $M = 64, K = 30$.

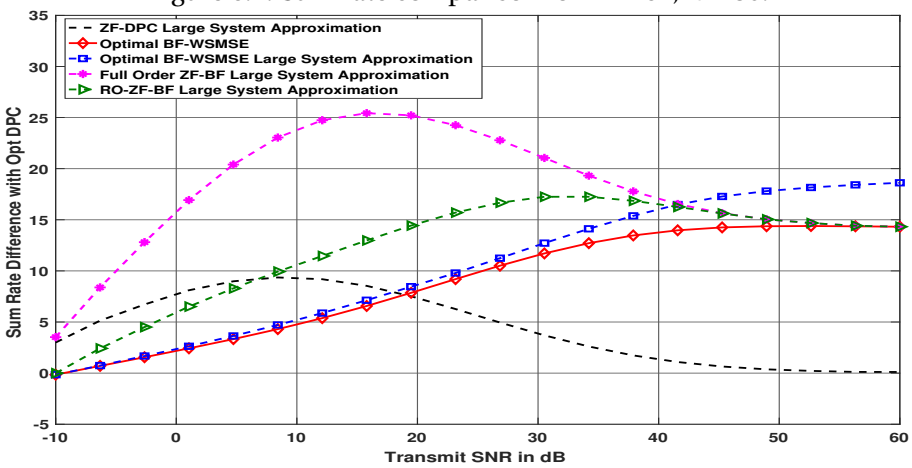


Figure 6.3: Sub-optimality compared to Optimal DPC for $M = 64, K = 30$.

Concluding Remarks 4

- In this chapter, we investigate the performance-complexity tradeoffs for the reduced order ZF BF. We propose the large system analysis for the RO-ZF BF, optimal BF, optimal DPC, ZF-DPC and full order ZF for the case of omnidirectional but differently attenuated user channels.
- Simulation results indicate that our RO-ZF BF scheme has a performance very close to the optimal BFs such as WSMSE and DPC, but with much lesser complexity compared to the full order ZF. We also propose an alternating optimization algorithm which computes the optimal ZF order for each user.
- We conjecture here one special scenario where the RO-ZF indeed can bring lower computational complexity. Consider a MISO single cell where channel $\mathbf{H}$ is $K \times M$ and $\mathbf{G}$ is $M \times K$. We actually obtain MMSE-ZF by $\arg\min_{\mathbf{G}} \|G\|^2$ s.t. $\mathbf{HG} = \mathbf{I}_K$ which leads to $\mathbf{G} = \mathbf{H}^H(\mathbf{H}\mathbf{H}^H)^{-1}$. Since we can still adjust the stream powers, the correct ZF requirement would be only $\text{offdiag}(\mathbf{HG}) = \mathbf{0}$ where $\text{offdiag}(\cdot)$ is the off-diagonal part of the matrix argument. This leads to $\mathbf{G} = \mathbf{H}^H(\mathbf{H}\mathbf{H}^H)^{-1}\mathbf{P}$ where $\mathbf{P}$ is a diagonal matrix of

Concluding Remarks 4 (cont.)

stream powers. Then $\arg\min_G \|\mathbf{G}\|^2$ criterion gets replaced by $\arg\max_P WSR$ to optimize the power. For RO-ZF in which a user gets ZF'd to or not, but when he does, he gets ZF'd to by all (other) BFs. Let after permutation $\mathbf{H}^H = \left[\mathbf{H}_z^H \, \mathbf{H}_n^H\right]$, where $\mathbf{H}_z$ = user channels to be ZF'd to, $\mathbf{H}_n$ are the others. Let $\mathbf{G} = \left[\mathbf{G}_z \, \mathbf{G}_n\right]$. Then $\mathbf{G} = \mathbf{H}^H\mathbf{A}$ where $\mathbf{A}$ is $K \times K$. As in traditional ZF, $\mathbf{H}_z\mathbf{G} = \left[\mathbf{I}_{K_z} \, \mathbf{0}\right]$, where $K = K_z + K_n$. In principle, have to do $\arg\min_G ||G||^2$ under the ZF constraint in the line above. The solution of the restructured problem above will be: $\mathbf{G}_z = \mathbf{H}_z^H(\mathbf{H}_z\mathbf{H}_z^H)^{-1}$ and $\mathbf{G}_n = \mathbf{P}^{\perp}_{\mathbf{H}_z^H}\mathbf{H}_n^H$ This leads to $\mathbf{A} = \begin{bmatrix} (\mathbf{H}_z\mathbf{H}_z^H)^{-1} & -(\mathbf{H}_z\mathbf{H}_z^H)^{-1}\mathbf{H}_z\mathbf{H}_n^H \\ \mathbf{0} & \mathbf{I}_{K_n} \end{bmatrix}$. The key point is that there is only one matrix inverse to be computed: $(\mathbf{H}_z\mathbf{H}_z^H)^{-1}$.

- RO-ZF is motivated by being simpler than MMSE BF and having performance close to it. For that to be true, the computation (including the optimization of the RO in fact) of RO-ZF should be simpler than the computation of the full-order ZF, ideally. Hence, a simplified algorithm for the RO-ZF order optimization still remains an open problem. That is not simple to achieve, especially the RO optimization, but that gets simpler if we require any user either to be ZF'd to by all or none.

6.9 Extension of RO-ZF BF to IBC under Partial CSIT

We consider an IBC with C cells with a total of K single antenna users. We shall consider a system-wide numbering of the users. User k is served by BS b_k. The received signal at user k in cell b_k is

$$\mathbf{y}_k = \underbrace{\mathbf{h}_{k,b_k}^H \mathbf{g}_k \, x_k}_{\text{signal}} + \underbrace{\sum_{\substack{i \neq k \\ b_i = b_k}} \mathbf{h}_{k,b_k}^H \mathbf{g}_i \, x_i}_{\text{intracell interf.}} + \underbrace{\sum_{j \neq b_k} \sum_{i:b_i=j} \mathbf{h}_{k,j}^H \mathbf{g}_i \, x_i}_{\text{intercell interf.}} + \mathbf{v}_k \tag{6.23}$$

where x_k is the intended (white, unit variance) scalar signal stream, $\mathbf{h}_{k,b_i}$ is the $M_{b_k} \times 1$ channel from BS b_i to user k. The Rx signal (and hence the channel) is assumed to be scaled so that we get for the noise $v_k \sim \mathcal{CN}(0,1)$. BS b_k serves $K_{b_k} = \sum_{i:b_i=b_k} 1$ users. The $M_{b_k} \times 1$ spatial Tx filter or beamformer (BF) is $\mathbf{g}_k$.

6.9.1 Channel and CSIT Model

For simplicity, we omit all the user indices k. We start from a deterministic Least-Squares (LS) channel estimate

$$\widehat{\mathbf{h}}_{LS} = \mathbf{h} + \widetilde{\mathbf{h}}, \tag{6.24}$$

where $\mathbf{h}$ is the true MISO channel, and the error is modeled as circularly symmetric white Gaussian noise $\widetilde{\mathbf{h}} \sim \mathcal{CN}(0, \widetilde{\sigma}^2\mathbf{I})$. Now each MISO channel is modeled according to a correlation struc-

ture (Karhunen-Loeve representation [21]) as follows,

$$\begin{aligned} \mathbf{h} &= \mathbf{C}\mathbf{c}, \\ \mathbf{c} &= \mathbf{D}^{1/2}\mathbf{c}', \end{aligned} \tag{6.25}$$

where $\mathbf{c}' \sim \mathscr{CN}(0, \mathbf{I}_L)$ and $\mathbf{D}$ is diagonal. Here $\mathbf{C}$ is the $M \times L$ eigenvector matrix of the reduced rank channel covariance $\mathbf{R_{hh}} = \mathbf{CDC}^H$. The total sum rank across all users $N_p = \sum_{k=1}^{K} L_{k,c}$ is assumed to be less than M_c, where $L_{k,c}$ is the channel rank between user k and BS c. Assuming the channel covariance subspace is known, the LMMSE channel estimate can be written as $\widehat{\mathbf{h}} = \mathbf{CDC}^H \left(\mathbf{CDC}^H + \widetilde{\sigma}^2\mathbf{I}\right)^{-1} \widehat{\mathbf{h}}_{LS}$. Applying the matrix inversion lemma and exploiting $\mathbf{C}^H\mathbf{C} = \mathbf{I}_L$, this simplifies to

$$\begin{aligned} \widehat{\mathbf{h}} &= \mathbf{C}\left(\widetilde{\sigma}^2\mathbf{D}^{-1} + \mathbf{I}\right)^{-1} \mathbf{C}^H \widehat{\mathbf{h}}_{LS} \\ &= \mathbf{C}\widehat{\mathbf{D}}^{1/2}\widehat{\mathbf{c}}, \end{aligned} \tag{6.26}$$

where

$$\begin{aligned} \widehat{\mathbf{D}} &= \left(\widetilde{\sigma}^2\mathbf{D}^{-1} + \mathbf{I}\right)^{-1} \mathbf{D}, \\ \text{and } \widehat{\mathbf{c}} &= \mathbf{D}^{-1/2}(\sigma^2\mathbf{D}^{-1} + \mathbf{I})^{-1/2}\mathbf{C}^H\widehat{\mathbf{h}}_{LS}. \end{aligned} \tag{6.27}$$

Note that a detailed derivation of the LMMSE estimate here appears in Appendix A (where (6.26) follows immediately by substituting for $\mathbf{C}_r = 1$).The posterior error covariance becomes

$$\mathbf{R}_{\widetilde{\mathbf{h}}\widetilde{\mathbf{h}}} = \mathbf{CDC}^H - \mathbf{CDC}^H \left(\mathbf{CDC}^H + \widetilde{\sigma}^2 I\right)^{-1} \mathbf{CDC}^H, \tag{6.28}$$

which the matrix inversion lemma allows to simplify to,

$$\begin{aligned} \mathbf{R}_{\widetilde{\mathbf{h}}\widetilde{\mathbf{h}}} &= \mathbf{C}\left[\mathbf{D} - \left(\widetilde{\sigma}^2\mathbf{D}^{-1} + \mathbf{I}\right)^{-1}\mathbf{D}\right]\mathbf{C}^H \\ &= \mathbf{C}\widetilde{\mathbf{D}}\mathbf{C}^H. \end{aligned} \tag{6.29}$$

So we can write for $\mathbf{S} = \mathrm{E}_{\mathbf{h}|\widehat{\mathbf{h}}}\left(\mathbf{hh}^H\right) = \widehat{\mathbf{h}}\widehat{\mathbf{h}}^H + \mathbf{R}_{\widetilde{\mathbf{h}}\widetilde{\mathbf{h}}} = \mathbf{CWC}^H$, where $\mathbf{W} = \widehat{\mathbf{D}}^{1/2}\widehat{\mathbf{c}}\widehat{\mathbf{c}}^H\widehat{\mathbf{D}}^{1/2} + \widetilde{\mathbf{D}}$.

6.9.2 Partial CSIT BF based on Different Channel Estimates

In the MaMIMO limit, BF design with partial CSIT will depend on the quantities $\mathbf{S} = \mathrm{E}_{\mathbf{h}|\widehat{\mathbf{h}}}\left(\mathbf{hh}^H\right) = \widehat{\mathbf{h}}\widehat{\mathbf{h}}^H + \mathbf{R}_{\widetilde{\mathbf{h}}\widetilde{\mathbf{h}}}$. We shall consider three possible channel estimates.

(i) *LS Channel Estimate*

We have $\widehat{\mathbf{h}}_{LS} = \mathbf{h} + \widetilde{\mathbf{h}}$, where $\mathbf{h}$ and $\widetilde{\mathbf{h}}$ are independent. In the LS case, $\mathbf{R}_{\widetilde{\mathbf{h}}\widetilde{\mathbf{h}}} = \widetilde{\sigma}^2\mathbf{I}$.

(ii) *LMMSE Channel Estimate*

We have $\mathbf{h} = \widehat{\mathbf{h}} + \widetilde{\mathbf{h}}$ in which $\widehat{\mathbf{h}}$ and $\widetilde{\mathbf{h}}$ are decorrelated and hence independent in the Gaussian case. In the LMMSE case, $\mathbf{R}_{\widetilde{\mathbf{h}}\widetilde{\mathbf{h}}}$ is the posterior covariance. The resulting $\mathbf{S} = \widehat{\mathbf{h}}\widehat{\mathbf{h}}^H + \mathbf{R}_{\widetilde{\mathbf{h}}\widetilde{\mathbf{h}}}$ now forms an unbiased estimate of $\mathbf{hh}^H$: $\mathrm{E}_{\widehat{\mathbf{h}}}\mathbf{S} = \mathbf{R_{hh}}$.

(iii) *Subspace Projection based Channel Estimate*

We also investigate the effect of limiting channel estimation error to the covariance subspace (without the LMMSE weighting, this is a simplification of the LMMSE estimate). The subspace channel estimate is given as,

$$\begin{aligned} \widehat{\mathbf{h}}_S &= \mathbf{P}_C\widehat{\mathbf{h}}_{LS} = \mathbf{h} + \mathbf{P}_C\widetilde{\mathbf{h}}_{LS}, \\ \mathbf{R}_{\widetilde{\mathbf{h}}_S\widetilde{\mathbf{h}}_S} &= \widetilde{\sigma}^2\mathbf{P}_C, \end{aligned} \tag{6.30}$$

where $\mathbf{P}_C = \mathbf{C}(\mathbf{C}^H\mathbf{C})^{-1}\mathbf{C}^H$ represents the projection onto the covariance subspace. Here, $\mathbf{S} = \widehat{\mathbf{h}}_S\widehat{\mathbf{h}}_S^H + \mathbf{R}_{\widetilde{h}_S\widetilde{h}_S} = \mathbf{C}(\widehat{\mathbf{c}}\widehat{\mathbf{c}}^H + \widetilde{\sigma}^2\mathbf{I})\mathbf{C}^H$.

6.9.3 BF with Partial CSIT

Three types of BF design with partial CSIT can be analyzed. In the case of partial CSIT we get for the Rx signal,

$$y_k = \widehat{\mathbf{h}}_{k,b_k}^H \mathbf{g}_k x_k + \underbrace{\widetilde{\mathbf{h}}_{k,b_k}^H \mathbf{g}_k x_k}_{\text{sig. ch. error}} + \sum_{i=1,\neq k}^{K} (\widehat{\mathbf{h}}_{k,b_i}^H \mathbf{g}_i x_i + \underbrace{\widetilde{\mathbf{h}}_{k,b_i}^H \mathbf{g}_i x_i}_{\text{interf. ch. error}}) + v_k . \tag{6.31}$$

1) Naive BF EWSR: just replace $\mathbf{h}$ by $\widehat{\mathbf{h}}$ in a perfect CSIT approach. Ignore $\widetilde{\mathbf{h}}$ everywhere. 2) Optimal BF EWSR: accounts for covariance CSIT in the signal and interference terms.

6.9.4 Max EWSR ZF BF in the MaMISO limit (ESEI-WSR)

The scenario of interest here is to design optimal beamformers when there is only partial CSIT. Once the CSIT is imperfect, various optimization criteria such as outage capacity can be considered. Here the design is based on expected weighted sum rate (EWSR) (and in a first instance with LMMSE channel estimates). The actual EWSR represents two rounds of averaging. In a first stage, the WSR is averaged over the channels given the channel estimates and covariance information (i.e. the partial CSIT), leading to a cost function that can be optimized by the Tx. The optimized result then needs to be averaged over the channel estimates to obtain the final ergodic WSR. In the MaMISO limit, due to the law of large numbers, a number of scalars converge to their expected value, facilitating averaging the WSR. From the law of total expectation and motivated from the ergodic capacity formulations [97] (point to point MIMO systems), [98] (multi user MISO systems),

$$\begin{aligned} EWSR &= \mathrm{E}_{\widehat{\mathbf{h}}} \max_{\mathbf{g}} EWSR(\mathbf{g}), \\ EWSR(\mathbf{g}) &= \mathrm{E}_{\mathbf{h}|\widehat{\mathbf{h}}} WSR(g) \\ &= \sum_{k=1}^{K} u_k \, \mathrm{E}_{\mathbf{h}|\widehat{\mathbf{h}}} \ln(s_k / s_{\overline{k}}) \\ &\overset{(a)}{=} \sum_{k=1}^{K} u_k \ln((\mathrm{E}_{\mathbf{h}|\widehat{\mathbf{h}}} s_k)/(\mathrm{E}_{\mathbf{h}|\widehat{\mathbf{h}}} s_{\overline{k}})) \\ &= \sum_{k=1}^{K} u_k \ln(r_{\overline{k}}^{-1} r_k), \end{aligned} \tag{6.32}$$

where transition (a) represents the MaMISO limit leading to ESEI-WSR (Expected Signal Expected Interference WSR), u_k are the rate weights, $\mathbf{g}$ represents the collection of BFs $\mathbf{g}_k$. $s_{\overline{k}}$ is the (channel dependent) interference plus noise power and s_k is the signal plus interference plus noise power. Their conditional expectations are

$$\begin{aligned} r_{\overline{k}} &= 1 + \sum_{i \neq k} \mathrm{E}_{\mathbf{h}|\widehat{\mathbf{h}}} |\mathbf{h}_{k,b_i}^H \mathbf{g}_i|^2 \\ &= 1 + \sum_{i \neq k} \mathbf{g}_i^H \mathbf{S}_{k,b_i} \mathbf{g}_i, \\ r_k &= r_{\overline{k}} + \mathbf{g}_k^H \mathbf{S}_{k,b_k} \mathbf{g}_k, \; \mathbf{S}_{k,b_k} \\ &= \mathbf{C}_{k,b_k} \mathbf{W}_{k,b_k} \mathbf{C}_{k,b_k}. \end{aligned} \tag{6.33}$$

For optimal ZF BF, all the interfering powers $\mathbf{g}_i^H \mathbf{S}_{k,b_i} \mathbf{g}_i = 0$ and thus $\mathbf{g}_k$ should belong to the orthogonal complement of the eigenvector subspace of all the interfering users. For this purpose, we define $\mathbf{C}_{\overline{k}}$ as the eigenvector space of all the users (except k) channel from b_k, $\mathbf{C}_{\overline{k}} =$

$[\mathbf{C}_{1,b_k},...,\mathbf{C}_{k-1,b_k},\mathbf{C}_{k+1,b_k},..,\mathbf{C}_{K,b_k}]$. Further we split $\mathbf{g}_k = \mathbf{g}'_k p_k^{1/2}$, where p_k is the power allocated to user k, and $||\mathbf{g}'_k|| = 1$. By adding the Lagrange terms for the BS power constraints, $\sum_{c=1}^{C} \mu_c (P_c - \sum_{k:b_k=c} ||\mathbf{g}_k||^2)$, to the EWSR in (6.32), we get the gradient (with $\alpha_k = \frac{u_k}{r_k}$),

$$\frac{\partial EWSR}{\partial \mathbf{g}_k^*} = \alpha_k \mathbf{S}_{k,b_k} \mathbf{g}_k - \mu_{b_k} \mathbf{g}_k = 0, \tag{6.34}$$

leading to $\mathbf{g}_k \propto \mathbf{V}_{max}(\mathbf{S}_{k,b_k})$. Finally we obtain the ZF BF as, $\mathbf{g}'_k = \mathbf{P}^{\perp}_{\mathbf{C}_{\overline{k}}} \mathbf{V}_{max}(\mathbf{S}_{k,b_k})$, where $\mathbf{P}^{\perp}_{\mathbf{C}_{\overline{k}}}$ represents the projection onto the orthogonal complement of $\mathbf{C}_{\overline{k}}$. To further simplify, consider the eigen decomposition of $\mathbf{W}_{k,b_k} = \mathbf{V}_{k,b_k} \mathbf{\Lambda}_{k,b_k} \mathbf{V}^H_{k,b_k}$. Then we can write $\mathbf{S}_{k,b_k} = \mathbf{C}_{k,b_k} \mathbf{V}_{k,b_k} \mathbf{\Lambda}_{k,b_k} \mathbf{V}^H_{k,b_k} \mathbf{C}_{k,b_k}$. Multiplication of the semi-unitary matrix $\mathbf{C}_{k,b_k}$ with the unitary matrix $\mathbf{V}_{k,b_k}$ results in a semi-unitary matrix itself and thus the eigenvalues of $\mathbf{S}_{k,b_k}$ are same as that of $\mathbf{W}_{k,b_k}$ and the corresponding eigenvectors become same as that of $\mathbf{W}_{k,b_k}$ left multiplied by $\mathbf{C}_{k,b_k}$. Finally we rewrite $\mathbf{g}_k$ as,

$$\mathbf{g}'_k = \mathbf{P}^{\perp}_{\mathbf{C}_{\overline{k}}} \mathbf{C}_{k,b_k} \mathbf{V}_{max}(\mathbf{W}_{k,b_k}) \,. \tag{6.35}$$

Further optimizing w.r.t p_k leads to the following water filling solution for the power,

$$p_k = \left(\frac{u_k}{\mu_{b_k}} - \frac{1}{\mathbf{g}'^H_k \mathbf{S}_{k,b_k} \mathbf{g}'_k} \right)^+, \tag{6.36}$$

where $(x)^+ = \max\{0, x\}$ and the Lagrange multipliers μ_c are adjusted (example by bisection) to satisfy the power constraints.

6.9.5 Reduced Order ZF with Partial CSIT

In this section, we consider the BF to be a reduced order ZF (RO-ZF) which is introduced in [99]. This can be interpreted as the number of interfering channels to be zero-forced for a user k is much less than K. The RO-ZF BF $\mathbf{g}_k$ can be written as, $\mathbf{g}_k = \frac{\mathbf{P}^{\perp}_{\mathbf{C}_{I_k}} \mathbf{C}_{k,b_k} \mathbf{V}_{max}(\mathbf{W}_{k,b_k})}{||\mathbf{P}^{\perp}_{\mathbf{C}_{I_k}} \mathbf{C}_{k,b_k} \mathbf{V}_{max}(\mathbf{W}_{k,b_k})||}$. Here, $\mathbf{P}_C = \mathbf{C}(\mathbf{C}^H\mathbf{C})^{\#}\mathbf{C}^H$ represent the projection onto the column space of $\mathbf{C}$, $\mathbf{P}^{\perp}_{\mathbf{C}} = \mathbf{I} - \mathbf{P}_C$ is the projection onto its orthogonal complement (# represents the Moore-Penrose pseudo-inverse). For the convenience of analysis, we define the following: I_k denotes the set of user indices for which the ZF is done. $\mathbf{C}_{I_k}$ represents the matrix of all the user eigenvector space in I_k. Complexity in the RO-ZF case will be about half of that of full ZF (multiplying the $M \times LK$ $\mathbf{C}$ by a triangular $LK \times LK$ instead of a full $LK \times LK$, computation of the $LK \times LK$ inverse or triangular factor takes $O((LK)^3)$ operations, with a smaller factor if only a triangular factor is needed and not a full inverse).

6.9.6 Large System Analysis for RO-ZF and Full Order ZF

In this section we consider the large system analysis for the RO-ZF scheme proposed in this chapter and also the full order ZF (full order means $|I_k| = K - 1, \forall k$). We assume that the LS channel estimation error $\widetilde{\sigma^2}$ remains finite with SNR. If for instance the error variance on the channel estimate would be inversely proportional to SNR, then at high SNR the channel estimate becomes exact and the covariance information does not bring any improvements. The channel estimation error remaining finite can be representative of the UL power being much less than the DL

power (channel estimation from UL pilots and using TDD reciprocity). The ESEINR (Expected Signal to Expected Interference plus Noise Ratio) can be written as,

$$\begin{aligned}\gamma_k^{RO-ZF} &= \frac{P_{S,k}}{P_{I,k}+1} \\ &= \frac{p_k \mathbf{g}_k'^H \mathbf{S}_{k,b_k} \mathbf{g}_k'}{\sum_{i=1,i\neq k}^{K} p_i \mathbf{g}_i'^H \mathbf{S}_{k,b_i} \mathbf{g}_i' + 1}, \\ \implies \mathbf{g}_k'^H \mathbf{S}_{k,b_k} \mathbf{g}_k' &= \frac{\mathbf{v}_{k,b_k}^H \mathbf{C}_{k,b_k}^H \mathbf{P}^{\perp}_{\mathbf{C}_{\bar{k}}} \mathbf{S}_{k,b_k} \mathbf{P}^{\perp}_{\mathbf{C}_{\bar{k}}} \mathbf{C}_{k,b_k} \mathbf{v}_{k,b_k}}{\left\| \mathbf{P}^{\perp}_{\mathbf{C}_{\bar{k}}} \mathbf{C}_{k,b_k} \mathbf{v}_{k,b_k} \right\|^2}.\end{aligned} \tag{6.37}$$

Consider the eigen decomposition of $\mathbf{W}_{k,b_k} = \mathbf{V}_{k,b_k} \mathbf{\Lambda}_{k,b_k} \mathbf{V}_{k,b_k}^H$ and we denote $\mathbf{V}_{max}(\mathbf{W}_{k,b_k}) = \mathbf{v}_{k,b_k}$,

$$\begin{aligned}\mathbf{v}_{k,b_k}^H \mathbf{C}_{k,b_k}^H \mathbf{P}^{\perp}_{\mathbf{C}_{\bar{k}}} \mathbf{S}_{k,b_k} \mathbf{P}^{\perp}_{\mathbf{C}_{\bar{k}}} \mathbf{C}_{k,b_k} \mathbf{v}_{k,b_k} &\overset{(a)}{=} \frac{1}{M_{b_k}^2} \mathrm{tr}\{\mathbf{P}^{\perp}_{\mathbf{C}_{\bar{k}}}\}^2 \mathbf{v}_{k,b_k}^H \mathbf{W}_{k,b_k} \mathbf{v}_{k,b_k} \\ &\overset{(b)}{=} \frac{1}{M_{b_k}^2} \mathrm{tr}\{\mathbf{P}^{\perp}_{\mathbf{C}_{\bar{k}}}\}^2 \lambda_{max}(\mathbf{W}_{k,b_k}), \\ \mathbf{g}_k'^H \mathbf{S}_{k,b_k} \mathbf{g}_k' &= \frac{1}{M_{b_k}} (M_{b_k} - \sum_{i=1,i\in I_k}^{K} L_{i,b_k}) \lambda_{max}(\mathbf{W}_{k,b_k}),\end{aligned} \tag{6.38}$$

where we substituted $\left\| \mathbf{P}^{\perp}_{\mathbf{C}_{\bar{k}}} \mathbf{C}_{k,b_k} \mathbf{v}_{k,b_k} \right\|^2 = \left\| \mathbf{v}_{k,b_k}^H \mathbf{C}_{k,b_k}^H \mathbf{P}^{\perp}_{\mathbf{C}_{\bar{k}}} \mathbf{C}_{k,b_k} \mathbf{v}_{k,b_k} \right\|$ using the property of projection matrices, $\mathbf{P}^{\perp}_{\mathbf{C}_{\bar{k}}} \mathbf{P}^{\perp}_{\mathbf{C}_{\bar{k}}} = \mathbf{P}^{\perp}_{\mathbf{C}_{\bar{k}}}$. Also, (a) in (6.38) follows from Lemma 4 in Appendix VI of [14], that $\mathbf{x}_N^H \mathbf{A}_N \mathbf{x}_N \xrightarrow{N\to\infty} (1/N) tr \mathbf{A}_N$ when the elements of $\mathbf{x}_N$ are i.i.d. with variance $1/N$ and independent of $\mathbf{A}_N$, and similarly when $\mathbf{y}_N$ is independent of $\mathbf{x}_N$, that $\mathbf{x}_N^H \mathbf{A}_N \mathbf{y}_N \xrightarrow{N\to\infty} 0$. Using this Lemma, $\mathbf{C}_{k,b_k}^H \mathbf{P}^{\perp}_{\mathbf{C}_{\bar{k}}} \mathbf{C}_{k,b_k} = \frac{1}{M_{b_k}} \mathrm{tr}\{\mathbf{P}^{\perp}_{\mathbf{C}_{\bar{k}}}\}$ and (b) follows from the fact that $\mathbf{v}_{k,b_k}$ (max eigenvector from $\mathbf{V}_{k,b_k}$) is orthogonal to all the other columns of $\mathbf{V}_{k,b_k}$ except the one corresponding to $\lambda_{max}(\mathbf{W}_{k,b_k})$. Further, by the law of large numbers, $P_{S,k} - \overline{P}_{S,k} \xrightarrow[a.s]{M\to\infty} 0$, where,

$$\overline{P}_{S,k} = (1 - \frac{\sum_{i=1,i\in I_k}^{K} L_{i,b_k}}{M_{b_k}}) \lambda_{max}(\mathbf{W}_{k,b_k}) p_k \tag{6.39}$$

Next, we consider the terms in $P_{I,k}$,

$$\mathbf{g}_i'^H \mathbf{S}_{k,b_i} \mathbf{g}_i' = \frac{\mathbf{v}_{i,b_i}^H \mathbf{C}_{i,b_i}^H \mathbf{P}^{\perp}_{\mathbf{C}_{\bar{i}}} \mathbf{S}_{k,b_i} \mathbf{P}^{\perp}_{\mathbf{C}_{\bar{i}}} \mathbf{C}_{i,b_i} \mathbf{v}_{i,b_i}}{\left\| \mathbf{P}^{\perp}_{\mathbf{C}_{\bar{i}}} \mathbf{C}_{i,b_i} \mathbf{v}_{i,b_i} \right\|^2}. \tag{6.40}$$

If $k \in I_i$, then $\mathbf{P}^{\perp}_{\mathbf{C}_{\bar{i}}}$ is orthogonal to the columns of $\mathbf{C}_{k,b_i}$ and thus $\mathbf{g}_i^H \mathbf{S}_{k,b_i} \mathbf{g}_i = 0$ else, using Lemma 4, we obtain $\mathbf{v}_{i,b_i}^H \mathbf{C}_{i,b_i}^H \mathbf{P}^{\perp}_{\mathbf{C}_{\bar{i}}} \mathbf{S}_{k,b_i} \mathbf{P}^{\perp}_{\mathbf{C}_{\bar{i}}} \mathbf{C}_{i,b_i} \mathbf{v}_{i,b_i} = \frac{1}{L_{i,b_i}} \mathrm{tr}\{\mathbf{C}_{i,b_i}^H \mathbf{P}^{\perp}_{\mathbf{C}_{\bar{i}}} \mathbf{S}_{k,b_i} \mathbf{P}^{\perp}_{\mathbf{C}_{\bar{i}}} \mathbf{C}_{i,b_i}\}$.

$$\begin{aligned}\frac{1}{L_{i,b_i}} \mathrm{tr}\{\mathbf{C}_{i,b_i}^H \mathbf{P}^{\perp}_{\mathbf{C}_{\bar{i}}} \mathbf{S}_{k,b_i} \mathbf{P}^{\perp}_{\mathbf{C}_{\bar{i}}} \mathbf{C}_{i,b_i}\} &\overset{(c)}{=} \frac{1}{M_{b_i}} \mathrm{tr}\{\mathbf{P}^{\perp}_{\mathbf{C}_{\bar{i}}} \mathbf{S}_{k,b_i} \mathbf{P}^{\perp}_{\mathbf{C}_{\bar{i}}}\} \\ \frac{1}{M_{b_i}} \mathrm{tr}\{\mathbf{W}_{k,b_i} \mathbf{C}_{k,b_i}^H \mathbf{P}^{\perp}_{\mathbf{C}_{\bar{i}}} \mathbf{C}_{k,b_i}\} &\overset{(d)}{=} \frac{1}{M_{b_i}^2} \mathrm{tr}\{\mathbf{P}^{\perp}_{\mathbf{C}_{\bar{i}}}\} \mathrm{tr}\{\mathbf{W}_{k,b_i}\} \\ &= \frac{1}{M_{b_i}} (1 - \frac{\sum_{r=1,r\in I_i}^{K} L_{r,b_i}}{M_{b_i}}) \sum_{l=1}^{L_{k,b_i}} \zeta_{k,b_i}^{(l)},\end{aligned} \tag{6.41}$$

where (c) and (d) are obtained by using Lemma 4 from [15]. Further we obtain $\mathbf{g}_i'^H \mathbf{S}_{k,b_i} \mathbf{g}_i' = \frac{1}{M_{b_i}} \sum_{l=1}^{L_{k,b_i}} \zeta_{k,b_i}^{(l)}$. Finally, we obtain the ESEINR in the large system limit as, $\gamma_k^{RO-ZF} - \overline{\gamma}_k^{RO-ZF} \xrightarrow[a.s]{M\to\infty} 0$,

$$\overline{\gamma}_k^{RO-ZF} = \frac{(1-\frac{\sum_{i=1,i\in I_k}^{K} L_{i,b_k}}{M_{b_k}})\lambda_{max}(\mathbf{W}_{k,b_k})p_k}{\frac{1}{M_{b_i}} \sum_{i=1,k\notin I_i}^{K} \sum_{l=1}^{L_{k,b_i}} \zeta_{k,b_i}^{(l)} p_i + 1} \tag{6.42}$$

For the full order ZF, the interference power vanishes from the ESEINR terms,

$$\overline{\gamma}_k^{ZF} = (1 - \frac{\sum_{i=1,i\neq k}^{K} L_{i,b_k}}{M_{b_k}})\lambda_{max}(\mathbf{W}_{k,b_k})p_k \tag{6.43}$$

The power updates for the RO-ZF BF can be shown to be as similar to the interference aware water filling as shown in [99] and the simplified expressions directly follow from the above equations as,

$$p_k = (\frac{u_k}{\mu_{b_k} + \sigma_k^{(2)}} - \frac{1}{\sigma_k^{(1)}})^+, \tag{6.44}$$

where, $\sigma_k^{(2)} = \frac{1}{M_{b_k}} \sum_{i=1,i\notin I_k}^{K} \beta_i \sum_{l=1}^{L_{i,b_k}} \zeta_{i,b_k}^{(l)}$, $\sigma_k^{(1)} = (1 - \frac{\sum_{i=1,i\in I_k}^{K} L_{i,b_k}}{M_{b_k}})\lambda_{max}(\mathbf{W}_{k,b_k})$, $\beta_i = u_k(\frac{1}{r_{\bar{k}}} - \frac{1}{r_k})$.
Computation of eigenvalues $\zeta_{k,b_i}^{(r)}$ of $\mathbf{W}_{k,b_i}$: from Section III,

$$\begin{aligned} \mathbf{W}_{k,b_i} &= \check{\mathbf{c}}_{k,b_i} \check{\mathbf{c}}_{k,b_i}^H + \widetilde{\mathbf{D}}_{k,b_i}, \\ \check{\mathbf{c}}_{k,b_i} &= \widehat{\mathbf{D}}_{k,b_i}^{1/2} \widehat{\mathbf{c}}_{k,b_i}, \forall i,k \end{aligned} \tag{6.45}$$

In (6.25), we assume that all the eigenvalues are equal and positive, i.e $\mathbf{D}_{k,b_i} = \eta_{k,b_i}\mathbf{I}$, $\widetilde{\mathbf{D}}_{k,b_i} = \widetilde{\eta}_{k,b_i}\mathbf{I}$. Thus the eigenvalues of $\mathbf{W}_{k,b_i}$ can be shown to be $\zeta_{k,b_i}^{(1)} = \lambda_{max}(\mathbf{W}_{k,b_i}) = \left\|\check{\mathbf{c}}_{k,b_i}\right\|^2 + \widetilde{\eta}_{k,b_i}$ and $\zeta_{k,b_i}^{(2)} = ... = \zeta_{k,b_i}^{(L_{k,b_i})} = \widetilde{\eta}_{k,b_i}$, where $\widetilde{\eta}_{k,b_i} = \frac{\widetilde{\sigma}_{k,b_i}^2 \eta_{k,b_i}}{\widetilde{\sigma}_{k,b_i}^2 + \eta_{k,b_i}}$, using the definition of $\widetilde{\mathbf{D}}_{k,b_i}$ from (6.29). $\lambda_{max}(\mathbf{W}_{k,b_i})$ is random since $\check{\mathbf{c}}_{k,b_i}$ is random. By the law of large numbers (assuming L_{k,b_i} is large but finite and $<< N_t$) we replace it by the expectation which can be computed as follows. $\mathrm{E}(\lambda_{max}(\mathbf{W}_{k,b_i})) = \mathrm{E}(\widehat{\mathbf{c}}_{k,b_i}^H \widehat{\mathbf{D}}_{k,b_i} \widehat{\mathbf{c}}_{k,b_i}) + \widetilde{\eta}_{k,b_i}$. This gets simplified as, $\mathrm{E}\left(\lambda_{max}(\mathbf{W}_{k,b_i})\right) = L_{k,b_i}\widehat{d}_{k,b_i} + \widetilde{\eta}_{k,b_i}$, where $\widehat{d}_{k,b_i} = \frac{\eta_{k,b_i}^2}{\eta_{k,b_i} + \widetilde{\sigma}_{k,b_i}^2}$ from (6.26) ($\widehat{\mathbf{D}}_{k,b_i} = \widehat{d}_{k,b_i}\mathbf{I}$) and $\mathrm{E}(\widehat{\mathbf{c}}_{k,b_i}^H \widehat{\mathbf{c}}_{k,b_i}) = L_{k,b_i}$ from (6.26).

6.9.7 Optimization of the ZF Order

In this section, we consider an alternating optimization algorithm (Algorithm 11) which computes the reduced ZF order for each user (I_k). We define here $\theta_{i,b_j} = \sum_{l=1}^{L_{i,b_j}} \zeta_{i,b_j}^{(l)}$, as the channel strength from BS b_j to user i. Note that at finite dimension MIMO, not only the channel strengths but also the relative orientation of the channel vectors count. However, in MaMIMO with multiple of identity covariances, there is no orientation issue, only the channel strengths count. So the user ordering is simple.

Algorithm 11: Reduced Zero-Forcing Order Determination

Given: $K, M, \sigma^2, \theta_{i,b_j}, \forall i, j$, with ordering $\theta_{1,b_j} \geq \theta_{2,b_j} \geq ... \geq \theta_{K,b_j}$. Start with $I_k = \emptyset, \forall k$, i.e., $\mathbf{g}_k^{(0)} = \mathbf{h}_{k,b_k}$.

for c = 1,..,C
Compute the interference powers received at all users from BS c. Find the link causing the maximum interference. Let it be BF $\mathbf{g}_k$ to user l.
Add ZF for the corresponding maximum interference causing channel link. i.e. $I_k = I_k \cap l$.
Update $\mathbf{g}_k^{(t)}$ ($\mathbf{g}_k$ corresponding to the updated I_k) , such that $b_k = c$.
Update the user powers p_k using (6.44).
Compute the WSR. If the WSR is decreased, exit the loop. Otherwise continue with next iteration $(t+1)$.
end for

6.9.8 Simulation Results

In this section, we present the Ergodic Sum Rate Evaluations for BF design for the various channel estimates. Monte Carlo evaluations of ergodic sum rates are done, where we consider a pathwise or low rank channel model as in section 6.9.1, with number of paths = channel covariance rank $L = 4$. In the figures, "LSA" refers to large system approximation and "Chnl Est" refers to

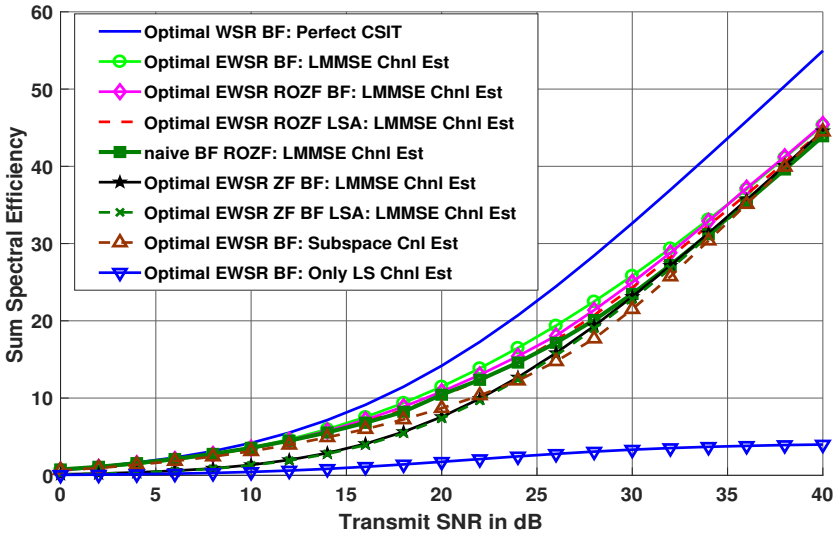


Figure 6.4: Sum Rates, $M = 64, K = 10, L = 4, \widetilde{\sigma^2} = 0.1$.

channel estimate. In these simulations, the deterministic channel estimation error (i.e, $\widetilde{\sigma}^2$) does not go to zero as SNR $-> \infty$ but remains constant. Otherwise, at high SNR it is the channel estimate that dominates, and the partial CSIT at high SNR will just become perfect CSIT. The simulations in Figure 6.4 show that exploiting also the channel error covariance information can lead to substantial performance gains compared to just using LS channel estimate. The naive channel estimate based partial CSIT BF approaches are suboptimal. We also compare optimal BF and full and reduced order ZF BF, based on LMMSE channel estimates plus error covariance.

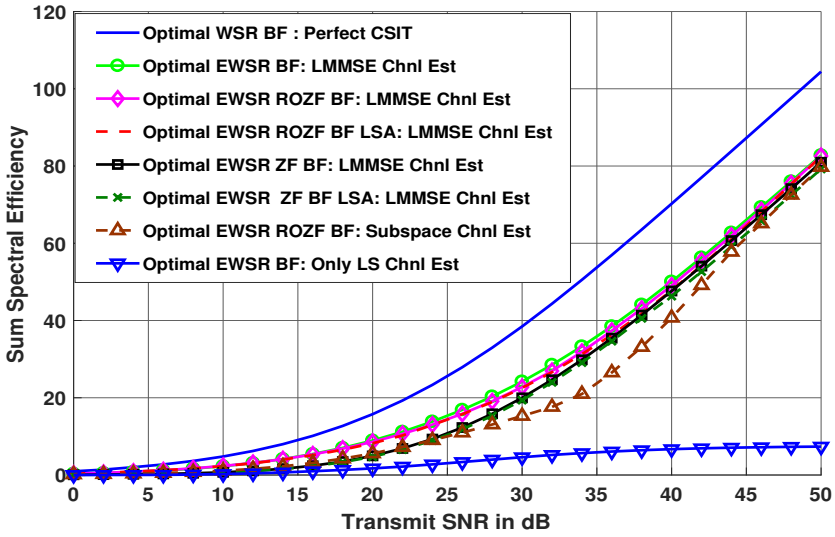


Figure 6.5: Sum Rates, $M = 128, K = 15, L = 4, \widetilde{\sigma^2} = 0.1$.

Note that in the case of reduced-rank channel covariances considered here, ZF BF may still be possible, even with partial CSIT. At high SNR, ZF BF is optimal. At low and intermediate SNRs, RO-ZF is able to outperform (full order) ZF and it is quite close to the optimal BF [100]. Figure 6.5 are for increased dimensions. Further these simulations suggest that the large system approximations for ROZF and ZF are accurate even for finite values for M, K, L.

6.9.9 Conclusions

Concluding Remarks 5

- In this chapter, we extend the concept of reduced order ZF BF to partial CSIT. Simulation results indicate that our RO-ZF BF scheme has a performance very close to the optimal BFs, but with much less complexity compared to the full order ZF.
- We also propose an alternating optimization algorithm which computes the optimal ZF order for each user.
- Moreover, we show (elsewhere) the improvement in performance by using an LMMSE channel estimate compared to just having LS estimates, and by furthermore properly exploiting all covariance information.
- Further work will include the exploitation of the large system analysis for the optimization of the reduced order for lesser complexity in RO-ZF BF.

Chapter 7

STOCHASTIC GEOMETRY BASED LARGE SYSTEM ANALYSIS

The development of Massive Multi-Input Multi-Output MIMO (MaMIMO) technology [66] enables high throughput for the next generation of wireless systems. However, MaMIMO systems have precise requirements for Channel State Information at the Tx (CSIT) which is more difficult to acquire than CSI at the Rx (CSIR). One of the pioneering works which talks about the effect of imperfect channel knowledge on the capacity is [101]. Indeed, in Massive Multiple-Input Single-Output (MaMISO) systems, the received interference and possibly signal powers converge to their expected value (channel hardening effect) due to the law of large numbers. The large system analysis becomes an important topic to consider since Monte-Carlo simulations involving large numbers of antennas and user equipments (UEs) become cumbersome in a MaMISO system. Also, simulations do not allow to see immediately how performance depends on various system parameters. A major breakthrough in the large system analysis (LSA) of MaMISO systems came in [14], where Wagner et al. develop results in random matrix theory to obtain deterministic equivalents for the signal-to-interference-plus-noise ratio (SINR) and thus the rate expression for regularized zero forcing (R-ZF) precoding under partial channel knowledge. Further to this, quite a few papers that build on the results from [14] for LSA appeared in [95, 102–105]. The very recent work [105] extends the LSA results in [14] to a Rician fading channel with perfect CSIT. However, to simplify the analysis, the authors therein consider identical correlation matrix for all the users in the system which is impractical in MaMIMO or mmWave system. Some of our own recent studies on large system analysis can be found in [99, 106, 107]. In [99], we focus on the asymptotic analysis for ZF and reduced order ZF BF under a simplified case of user channel covariances which are multiple of identity (with distinct scale factors for different users).

It is true that the asymptotic analysis results for MaMIMO system have evolved a lot since [14] with more practical channel models involving Rician fading etc. However, we observe that none of these works give analytical insights into the behavior of the system with respect to the different types of partial CSIT and different (including optimal) BF designs. Note that assumptions on the quality of the CSIT greatly impact the low or high signal-to-noise ratio (SNR) behavior of the ergodic capacity. The impact of quantized CSIT in the case of finite rate feedback channels is reported in [108, 109], where they analyze the per-user spectral efficiency (SE) for the MISO broadcast channel (BC) for conventional ZF BF.

An introduction to the literature on stochastic geometry can be found in [110]. In stochastic geometry, generally the location of the nodes in the wireless network is modeled as random, fol-

lowing for example a poisson point process. In stochastic geometry based methods [111,112], the location of the users being random, their geographic distribution then induces a certain probability distribution for the channel attenuation. This leads to results on the coverage probability, the capacity, the outage probability and other fundamental limits in wireless networks. Whereas most stochastic geometry work focuses on the distribution of the attenuation, here we consider an extension to multi-antenna systems. The multipath propagation for the various users leads to randomized angles of arrival at the base station (BS) which can be translated into spatial channel response contributions that depend on the antenna array response. In the MaMISO regime, it has been observed and exploited that despite complex multipath propagation, the channel covariance matrix tends to be low rank. Exploiting the randomized nature of the user and scatterer positions and making abstraction of the antenna array response, we propose to model the user channel subspaces as isotropically randomly oriented. This allows us to assume the eigenvectors of the channel covariance matrix to be Haar distributed. Moreover, this is identically and independently distributed for all users. The experimental studies conducted in [113, 114] show that for a typical cellular configuration with a tower-mounted BS, the angular spread of the incoming or outgoing rays from the BS to a UE is very small, resulting in a sparse representation of the user channel in the virtual angular domain. This has been observed even in below 6−GHz bands [113]. To further justify the channel model used in this chapter, we refer to [115], where they evaluate the sum rate performance of ZF precoding in MU-MIMO system with diverse correlation patterns across the UEs. The correlation models used there are parameterized by the measured data from a recent 2.53 GHz urban macrocellular compaign conducted in Cologne, Germany. Their measurements show the diverse angular patterns across different user terminals and their analysis further validated the importance of having more physically motivated models to evaluate accurately the SE performance.

The majority of the existing work on MaMISO SE analysis focus on spatially uncorrelated user channel covariance matrices [116, 117]. While this assumption makes the asymptotic analysis very simple, a more realistic approach in a MaMISO or a mmWave system is when the user channel covariance matrix is spatially correlated. Even though for a single user MIMO system, the spatial correlation can be detrimental to the system performance, for a multi-user MISO system having different channel covariance matrices spanning mutually orthogonal subspaces can be advantageous [118, 119]. The authors in [118] introduce the term transmit correlation diversity to capture this effect. The spatial correlation can be induced by either the significant multipath components originating from some spatial direction or by spatially dependent antenna patterns and polarization at the BS. The achievable SE of MaMIMO systems are studied under spatially uncorrelated [120] or spatially correlated Rayleigh fading [61,102]. In a very recent study done by Özgecan Özdogan et al. in [121], the authors go one step further in analyzing MaMIMO system with Rician fading channels. The channel model being considered is composed of a deterministic line of sight component and a stochastic non-line-of-sight component modeled using the spatially correlated covariance matrices. They analyze the large system SE behavior of linear minimum mean square error (LMMSE), EW LMMSE and LS channel estimates and shows that the LMMSE estimate performs better than other sub-optimal estimates through simulations. Nevertheless, they consider the simpler and sub-optimal BF, maximal ratio precoding, for the DL SE analysis. In [16] we have extended the LSA of [14] to a scenario with users having different channel covariance matrices and BF techniques with partial CSIT. However, due to the abundance of different covariance matrices, the resulting deterministic analysis does not allow for much insight. The multi-antenna stochastic geometry aspect introduced here reduces such LSA analysis back to the simplicity of the case of multiple of identity covariance matrices.

7.0.1 Summary of this Chapter

In this chapter, we focus on two important questions:
1) Under transmit correlation diversity in a multiuser multicell MaMIMO system, how is the SE affected by channel estimation error for different MaMISO limit approximations of the ergodic capacity under different schemes for channel estimation?
2) How do the random matrix theory tools can be exploited to analyze the SE behaviour under extreme SNR regions?
Following are the contributions in the chapter to tackle these questions:

- We first consider the various channel estimates: linear minimum mean square error (LMMSE), least squares (LS) and subspace projection. Further we review the Expected Signal and Interference Power WSR (ESIP-WSR) BF design for the expected weighted sum rate (EWSR) criterion in the MaMISO limit.

- We evaluate the ergodic sum rate performance for LS, LMMSE and subspace projection channel estimators with an upper bound of the EWSR BF (which we call ESIP-WSR) which is tight in (a certain) massive MISO limit. Simulation results suggest that there is substantial gain by exploiting the channel covariance information compared to just using the LS estimates.

- The analysis presented in the chapter provides accurate (also validated in the simulations) SE expressions under realistic channel estimation quality which are useful at any operating SNR. Moreover, these SE expressions are very simple and provide analytical insights into the system behavior and depends only on few system parameters such as channel power across multipaths, Tx power, rank of the channel covariance matrix and number of Tx antennas. Compared to our previous work [106], we derive simplified sum rate expressions at low and high SNR for the various BFs (ESIP-WSR, naive and EWSMSE) for the various channel estimates, which clearly shows the SNR offset for the sub-optimal BFs compared to the proposed optimal EWSR BF.

- We furthermore provide certain illustrative examples which are special cases of the ESIP-WSR BF such as perfect channel CSIT case and covariance only CSIT (CoCSIT) scenario where only the channel covariance information is known at the BS. We show that we can obtain analytical expressions for the implicit equations which need to be solved as part of the large system analysis, which in fact provide analytical insights into the system behavior. We also provide simplified sum rate expressions at high SNR for the CoCSIT case and obtain the rate offset with respect to the perfect CSIT case under different channnel estimation quality.

- For the sum rate analysis at extreme SNR regions, we consider two scenarios. One where the channel estimation error is inversely proportional the SNR and the second scenario is where the channel estimation error remains constant with SNR (finite rate feedback channels). When the channel estimation error is inversely proportional to the SNR, at high SNR, ESIP-WSR/EWSMSE/naive BFs converge to the perfect CSIT BF performance since channel estimation error converges to zero. With constant channel estimation error, it is observed that the EWSMSE and naive BFs with LMMSE/Subspace channel estimates saturate at high SNR which is explained by the derived SNR offset. The SNR offset for the EWSMSE or naive BFs shows that at high SNR, the interference power also increases along

with the SNR, since no ZF to the interfering channels happen at high SNR. However, the ESIP-WSR design does not exhibit a saturation.

In addition, since the publication of the conference version of this work [64, 106], another work on MaMISO SE analysis using similar stochastic geometry based randomization has been noted [122] and we would like to discuss the differences of our work from them. We observe that [122] deals with the deterministic equivalents of the upper and lower bounds to the ergodic capacity and not on the asymptotic tightness of the approximations. Moreover, for the simulations, they consider a random partial Fourier correlation model which is motivated by the typical uniform linear array (ULA) in MIMO systems. However, this channel model is too approximate. We remark that, the analysis presented here can be readily extended to the case of a Rician fading channel model as is considered in [121].

7.1 Massive MISO Stochastic Geometry based Large System Analysis

7.1.1 MISO IBC Signal Model

We consider here an Interfering Broadcast Channel (IBC) with C cells and a total of K single antenna users. We shall consider a system-wide numbering of the users. User k is served by BS b_k. $\mathbf{h}_{k,b_i}$ is the $M_{b_k} \times 1$ channel from BS b_i to user k. For notational convenience, we use an abbreviated notation for the direct channels (channel from BS b_k to the serving user k), i.e., $\mathbf{h}_{k,b_k}$ will be denoted as $\mathbf{h}_k$. The received signal at user k in cell b_k is

$$\mathbf{y}_k = \underbrace{\mathbf{h}_k^H \mathbf{g}_k x_k}_{\text{signal}} + \underbrace{\sum_{\substack{i \neq k \\ b_i = b_k}} \mathbf{h}_k^H \mathbf{g}_i x_i}_{\text{intracell interf.}} + \underbrace{\sum_{j \neq b_k} \sum_{i:b_i=j} \mathbf{h}_{k,j}^H \mathbf{g}_i x_i}_{\text{intercell interf.}} + \mathbf{v}_k \tag{7.1}$$

where x_k is the intended (white, unit variance) scalar signal stream, The Rx signal (and hence the channel) is assumed to be scaled so that we get for the noise $v_k \sim \mathcal{CN}(0,1)$. BS c serves $K_c = \sum_{i:b_i=c} 1$ users. The $M_{b_k} \times 1$ spatial Tx filter or beamformer (BF) is $\mathbf{g}_k$. The Tx power constraint at BS c is, $\sum_{i:b_i=c} \left\|\mathbf{g}_i\right\|^2 \leq P_c$.

7.1.2 Channel and CSIT Model

For simplicity, we omit all the user indices k. Each zero mean MISO channel is modeled according to Karhunen-Loeve representation [21] as

$$\begin{aligned} \mathbf{h} &= \mathbf{C}\mathbf{D}^{1/2}\mathbf{c}, \\ \mathbf{R_{hh}} &= \mathbf{C}\mathbf{D}\mathbf{C}^H, \end{aligned} \tag{7.2}$$

where $\mathbf{R_{hh}}$ is the covariance matrix and $\mathbf{c} \sim \mathcal{CN}(0, \mathbf{I}_L)$ are the Rayleigh fading multipath gains in the eigen domain. Here $\mathbf{C}$ is the $M \times L$ eigenvector matrix of the reduced rank channel covariance $\mathbf{R_{hh}}$ with diagonal eigenvalue matrix $\mathbf{D}$. This reduced rank covariance matrix of user channels typically occurs in realistic MaMISO channels due to the limited angular spread of the multipath components [123]. The rank corresponds to an equivalent number of linearly independent multipath components. The total sum rank across all user channels from BS c, $\sum_{k=1}^{K} L_{k,c}$ is assumed to be less than M_c, where $L_{k,c}$ is the channel rank between user k and BS c.

Since the focus of this chapter is to study the effect of channel estimation error, we assume that we are given a deterministic Least-Squares (LS) channel estimate

$$\widehat{\mathbf{h}}_{LS} = \mathbf{h} + \widetilde{\mathbf{h}}, \tag{7.3}$$

where $\mathbf{h}$ is the true MISO channel, and the error is modeled as circularly symmetric white Gaussian noise $\widetilde{\mathbf{h}} \sim \mathcal{CN}(0, \widetilde{\sigma}^2\mathbf{I})$. The error $\widetilde{\sigma}^2$ is given a priori. Now, assuming the channel covariance subspace is known, the LMMSE channel estimate can be obtained as

$$\widehat{\mathbf{h}} = \mathbf{C}\mathbf{D}\mathbf{C}^H\left(\mathbf{C}\mathbf{D}\mathbf{C}^H + \widetilde{\sigma}^2\mathbf{I}\right)^{-1}\widehat{\mathbf{h}}_{LS}. \tag{7.4}$$

As in the previous chapters, we refer to the Appendix A for a detailed derivation of the LMMSE channel estimate above. Applying the matrix inversion lemma and exploiting $\mathbf{C}^H\mathbf{C} = \mathbf{I}_L$, this simplifies to

$$\begin{aligned}\widehat{\mathbf{h}} &= \mathbf{C}\left(\widetilde{\sigma}^2\mathbf{D}^{-1} + \mathbf{I}\right)^{-1}\mathbf{C}^H\widehat{\mathbf{h}}_{LS} \\ &= \mathbf{C}\widehat{\mathbf{D}}^{1/2}\widehat{\mathbf{c}},\end{aligned} \tag{7.5}$$

where $\widehat{\mathbf{D}} = \left(\widetilde{\sigma}^2\mathbf{D}^{-1} + \mathbf{I}\right)^{-1}\mathbf{D}$ and $\widehat{\mathbf{c}} = \mathbf{D}^{-1/2}(\sigma^2\mathbf{D}^{-1} + \mathbf{I})^{-1/2}\mathbf{C}^H\widehat{\mathbf{h}}_{LS}$ with $\mathbf{R}_{\widehat{\mathbf{c}}\widehat{\mathbf{c}}} = \mathbf{I}$.

$$\begin{aligned}\mathbf{R}_{\widetilde{\mathbf{h}}\widetilde{\mathbf{h}}} &= \mathbf{C}\widetilde{\mathbf{D}}\mathbf{C}^H \\ &= \mathbf{C}[\mathbf{D} - \left(\widetilde{\sigma}^2\mathbf{D}^{-1} + \mathbf{I}\right)^{-1}\mathbf{D}]\mathbf{C}^H.\end{aligned} \tag{7.6}$$

Further exploiting the orthogonality property of the LMMSE channel estimate, we can write

$$\begin{aligned}\mathbf{S} = \mathrm{E}_{\mathbf{h}|\widehat{\mathbf{h}}}\left(\mathbf{h}\mathbf{h}^H\right) &= \widehat{\mathbf{h}}\widehat{\mathbf{h}}^H + R_{\widetilde{\mathbf{h}}\widetilde{\mathbf{h}}} \\ &= \mathbf{C}\mathbf{W}_L\mathbf{C}^H, \text{ where} \\ \mathbf{W}_L &= \widehat{\mathbf{D}}^{1/2}\widehat{\mathbf{c}}\widehat{\mathbf{c}}^H\widehat{\mathbf{D}}^{1/2} + \widetilde{\mathbf{D}}.\end{aligned} \tag{7.7}$$

For the convenience of analysis in the following sections, we define the following quantities, $\mathbf{C}^H\mathbf{h}_{LS} = \widehat{\mathbf{d}} = \mathbf{d} + \widetilde{\mathbf{d}}$, where $\mathbf{d} = \mathbf{C}^H\mathbf{h} \sim \mathcal{CN}(\mathbf{0}, \mathbf{D})$ and $\widetilde{\mathbf{d}} \sim \mathcal{CN}(\mathbf{0}, \widetilde{\sigma}^2\mathbf{I}_L)$.

In this chapter, we analyze two scenarios where the channel estimation quality indicated by $\widetilde{\sigma}^2$ behaves differently. First, we consider the case when the channel estimation error is inversely proportional to the SNR, so $\widetilde{\sigma}^2 \propto \frac{1}{P}$ since noise variance is assumed to be 1. However, it is difficult to meet the required CSIT quality particularly in the frequency division duplexed (FDD) systems. At the UE, DL training can be used to obtain the CSIT. But obtaining CSIT in the uplink requires feedback from the UE due to the lack of channel reciprocity. This leads to the finite rate feedback model [108], where each UE feedbacks the estimated channel information through finite number of bits. Motivated by this, we also consider the case of constant channel estimation error in the uplink. Even though in this chapter, we do not explicitly consider the pilot contamination effects in the channel estimation phase as in [121], the constant channel estimation scenario considered herein can also be interpreted as representing the case of pilot contamination assuming UL powers are lesser than or not proportional to that of the DL Tx power.

7.1.3 Various Channel Estimates for Partial CSIT

In the MaMISO limit, the EWSR upper bound based BF design with partial CSIT will depend on the quantities $\mathbf{S} = \mathrm{E}_{\mathbf{h}|\widehat{\mathbf{h}}}\left(\mathbf{h}\mathbf{h}^H\right) = \widehat{\mathbf{h}}\widehat{\mathbf{h}}^H + \widetilde{\Theta}$, which will be shown in Section 7.1.4. $\widetilde{\Theta}$ represents the estimation error covariance matrix. In this chapter, we evaluate the SE performance for three

possible channel estimates which depend on different levels of statistical channel knowledge.
(i) *LS Channel Estimate*: We have $\widetilde{\Theta} = \widetilde{\sigma}^2\mathbf{I}$, $\widehat{\mathbf{h}}_{LS} = \mathbf{h}+\widetilde{\mathbf{h}}$ where $\mathbf{h}$ and $\widetilde{\mathbf{h}}$ are independent. In this case, we do not assume any statistical knowledge of the channel.
(ii) *LMMSE Channel Estimate*: In this case, the channel covariance matrix is known (both the covariance subspace $\mathbf{C}$ and eigenvalue matrix $\mathbf{D}$). Then from Section 7.1.2, we have $\mathbf{h} = \widehat{\mathbf{h}} + \widetilde{\mathbf{h}}$ in which $\widehat{\mathbf{h}}$ and $\widetilde{\mathbf{h}}$ are decorrelated and hence independent in the Gaussian case. $\widetilde{\Theta} = \mathbf{R}_{\widetilde{\mathbf{h}}\widetilde{\mathbf{h}}}$ is the posterior covariance. The resulting $\mathbf{S} = \mathrm{E}_{\mathbf{h}|\widehat{\mathbf{h}}}\mathbf{h}\mathbf{h}^H = \widehat{\mathbf{h}}\widehat{\mathbf{h}}^H + \widetilde{\Theta}$ is the (nonlinear) MMSE estimate of $\mathbf{h}\mathbf{h}^H$ (nonlinear because quadratic in $\widehat{\mathbf{h}} = \widehat{\mathbf{h}}_{LS}$ plus a constant). It is unbiased : $E_{\widehat{\mathbf{h}}}\mathbf{S} = E_{\widehat{\mathbf{h}}}E_{\mathbf{h}|\widehat{\mathbf{h}}}\mathbf{h}\mathbf{h}^H = \mathrm{E}_{\mathbf{h}}\mathbf{h}\mathbf{h}^H$ and it is MMSE, hence minimum variance since unbiased. In particular, it also minimizes the variance of $|\mathbf{g}^H\mathbf{h}|^2 = \mathbf{g}^H\mathbf{h}\mathbf{h}^H\mathbf{g} = \mathbf{g}^T \otimes \mathbf{g}^H vec(\mathbf{h}\mathbf{h}^H)$ where $vec(\mathbf{h}\mathbf{h}^H) = \mathbf{h}^* \otimes \mathbf{h}$. Furthermore, we assume that BS c has a prior knowledge on the covariance subspaces $\mathbf{C}$ as well as the eigenvalue matrix $\mathbf{D}$ of its own users and that of the user channels causing inter-cell interference. The user channel covariance subspaces and the eigenvalue matrix can be estimated if the multipath parameters in a mmWave or a MaMIMO channel can be estimated. Multipath parameter estimation can be effectively computed using advanced tensor signal processing based methods as outlined in [124] or in our own work [125] which utilizes variational Bayesian inference methods applied to tensor signal model.
(iii) *Subspace Projection based Channel Estimate*: We also investigate the effect of limiting channel estimation error to the covariance subspace (LMMSE without weighting). The subspace channel estimate is given as

$$\begin{aligned} \widehat{\mathbf{h}}_S &= \mathbf{P}_C\widehat{\mathbf{h}}_{LS} = \mathbf{h} + \mathbf{P}_C\widetilde{\mathbf{h}}_{LS}, \\ \mathbf{R}_{\widetilde{\mathbf{h}}_S\widetilde{\mathbf{h}}_S} &= \widetilde{\sigma}^2\mathbf{P}_C, \end{aligned} \tag{7.8}$$

where $\mathbf{P}_C = \mathbf{C}\mathbf{C}^H$ represents the projection onto the covariance subspace. Further we can write the estimate for $\mathbf{h}\mathbf{h}^H$

$$\begin{aligned} \mathbf{S} &= \widehat{\mathbf{h}}_S\widehat{\mathbf{h}}_S^H + \mathbf{R}_{\widetilde{\mathbf{h}}_S\widetilde{\mathbf{h}}_S} \\ &= \mathbf{C}\mathbf{W}_S\mathbf{C}^H \ \text{ with} \\ \mathbf{W}_S &= \widehat{\mathbf{d}}\widehat{\mathbf{d}}^H + \widetilde{\sigma}^2\mathbf{I}. \end{aligned} \tag{7.9}$$

One remark here is that subspace channel estimator represents a simplification of the LMMSE channel estimator, since it does not require the knowledge of the eigenvalue matrix $\mathbf{D}$ and without negligible performance loss as is validated in our numerical simulations. Another point to be noted is that, combining subspace channel estimator and LMMSE estimator, from (7.5), we can write $\widehat{\mathbf{h}} = \mathbf{C}\mathbf{U}\mathbf{C}^H\widehat{\mathbf{h}}_{LS} = \mathbf{C}\mathbf{U}\widehat{\mathbf{d}}$, where for LMMSE $\mathbf{U}_L = \left(\mathbf{I} + \widetilde{\sigma}^2\mathbf{D}^{-1}\right)^{-1}$ and for subspace $\mathbf{U}_S = \mathbf{I}$. This observation also hints at the possibility of optimizing $\mathbf{U}$ (LMMSE is not necessarily the best) to maximize the ergodic capacity, but this is left for future work.

7.1.4 Beamforming with Partial CSIT

In the following $\mathbf{h}_{k,b_i}$, $\widehat{\mathbf{h}}_{k,b_i}$, $\widetilde{\mathbf{h}}_{k,b_i}$ denote the actual channel, channel estimate and estimation error, respectively between user k and BS b_i. Similarly, we define the quantities, $\widehat{\mathbf{d}}_{k,b_i}$, $\widetilde{\mathbf{d}}_{k,b_i}$, $\mathbf{U}_{k,b_i}$, $\mathbf{C}_{k,b_i}$, $\mathbf{W}_{k,b_i}$, $\mathbf{S}_{k,b_i}$ for the channel between user k and BS b_i. Again, for notational convenience, variables corresponding to the direct channels, $\widehat{\mathbf{d}}_{k,b_k}$, $\widetilde{\mathbf{d}}_{k,b_k}$, $\mathbf{D}_{k,b_k}$, $\widetilde{\mathbf{D}}_{k,b_k}$, $\mathbf{U}_{k,b_k}$, $\mathbf{C}_{k,b_k}$, $\mathbf{W}_{k,b_k}$, $\mathbf{S}_{k,b_k}$ will be denoted as $\widehat{\mathbf{d}}_k$, $\widetilde{\mathbf{d}}_k$, $\mathbf{D}_k$, $\widetilde{\mathbf{D}}_k$, $\mathbf{U}_k$, $\mathbf{C}_k$, $\mathbf{W}_k$, $\mathbf{S}_k$, respectively. Once the CSIT is imperfect, various optimization criteria such as outage capacity can be considered. Motivated by

the ergodic capacity formulations in [97] for point to point MIMO systems, and in [98] for multi-user MISO systems, the design here is based on expected weighted sum rate (EWSR) (and normally with LMMSE channel estimates). In a first stage, the WSR is averaged over the channels given the channel estimates and covariance information (i.e. the partial CSIT), leading to a cost function that can be optimized by the Tx. The optimized result then needs to be averaged over the channel estimates to obtain the final ergodic WSR. From the law of total expectation, we formulate the BF design with a sum power constaint at each BS (P_c) as follows,

$$
\begin{aligned}
EWSR &= \mathrm{E}_{\widehat{\mathbf{h}}} \max_{\mathbf{g}} EWSR(\mathbf{g}), \text{ with } \sum_{i=1,b_i=c}^{K_i} \left\|\mathbf{g}_i\right\|^2 \leq P_c, \text{ where} \\
EWSR(\mathbf{g}) &= \mathrm{E}_{\mathbf{h}|\widehat{\mathbf{h}}} WSR(g) \\
&= \sum_{k=1}^{K} u_k \, \mathrm{E}_{\mathbf{h}|\widehat{\mathbf{h}}} \ln\left(s_k / s_{\overline{k}}\right) \\
&= \mathrm{E}_{\mathbf{h}|\widehat{\mathbf{h}}} \sum_{k=1}^{K} u_k \ln\left(1 + \frac{|\mathbf{h}_k^H \mathbf{g}_k|^2}{s_{\overline{k}}}\right) \\
&\overset{(a)}{\approx} \mathrm{E}_{\mathbf{h}|\widehat{\mathbf{h}}} \sum_{k=1}^{K} u_k \ln\left(1 + \frac{|\mathbf{h}_k^H \mathbf{g}_k|^2}{\mathrm{E}_{\mathbf{h}}\, s_{\overline{k}}}\right) \\
&\overset{(b)}{\leq} \sum_{k=1}^{K} u_k \ln\left(1 + \frac{\mathrm{E}_{\mathbf{h}|\widehat{\mathbf{h}}} |\mathbf{h}_k^H \mathbf{g}_k|^2}{\mathrm{E}_{\mathbf{h}|\widehat{\mathbf{h}}}\, s_{\overline{k}}}\right) \\
&= \sum_{k=1}^{K} u_k \ln(r_{\overline{k}}^{-1} r_k) \\
&= ESIP - WSR(\mathbf{g})
\end{aligned}
\tag{7.10}
$$

where u_k are the rate weights, $\mathbf{g}$ represents the collection of BFs $\mathbf{g}_k$. Transition (a) is due to the MaMISO limit ($K \to \infty$) and (b) is due to the concavity of ln(.) and Jensen's inequality. This leads to the ESIP-WSR upper bound. $s_{\overline{k}}$ is the (channel dependent) interference plus noise power and s_k is the total received power, with conditional expectations $r_{\overline{k}}$, r_k:

$$
\begin{aligned}
s_{\overline{k}} &= 1 + \sum_{i \neq k} |\mathbf{h}_{k,b_i}^H \mathbf{g}_i|^2, \\
s_k &= s_{\overline{k}} + |\mathbf{h}_k^H \mathbf{g}_k|^2, \\
r_{\overline{k}} &= \mathrm{E}_{\mathbf{h}|\widehat{\mathbf{h}}} s_{\overline{k}} = 1 + \sum_{i \neq k} \mathbf{g}_i^H \mathbf{S}_{k,b_i} \mathbf{g}_i, \\
r_k &= \mathrm{E}_{\mathbf{h}|\widehat{\mathbf{h}}} s_k = r_{\overline{k}} + \mathbf{g}_k^H \mathbf{S}_k \mathbf{g}_k, \\
\mathbf{S}_k &= \mathbf{C}_k \mathbf{W}_k \mathbf{C}_k.
\end{aligned}
\tag{7.11}
$$

The ESIP-WSR upper bound can be somewhat loose (gap is maximal at high SNR, see [126]), because inspite of $\mathbf{h}_k$ being MISO, $\mathbf{g}_k^H \mathbf{h}_k$ is only a simple complex Gaussian scalar. Nevertheless, this gap is upper bounded by the Euler constant $\gamma = 0.58$, regardless of SNR. And for the case of only coCSIT ($\widehat{\mathbf{h}} = 0$), the gap is exactly γ, which means that it has no influence on the optimal $\mathbf{g}_k$. By adding the Lagrange terms for the BS power constraints, $\sum_{c=1}^{C} \mu_c (P_c - \sum_{k:b_k=c} ||\mathbf{g}_k||^2)$, to the EWSR in (4.15), we get the gradient (with $\alpha_k = \frac{u_k}{r_k}$, $\beta_k = u_k(\frac{1}{r_{\overline{k}}} - \frac{1}{r_k})$)

$$
\frac{\partial EWSR}{\partial \mathbf{g}_k^*} = \alpha_k \mathbf{S}_k \mathbf{g}_k - (\sum_{i \neq k} \beta_i \mathbf{S}_{i,b_k} + \mu_{b_k} \mathbf{I}) \mathbf{g}_k = 0,
\tag{7.12}
$$

with $\mathbf{S}_{i,b_k} = \mathbf{C}_{i,b_k}\mathbf{W}_{i,b_k}\mathbf{C}_{i,b_k}^H$. This leads to the generalized eigen vector,

$$\mathbf{g}_k' = \mathbf{V}_{max}(\mathbf{S}_k, \sum_{i\neq k} \beta_i \mathbf{S}_{i,b_k} + \mu_{b_k}\mathbf{I}) \,. \tag{7.13}$$

While (7.12) can be interpreted many ways, (7.13) comes from the following DC programming. Introducing the Tx covariance matrices $\mathbf{Q}_i = \mathbf{g}_i\mathbf{g}_i^H$, the power constraints can be written as $\sum_{k:b_k=c} tr\{\mathbf{Q}_k\} \leq P_c$. The EWSR problem is non-concave in the $\mathbf{Q}_k$ due to the interference terms. Therefore finding the global optimum is challenging. In order to find at least a local optimum, we consider the difference of convex functions programming (DCP) approach as in [30]. Whereas [30] however solves the Lagrange multipliers by Lagrangean duality, here we solve them together with the powers (as in standard water filling) in an alternating optimization approach (alternating with optimizing the $\mathbf{g}_k'$). In DCP one keeps the concave signal term and linearizes the convex term, leading to a concave cost function in the $\mathbf{Q}_i$ (or a minorizer actually in the $\mathbf{g}_i'$ and p_i), which can be optimized iteratively.

$$\begin{aligned} EWSR &= u_k \ln\det(r_{\bar{k}}^{-1} r_k) + EWSR_{\bar{k}}, \\ EWSR_{\bar{k}} &= \sum_{i=1,\neq k}^{K} u_i \ln(r_{\bar{i}}^{-1} r_i), \end{aligned} \tag{7.14}$$

where $\ln(r_{\bar{k}}^{-1} r_k)$ is concave in $\mathbf{Q}_k$ and $WSR_{\bar{k}}$ is convex in $\mathbf{Q}_k$. Since a linear function is simultaneously convex and concave, consider the first order Taylor series expansion of $WSR_{\bar{k}}$ in $\mathbf{Q}_k$ around $\widehat{\mathbf{Q}}$ (i.e. all $\widehat{\mathbf{Q}}_i$).

$$EWSR_{\bar{k}}(\mathbf{Q}_k, \widehat{\mathbf{Q}}) \approx EWSR_{\bar{k}}(\widehat{\mathbf{Q}}_k, \widehat{\mathbf{Q}}) - tr\left\{(\mathbf{Q}_k - \widehat{\mathbf{Q}}_k)\widehat{\mathbf{A}}_k\right\}, \tag{7.15}$$

$$\begin{aligned} \text{where, } \widehat{\mathbf{A}}_k &= -\left.\frac{\partial EWSR_{\bar{k}}(\mathbf{Q}_k, \widehat{\mathbf{Q}})}{\partial \mathbf{Q}_k}\right|_{\widehat{\mathbf{Q}}_k, \widehat{\mathbf{Q}}} \\ &= \sum_{i=1,\neq k}^{K} \widehat{\beta}_i \mathbf{S}_{i,b_k} \,. \end{aligned} \tag{7.16}$$

Note that the linearized tangent expression for $EWSR_{\bar{k}}$ constitutes a lower bound for it and hence the DC approach is also a minorization approach. Now dropping the constant terms and reparameterizing the $\mathbf{Q}_k$ in terms of the $\mathbf{g}_k$, we can write the original WSR as the Lagrangian,

$$\begin{aligned} EWSR(\mathbf{g}) = \sum_{k=1}^{K} \Big[& u_k \ln\left(1 + \mathbf{g}_k^H \widehat{\mathbf{B}}_k \mathbf{g}_k\right) \\ & - \mathrm{tr}\left\{\mathbf{g}_k^H \left(\widehat{\mathbf{A}}_k + \mu_{b_k}\mathbf{I}\right)\mathbf{g}_k\right\}\Big] + \sum_{j=1}^{C} \mu_j P_j, \\ \widehat{\mathbf{B}}_k = \widehat{r}_{\bar{k}}^{-1} \mathbf{S}_k. & \end{aligned} \tag{7.17}$$

(7.17) leads again to (7.12) and esp. (7.13). The advantage of formulation (7.17) is that it allows straightforward power adaptation: substituting $\mathbf{g}_k = \sqrt{p_k}\,\mathbf{g}_k'$ in (7.17) and optimizing leads to the following interference leakage ($\sigma_k^{(2)}$) aware water filling

$$p_k = \left(\frac{u_k}{\sigma_k^{(2)} + \mu_{b_k}} - \frac{1}{\sigma_k^{(1)}}\right)^+, \tag{7.18}$$

where $(x)^+ = \max\{0, x\}$ and the Lagrange multipliers μ_c are adjusted (e.g. by bisection) to satisfy the power constraints. Also, $\sigma_k^{(1)} = \mathbf{g}_k'^H \widehat{\mathbf{B}}_k \mathbf{g}_k'$, $\sigma_k^{(2)} = \mathbf{g}_k'^H \widehat{\mathbf{A}}_k \mathbf{g}_k'$. With $\sigma_k^{(2)} = 0$ this would be standard waterfilling. (7.13) comes from the difference of convex functions programming (DCP) [30] and $\beta_k = u_k \left(\frac{1}{r_{\bar{k}}} - \frac{1}{r_k} \right)$. Substituting $\mathbf{g}_k = \sqrt{p_k}\mathbf{g}_k'$ and optimizing using DCP leads to the following *interference leakage* ($\sigma_k^{(2)}$) *aware water filling* (ILA-WF)

$$p_k = \left(\frac{u_k}{\sigma_k^{(2)} + \mu_{b_k}} - \frac{1}{\sigma_k^{(1)}} \right)^+, \tag{7.19}$$

where $(x)^+ = \max\{0, x\}$, $\sigma_k^{(1)} = \mathbf{g}_k'^H \widehat{\mathbf{B}}_k \mathbf{g}_k'$, $\sigma_k^{(2)} = \mathbf{g}_k'^H \widehat{\mathbf{A}}_k \mathbf{g}_k'$, and the Lagrange multipliers μ_c are adjusted (example by bisection) to satisfy the power constraints. Note that intuitively the BF expression (7.13) and the power updates (7.19) represents a optimal compromise between minimizing the leakage power part (leakage channels represent the channels to which $\mathbf{g}_k$ causes interference) and the desired signal power part. Also, this BF design can easily be extended to a MIMO case with multiple streams per-user and ILA-WF power optimization step implicitly selects the required number of streams to be transmitted per-user. We present below in Algorithm 12 the alternating optimization algorithm for BF design using ESIP-WSR. Note that the steps outlined are applicable to EWSMSE/naive BFs, except that the computation of the auxiliary variables $\mathbf{S}_{k,c}$ differ as explained in Section 7.1.5. In the following sections, we represent $\widetilde{\sigma}_k^2$ to be the channel estimation error for all the channels to user k ($\mathbf{h}_{k,c}, \forall c$). Next, we consider the case of CoCSIT (Covariance only CSIT), which corresponds to the case of channel estimate part being absent, so $\mathbf{S} = \mathrm{E}_{\mathbf{h}|\mathbf{C}}\mathbf{C}\mathbf{D}\mathbf{C}^H$. The corresponding BF optimization problem in (7.10) becomes

$$\max_{\mathbf{g}} \sum_{k=1}^{K} u_k \ln\left(1 + \frac{\mathrm{E}_{\mathbf{h}|\mathbf{C}} |\mathbf{h}_k^H \mathbf{g}_k|^2}{\mathrm{E}_{\mathbf{h}|\mathbf{C}}\, s_{\bar{k}}}\right). \tag{7.20}$$

Directly optimizing (7.20) leads to the BF expression as,

$$\begin{aligned} \mathbf{g}_k'' &= \mathbf{V}_{max}(\widehat{\mathbf{B}}_k, \widehat{\mathbf{A}}_k + \mu_{b_k}\mathbf{I}), \\ \text{where,} \quad \widehat{\mathbf{A}}_k &= \sum_{i=1,\neq k}^{K} \beta_i \mathbf{C}_{i,b_k} \mathbf{D}_{i,b_k} \mathbf{C}_{i,b_k}^H, \\ \widehat{\mathbf{B}}_k &= r_{\bar{k}}^{-1} \mathbf{C}_k \mathbf{D}_k \mathbf{C}_k^H \end{aligned} \tag{7.21}$$

For the perfect CSIT case, we optimize the WSR w.r.t $\mathbf{g}$, which leads to

$$\begin{aligned} \mathbf{g}_k' &= \mathbf{V}_{max}(\widehat{\mathbf{B}}_k, \widehat{\mathbf{A}}_k + \mu_{b_k}\mathbf{I}), \\ \text{where,} \quad \widehat{\mathbf{A}}_k &= \sum_{i=1,\neq k}^{K} \beta_i \mathbf{h}_{i,b_k} \mathbf{h}_{i,b_k}^H, \widehat{\mathbf{B}}_k \\ &= s_{\bar{k}}^{-1} \mathbf{h}_k \mathbf{h}_k^H. \end{aligned} \tag{7.22}$$

For these two special cases, the power optimization expression remains the same as the ILA-WF (7.19) with $\widehat{\mathbf{B}}_k, \widehat{\mathbf{A}}_k$ replaced by the expressions in (7.21), (7.22).

7.1.5 Further Considerations on EWSR Bounds

Also, note that the BF expression above (7.34) is quite generic and holds for all the cases 0)-4) described below, with different $\mathbf{A}_k$ for each case.

Algorithm 12: BF Design via ESIP-WSR Optimization

Given: $P_c, \widetilde{\sigma}_k^2, \widehat{\mathbf{h}}_{k,c,LS}, \mathbf{C}_{k,c}, \mathbf{D}_{k,c}, u_k, \forall k, c$.
Initialization:
The $\mathbf{g}_k^{(0)}$ are taken as the ZF precoders with uniform powers. Compute the channel estimates (LS/LMMSE/Subspace projected) $\widehat{\mathbf{h}}_{k,c}, \forall k, c$ as in Section 7.1.2. Compute $\mathbf{S}_{k,c}$ from Section 7.1.3.
Iteration (j) :

1. Compute $\widehat{\mathbf{B}}_k, \widehat{\mathbf{A}}_k$ from (7.13).
2. Update $\mathbf{g}_k''^{(j)}$, $\forall k$, from (7.13).
3. Update $\mu_c^{(j)}$ $\forall c$ using bisection and update the user powers p_k from (7.19).
4. Check for convergence of the WSR: if not go to step 1).

0) **Perfect CSIT:** This corresponds to the case when we replace $\widetilde{\mathbf{h}}_k = 0, \forall k$ in the optimizing function $EWSR(\mathbf{g})$. For the perfect CSIT case, we can obtain

$$\mathbf{A}_k = \sum_{i \neq k} \beta_i \mathbf{h}_{i,b_k} \mathbf{h}_{i,b_k}^H. \tag{7.23}$$

1) **Naive BF EWSR:** In the optimizing function $EWSR(\mathbf{g})$, just replace $\mathbf{h}$ by $\widehat{\mathbf{h}}$ in a perfect CSIT approach, i.e., ignore $\widetilde{\mathbf{h}}$ everywhere. For naive BF

$$\mathbf{A}_k = \sum_{i \neq k} \beta_i \widehat{\mathbf{h}}_{i,b_k} \widehat{\mathbf{h}}_{i,b_k}^H. \tag{7.24}$$

2) **EWSMSE BF [29]**: It accounts for covariance CSIT in the interference terms, but also associates the signal $\widetilde{\mathbf{h}}$ term with the interference. EWSMSE, also called the "use and forget lower bound" in [127], can indeed be shown to be a lower bound for EWSR. For EWSMSE, we just have to replace

$$\mathbf{A}_k = \sum_{i \neq k} \beta_i \mathbf{S}_{i,b_k} - \beta_k \mathbf{C}_k \widetilde{\mathbf{D}}_k \mathbf{C}_k^H. \tag{7.25}$$

For EWSMSE criterion, it is suboptimal in that it exploits the rate-MSE (mean squared error) relation to transform weighted sum rate (WSR) in weighted sum MSE (WSMSE) but the order of expectation and optimization over weights is reversed, to simplify the cost function.
3) **EWSR upper bound ESIP-WSR:** It also accounts for covariance CSIT in the interference term but, unlike EWSMSE, associates the signal $\widetilde{\mathbf{h}}$ term with the signal power.
4) **Covariance CSIT (CoCSIT):** CoCSIT represents the case when only the channel covariance information (of all the users in the system) is known at the BS, i.e the knowledge of $\mathbf{C}$ and $\mathbf{D}$. For the CoCSIT case, we obtain

$$\mathbf{A}_k = \sum_{i \neq k} \beta_i \mathbf{C}_{i,b_k} \mathbf{D}_{i,b_k} \mathbf{C}_{i,b_k}^H, \ \mathbf{v}_k = \mathbf{e}_{i,max} \tag{7.26}$$

In fact, (7.23) tells us that BF is along only one of the unitary vectors (or dominant eigenvector) in the covariance subspace $\mathbf{C}_k$, which leads to a reduction signal power which accounts for the rate offset for the CoCSIT case, see Section 7.1.10.4.

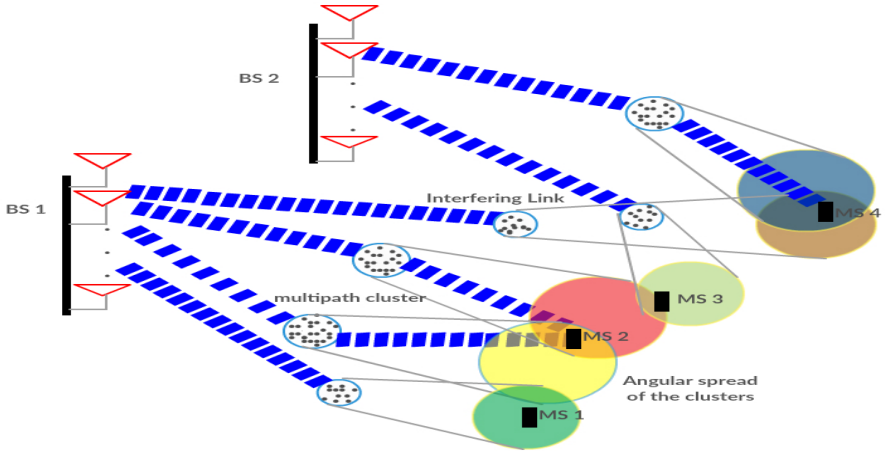


Figure 7.1: COST2100 MIMO Channel Model.

7.1.6 Asymptotic Analysis: Stochastic Geometry MaMISO Regime

In this section, we analyze the asymptotic SE behaviour of MaMISO systems using ESIP-WSR BFs solved using ergodic capacity formulation under partial CSIT. Our analysis is based on the following technical assumptions.

Assumption 5. $\forall k, c, \lim_{M\to\infty} \inf \frac{tr\{\mathbf{D}_{k,c}\}}{M_c} > 0,$

Assumption 6. $0 < \frac{\sum_{k=1}^{K_c} L_{k,c}}{M_c} \overset{\Delta}{=} \alpha_c \leq 1, \forall c.$

Assumption 7. *The long term channel energy captured by* $tr\{\mathbf{D}_{k,c}\}$ *(representative of the large scaling fading factors such as path loss and shadowing) is a constant for a fixed number of BS antennnas* M_c*, i.e.* $tr\{\mathbf{D}_{k,c}\} = \eta_{k,c}(M_{k,c}), \forall k, c$*. Also, we assume that* $\lim_{M\to\infty} \frac{tr\{\mathbf{D}_{k,c}\}}{M_c} \overset{\Delta}{=} \delta_{k,c} < \infty.$

The first assumption (uniform boundedness on the spatial covariance matrix) is essential for the computation of deterministic equivalents using large system analysis results from [14]. This assumption also means that the antenna array gathers an amount of energy which is proportional to the number of antennas and moreover they come from different spatial directions. In the following sections, we use an abuse of notation for convenience, when we refer to $\xrightarrow[a.s]{M\to\infty}$, we refer to the almost sure convergence in the large system limit where $M, L, K \to \infty$ at a fixed ratio (Assumption 6).

The channel model (7.2) results from multipath propagation and the use of BS side antenna arrays. An example of a geometry based stochastic channel model is provided by the COST2100 channel model [128], and can be depicted as in Figure 7.1. This model represents the propagation channel between a BS and user through multipath components (MPCs) arriving at the user terminals and are resulting from the interaction of the transmitted waveform with a set of objects (also called as scatterers). More recent studies with valid measurement data to justify the significance of the COST2100 channel model can be found in [129, 130]. One particular application of this model for the user covariance matrices is considered in [131]. Wherein, the authors consider the scenario in which the support of the multipath angle of arrival or departure (AoA/AoD) for any desired user does not overlap with that of the interfering users. The authors show that the multipath components with AoA/AoD outside the angular support of the desired user tend

to fall in the null space of its covariance matrix in the large antenna limit, leading to orthogonal subspaces $\mathbf{C}$ in the MaMISO regime. Here we add a stochastic geometry regime, in which the random positions of users and scatterers lead to antenna array responses at random angles. In reality, the antenna array responses will be more complex than the Vandermonde vectors for Uniform Linear Arrays considered in [131] due to mutual antenna coupling and various other effects. As a result of this randomness of angles and antenna array responses, and due to limited angular support, the multipath channels live in subspaces that are of limited dimension and uniformly randomly oriented in array response space. As a result, an appropriate random model for the semi-unitary matrices $\mathbf{C}$ spanning these subspaces is a Haar distribution. We shall consider that as the number of antennas M grows unboundedly, the subspace dimensions L also go to infinity (leading to hardening of the signal power), but slower than M. As a result, for the large system analysis we may equivalently consider the elements of $\mathbf{C}$ as i.i.d. with zero mean and variance $1/M$ so that asymptotically such a $\mathbf{C}$ is still semi-unitary: $\mathbf{C}^H\mathbf{C} \overset{M\to\infty}{\longrightarrow} \mathbf{I}_L$. The subspaces $\mathbf{C}$ of different channels will be considered independent.

Before we proceed further, in this section, we recall some of the large system results from [14] we use.

Theorem 8 ([14, Theorem 1]). *Let $\mathbf{Q}_N \in \mathscr{C}^{N\times N}$ be a deterministic matrix and $\mathbf{A}_N = \mathbf{X}_N\mathbf{X}_N^H + \mathbf{S}_N$, with $\mathbf{X}_N$ contains n independent columns with covariance matrix $\mathbf{\Theta}_i$ for i^{th} column and $\mathbf{S}_N \in \mathscr{C}^{N\times N}$ is a Hermitian non-negative definite matrix. Also, assume that $\mathbf{Q}_N, \mathbf{\Theta}_i$ have uniformly bounded spectral norms. Then, for any $z > 0$,*

$$\begin{aligned} &\frac{1}{N} tr\{\mathbf{Q}_N (\mathbf{A}_N - z\mathbf{I}_N)^{-1}\} - \frac{1}{N} tr\{\mathbf{Q}_N \mathbf{T}(z)\} \xrightarrow[a.s]{M\to\infty} 0, \\ &\text{with, } \mathbf{T}(z) = \left(\frac{1}{N}\sum_{i=1}^{n} \frac{\mathbf{\Theta}_i}{1+\delta_i(z)} - z\mathbf{I}_N\right)^{-1}, \text{ where,} \\ &\delta_i(z) = \delta_i^{(\infty)}(z) \text{ is defined as the unique positive solution of} \\ &\delta_i^{(t)}(z) = \frac{1}{N} tr\{\mathbf{\Theta}_i \left(\frac{1}{N}\sum_{i=1}^{n} \frac{\mathbf{\Theta}_i}{1+\delta_i^{(t-1)}(z)} - z\mathbf{I}_N\right)^{-1}\}. \end{aligned} \tag{7.27}$$

We briefly summarize the Lemma's here.

Lemma 3 ([14, Lemma 4, Appendix VI]). *$\mathbf{x}_N^H\mathbf{A}_N\mathbf{x}_N - \frac{1}{N} tr\{\mathbf{A}_N\} \xrightarrow{N\to\infty} 0$ when the elements of $\mathbf{x}_N$ are i.i.d with zero mean and variance $1/N$ and independent of $\mathbf{A}_N$, and similarly when $\mathbf{y}_N$ is independent of $\mathbf{x}_N$, that $\mathbf{x}_N^H\mathbf{A}_N\mathbf{y}_N \xrightarrow[a.s]{N\to\infty} 0$.*

Lemma 4 ([14, Lemma 6, Appendix VI]). *Let $\mathbf{A}_N$ be a deterministic matrix with uniformly bounded spectral norm and $\mathbf{B}_1, ..., \mathbf{B}_N$ be random Hermitian matrices with $\mathbf{B}_N \in \mathscr{C}^{N\times N}$ and eigenvalues $\lambda_1 \leq ....\leq \lambda_N$. Then rank 1 perturbation lemma states that for $\mathbf{v} \in \mathbf{C}^N$,*

$$\frac{1}{N} tr\{\mathbf{A}_N\mathbf{B}_N^{-1}\} - \frac{1}{N} tr\{\mathbf{A}_N \left(\mathbf{B}_N + \mathbf{v}\mathbf{v}^H\right)^{-1}\} \xrightarrow[a.s]{N\to\infty} 0, \tag{7.28}$$

where we assume that $\mathbf{B}_N^{-1}$ and $\left(\mathbf{B}_N + \mathbf{v}\mathbf{v}^H\right)^{-1}$ exist with probability 1.

Lemma 5 ([14, Lemma 1, Appendix VI]). *We also use the matrix inversion lemma (MIL) throughout the paper. Let $\mathbf{A}, \mathbf{C}$ are invertible matrices of size $N \times N$ and $K \times K$, with $\mathbf{B}$ being of size $N \times K$, then MIL states that,*

$$(\mathbf{A} + \mathbf{BCD})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}\left(\mathbf{C}^{-1} + \mathbf{DA}^{-1}\mathbf{B}\right)\mathbf{DA}^{-1}. \tag{7.29}$$

Theorem 9. *In Theorem 8, let* $\mathbf{Q}_k = \mathbf{C}_k \mathbf{D}_k \mathbf{C}_k^H \in \mathscr{C}^{M_{b_k} \times M_{b_k}}$ *be a Hermitian deterministic matrix and* $\mathbf{\Gamma}_k = \sum_{i=1}^{K} \mathbf{C}_{i,b_k} \mathbf{V}_{i,b_k} \mathbf{\Lambda}_{i,b_k} \mathbf{V}_{i,b_k}^H \mathbf{C}_{i,b_k}^H$, *with* $\mathbf{C}_{i,b_k}\mathbf{V}_{i,b_k}$ *contains* L_{i,b_k} *independent columns with covariance matrix* $\mathbf{\Theta}_{i,b_k} = \frac{1}{M}\mathbf{I}_M$ *for* r^{th} *column.* $\mathbf{\Lambda}_{i,b_k}$ *is a diagonal matrix, with* r^{th} *diagonal element being* $\lambda_{i,b_k}^{(r)}$. *Then, for any* $z > 0$

$$\begin{aligned} &\tfrac{1}{M_{b_k}} tr\{\mathbf{Q}_k (\mathbf{\Gamma}_k + z\mathbf{I}_M)^{-1}\} - \tfrac{1}{M_{b_k}} tr\{\mathbf{D}_k\} e_c \xrightarrow[a.s]{M\to\infty} 0, \\ &\textit{with, } b_k = c, \textit{ } e_c, \textit{ is defined as the unique positive solution of} \\ &e_c = \left(\tfrac{1}{M_{b_k}} \sum_{i=1}^{K} \sum_{r=1}^{L_{i,c}} \frac{\beta_i \lambda_{i,c}^{(r)}}{1+\beta_i \lambda_{i,c}^{(r)} e_c} + z \right)^{-1}. \end{aligned} \tag{7.30}$$

From (7.13), we can see that $\mathbf{g}_k'$ will be of the form $\left[\sum_{i\neq k} \beta_i \mathbf{S}_{i,b_k} + \mu_{b_k}\mathbf{I}\right]^{-1} \mathbf{C}_k \mathbf{b}_k$, where $\mathbf{b}_k \propto \mathbf{W}_k \mathbf{C}_k^H \mathbf{g}_k'$ is of size $L_{k,b_k} \times 1$. $\mathbf{b}_k$ can be seen to satisfy,

$$\begin{aligned} &\mathbf{b}_k \propto \mathbf{W}_k \mathbf{C}_k^H \mathbf{\Gamma}_k^{-1} \mathbf{C}_k \mathbf{b}_k, \\ &\mathbf{\Gamma}_k = \sum_{i\neq k} \beta_i \mathbf{S}_{i,b_k} + \mu_{b_k}\mathbf{I}. \end{aligned} \tag{7.31}$$

Hence, $\mathbf{b}_k$ is the the eigenvector corresponding to the maximum eigenvalue, or max eigenvector for short, of $\mathbf{W}_k \mathbf{C}_k^H \left(\sum_{i\neq k} \beta_i \mathbf{S}_{i,b_k} + \mu_{b_k}\mathbf{I}\right)^{-1} \mathbf{C}_k$. Asymptotically $\mathbf{C}_k^H \left(\sum_{i\neq k} \beta_i \mathbf{S}_{i,b_k} + \mu_{b_k}\mathbf{I}\right)^{-1} \mathbf{C}_k$ converges to a deterministic limit which is a multiple of identity, $e_{b_k}\mathbf{I}$, where e_{b_k} is obtained as follows, by applying Theorem 1 and other Lemmas described above. First, we consider the eigen decomposition of $\mathbf{W}_{k,b_i} = \mathbf{V}_{k,b_i} \mathbf{\Lambda}_{k,b_i} \mathbf{V}_{k,b_i}^H$, where $\mathbf{\Lambda}_{k,b_i} = \text{diag}(\lambda_{k,b_i}^{(1)}, ..., \lambda_{k,b_i}^{(L_{k,b_i})})$ and let $\mathbf{C}_{k,b_i}\mathbf{V}_{k,b_i} = \mathbf{C}_{k,b_i}'$. We remark that $\mathbf{C}_{k,b_i}'$ as the product of a Haar matrix and a unitary matrix remains Haar, so Theorem 1 remains applicable. Since the columns of $\mathbf{C}_k$ are independent, we use Lemma 3 to obtain $\mathbf{C}_k^H \mathbf{\Gamma}_k^{-1} \mathbf{C}_k \xrightarrow[a.s]{M\to\infty} \frac{1}{M_{b_k}} \text{tr}\{\mathbf{\Gamma}_k^{-1}\} \mathbf{I}_{L_{k,b_k}}$ (non-diagonal elements converge to zero). Further, we use Lemma 4 to approximate terms of the form $\mathbf{\Gamma}_k^{-1} \approx \left[\mathbf{\Gamma}_k + \beta_k \mathbf{C}_k \mathbf{W}_k \mathbf{C}_k^H\right]^{-1}$. Further using Theorem 8, we obtain a deterministic equivalent as $\frac{1}{M_c}\text{tr}\{\mathbf{\Gamma}_k^{-1}\} - \frac{1}{M_c}\text{tr}\{\mathbf{T}_c(z)\} \xrightarrow[a.s]{M\to\infty} 0$ (with $b_k = c, z = \mu_c$),

$$\begin{aligned} \mathbf{T}_c(z) &= \left(\frac{1}{M_c} \sum_{i=1}^{K} \sum_{r=1}^{L_{i,c}} \frac{\beta_i \lambda_{i,c}^{(r)}}{1+\delta_{ir,c}(z)} + z \right)^{-1} \mathbf{I}_{M_c}, \\ \delta_{ir,c}(z) &= \beta_i \lambda_{i,c}^{(r)} \left(\frac{1}{M_c} \sum_{i=1}^{K} \sum_{r=1}^{L_{i,c}} \frac{\beta_i \lambda_{i,c}^{(r)}}{1+\delta_{ir,c}(z)} + z \right)^{-1}, \\ \text{Define,} \quad e_c(z) &= \left(\frac{1}{M_c} \sum_{i=1}^{K} \sum_{r=1}^{L_{i,c}} \frac{\beta_i \lambda_{i,c}^{(r)}}{1+\delta_{ir,c}(z)} + z \right)^{-1}, \Longrightarrow \\ \delta_{ir,c}(z) &= \beta_i \lambda_{i,c}^{(r)} e_c(z). \end{aligned} \tag{7.32}$$

Note that in (7.32) above, $\mathbf{T}_c(z) = e_c(z)\mathbf{I}$, hence we obtain $\frac{1}{M_{b_k}}\text{tr}\{\mathbf{\Gamma}_k^{-1}\} = e_c(z)$, which is obtained as the unique positive solution of the last expression above, i.e.

$$e_c = \left(\frac{1}{M_c} \sum_{i=1}^{K} \sum_{r=1}^{L_{i,c}} \frac{\beta_i \lambda_{i,c}^{(r)}}{1+\beta_i \lambda_{i,c}^{(r)} e_c} + \mu_c \right)^{-1}. \tag{7.33}$$

Computation of analytical solution of e_c is not feasible except in a simplified case of multiple of identity eigenvalue matrix $\mathbf{D}$ for all users, as we do in Appendix I. However, we remark that if we consider randomized position of users across the multi-cell system, we can assume a probability distribution for the user location information (possibly poisson distributed as in classical stochastic geometry [110]). Then this randomized location induces different attenuation (represented by the eigenvalues in $\mathbf{D}$) for the UE channels due to different path loss between them and the BS and can be modeled as having a spatial distribution. In this case, in the MaMISO limit, we can replace the summation over users by its expectation, i.e. $\frac{1}{M_c}\sum_{i=1}^{K}\sum_{r=1}^{L_{i,c}}\frac{\beta_i\lambda_{i,c}^{(r)}}{1+\beta_i\lambda_{i,c}^{(r)}e_c} \xrightarrow[a.s]{M\to\infty} \frac{\sum_i L_{i,c}}{M_c}\mathrm{E}\left(\frac{\beta_i\lambda_{i,c}^{(r)}}{1+\beta_i\lambda_{i,c}^{(r)}e_c}\right)$, where the expectation is over the probability distribution of the attenuation factor. We remark that it would be of sufficient interest to analyze the SE behaviour with a random attenuation factor and it is left for future work. Further, this leads to $\mathbf{b_k} = \mathbf{V}_{max}(\mathbf{W}_k) \overset{\Delta}{=} \mathbf{v}_{k,b_k}$. Finally, we write the optimized BF w.r.t. partial CSIT, in the stochastic geometry MaMISO regime as,

$$\mathbf{g}_k' = \frac{\mathbf{g}_k''}{\|\mathbf{g}_k''\|},\ \mathbf{g}_k'' = [\sum_{i\neq k}\beta_i\mathbf{S}_{i,b_k}+\mu_{b_k}\mathbf{I}]^{-1}\mathbf{C}_k\mathbf{v}_{k,b_k}. \tag{7.34}$$

It can be intuitively interpreted as follows: The term $\mathbf{C}_k\mathbf{v}_{k,b_k}$ represents beamforming (matched filtering) in the covariance subspace $\mathbf{C}_k$ of the channel from BS b_k to user k. The first term $\mathbf{\Gamma}_k$ represents a weighting matrix which converges to the projection matrix on the orthogonal complement of the covariance subspace of the leakage channels at high SNR. At any intermediate SNR, BF choses to approximately ZF to a subset of leakage channels. A more detailed interpretation using the concept of reduced order ZF BF can be found in [99]. At low SNR, $\mathbf{\Gamma}_k$ reduces to $\frac{1}{\mu_{b_k}}\mathbf{I}$. Thus at low SNR, BF reduces to $\mathbf{C}_k\mathbf{v}_{k,b_k}$, which is just the matched filter. Further, we deduce the optimized BF under the special cases such as perfect CSIT and CoCSIT case. For the perfect CSIT case,

$$\begin{aligned}\mathbf{g}_k' &= \frac{\mathbf{g}_k''}{\|\mathbf{g}_k''\|},\\ \mathbf{g}_k'' &= [\sum_{i\neq k}\beta_i\mathbf{h}_{i,b_k}\mathbf{h}_{i,b_k}^H+\mu_{b_k}\mathbf{I}]^{-1}\mathbf{h}_k.\end{aligned} \tag{7.35}$$

For the CoCSIT case, we obtain,

$$\begin{aligned}\mathbf{g}_k' &= \frac{\mathbf{g}_k''}{\|\mathbf{g}_k''\|},\\ \mathbf{g}_k'' &= [\sum_{i\neq k}\beta_i\mathbf{C}_{i,b_k}\mathbf{D}_{i,b_k}\mathbf{C}_{i,b_k}^H+\mu_{b_k}\mathbf{I}]^{-1}\mathbf{C}_k\mathbf{e}_{i,max},\end{aligned} \tag{7.36}$$

where $\mathbf{e}_{i,max}$ represents the unit vector $\mathbf{e}_i$ corresponding to $\mathbf{D}_{i,i}$ which is the maximum among all the eigenvalues. In fact, (7.35) tells us that BF is along only one of the unitary vectors in the covariance subspace $\mathbf{C}_k$, which leads to a reduction signal power which accounts for the rate offset as in described in the Corollary 12.4.

7.1.7 Computation of eigenvalues of $\mathbf{W}_{k,b_i}$

To compute $e_c(z)$ in the large system expressions as in (7.32), the eigenvalues of $\mathbf{W}_{k,b_i}$ need to be known which we discuss here. For the convenience of analysis we omit the user and BS index

here. We represent $\mathbf{W}$ by $\mathbf{W}_L, \mathbf{W}_S$ for LMMSE and subspace channel estimators respectively. From Section III, for the LMMSE,

$$\mathbf{W}_L = \mathbf{U}_L \left(\widehat{\mathbf{d}}\widehat{\mathbf{d}}^H + \widetilde{\sigma}^2\mathbf{I}\right)\mathbf{U}_L^H + (\mathbf{I}-\mathbf{U}_L)\,\mathbf{D}\,(\mathbf{I}-\mathbf{U}_L)^H \tag{7.37}$$

where $\mathbf{U}_L = \left(\mathbf{I}+\widetilde{\sigma}^2\mathbf{D}^{-1}\right)^{-1}$, $\mathbf{U}_L\widehat{\mathbf{d}} = \widehat{\mathbf{c}}_L$ for short and we obtain $\mathbf{W}_L = \left(\mathbf{I}+\widetilde{\sigma}^2\mathbf{D}^{-1}\right)^{-1}\widehat{\mathbf{d}}\widehat{\mathbf{d}}^H\left(\mathbf{I}+\widetilde{\sigma}^2\mathbf{D}^{-1}\right)^{-1} + \mathbf{D}-\mathbf{D}\left(\widetilde{\sigma}^2\mathbf{I}+\mathbf{D}\right)^{-1}\mathbf{D}$. At both high and low SNR, we can replace the error covariance $\mathbf{D}-\mathbf{D}\left(\widetilde{\sigma}^2\mathbf{I}+\mathbf{D}\right)^{-1}\mathbf{D}$ by its dominating term $\mathbf{D}_{1,1}\widetilde{\sigma}^2/\left(\mathbf{D}_{1,1}+\widetilde{\sigma}^2\right)\mathbf{e}_1\mathbf{e}_1^H$, assuming $\mathbf{D}_{1,1}$ is the largest diagonal element of $\mathbf{D}$. Thus $\mathbf{W}_L$ becomes the sum of two rank one matrices and we propose the resulting rank 2 approximation of $\mathbf{W}_L$ for all SNR. It will be all the more precise at intermediate SNR if $\mathbf{D}_{1,1}$ dominates the rest of $\mathbf{D}$. Further we look at the computation of the eigenvalue matrix $\mathbf{\Lambda}$ of $\mathbf{W}_L$. At high SNR, we can approximate $\left(\mathbf{I}+\widetilde{\sigma}^2\mathbf{D}^{-1}\right)^{-1} = \mathbf{I}-\widetilde{\sigma}^2\mathbf{D}^{-1}$. So up to first order in $\widetilde{\sigma}^2$, we obtain $\mathbf{W}_L \approx \widehat{\mathbf{c}}_L\widehat{\mathbf{c}}_L^H + \widetilde{\sigma}^2\mathbf{I}$, where the first term contains first-order terms also. At low SNR, $\mathbf{U}_L \approx \widetilde{\sigma}^{-2}\mathbf{D}$ and we can obtain $\mathbf{W}_L \approx \widehat{\mathbf{c}}_L\widehat{\mathbf{c}}_L^H + \mathbf{D}$. For any SNR, these two extremes can be connected by the following approximation:

$$\begin{aligned}\mathbf{\Lambda} &= \widetilde{\sigma}^2\mathbf{D}\left(\widetilde{\sigma}^2\mathbf{I}+\mathbf{D}\right)^{-1} + ||\widehat{\mathbf{c}}_L||^2\mathbf{e}_1\mathbf{e}_1^H \\ &= \widetilde{\sigma}^2\mathbf{D}\left(\widetilde{\sigma}^2\mathbf{I}+\mathbf{D}\right)^{-1} + \mathrm{tr}\{\mathbf{D}\left(\mathbf{I}+\widetilde{\sigma}^2\mathbf{D}^{-1}\right)^{-1}\}\,\mathbf{e}_1\mathbf{e}_1^H\end{aligned} \tag{7.38}$$

where the last equality is due to the law of large numbers. For subspace channel estimator, $\mathbf{W}_S = \widehat{\mathbf{d}}\widehat{\mathbf{d}}^H + \widetilde{\sigma}^2\mathbf{I}$, the eigenvalue matrix becomes $\mathbf{\Lambda} = \widetilde{\sigma}^2\mathbf{I} + \mathrm{tr}\{\mathbf{D}\}\,\mathbf{e}_1\mathbf{e}_1^H$.

Further we look at the deterministic equivalent of the SINR and rate. Using the BF expression derived under partial CSIT and stochastic geometry regime (7.34), we evaluate the WSR using the actual channel distribution and arrive at the following result.

Theorem 10. *In the large system limit, the quantities $\gamma_k - \overline{\gamma}_k \xrightarrow[a.s]{M\to\infty} 0$, where $\overline{\gamma}_k$ is the deterministic equivalent of the SINR. Similarly, the quantities defined in* (7.11) *also converges to their deterministic equivalents, $r_k - \overline{r}_k \xrightarrow[a.s]{M\to\infty} 0, r_{\overline{k}} - \overline{r}_{\overline{k}} \xrightarrow[a.s]{M\to\infty} 0$. Further we can show that, since the logarithm is a continuous function, by applying the continuous mapping theorem [96], it follows from the almost sure convergence of γ_k that, $R_k - \overline{R}_k \xrightarrow[a.s]{M\to\infty} 0$, where R_k is the rate of user k, with $\overline{R}_k = \ln(1+\overline{\gamma}_k)$. By using similar argument, we state that $\beta_k - \overline{\beta}_k \xrightarrow[a.s]{M\to\infty} 0$. The deterministic limits for the ESIP-WSR BF with LMMSE and subspace channel estimates are obtained as,*

$$\begin{aligned}\overline{\gamma}_{k,L}^{(Opt)} &= \frac{p_k\left(1-x_c^{(L,Opt)}\right)tr\{\mathbf{D}_k\}}{\frac{1}{M_{b_i}}\sum_{i\neq k}p_i\,tr\{\mathbf{D}_{k,b_i}\mathbf{B}_{k,b_i}^{-2}\}+1},\ b_k = c \\ \overline{\beta}_k &= u_k\left(\frac{1}{\overline{r}_{\overline{k}}} - \frac{1}{\overline{r}_k}\right), \\ \textit{where, } x_c^{(L)} &= \frac{e_c^2}{M_c}\sum_{i=1}^{K}\sum_{r=1}^{L_{i,c}}\frac{\beta_i^2\lambda_{i,c}^{(r),2}}{\left(1+\beta_i\lambda_{i,c}^{(r)}e_c\right)^2}, \\ \gamma_{k,S}^{(Opt)} &= \frac{\left(1-x_c^{(S,Opt)}\right)(tr\{\mathbf{D}_k\})^2}{tr\{\left(\mathbf{D}_k+\widetilde{\sigma}_k^2\mathbf{I}\right\}\left(\frac{1}{M}\sum_{i\neq k}^{K}\frac{tr\{\mathbf{B}_{k,b_i}^{-2}\mathbf{D}_{k,b_i}\}}{tr\{\mathbf{D}_i+\widetilde{\sigma}_k^2\mathbf{I}\}}p_i+1\right)}.\end{aligned} \tag{7.39}$$

In the above expression, in the notation $\gamma_{k,L}^{(Opt)}$, subscript L indicates LMMSE channel estimator and superscript (Opt) indicates ESIP-WSR BF. Similarly, in the notation $\gamma_{k,S}^{(Opt)}$, subscript S indi-

cates Subspace channel estimator and superscript (Opt) indicates ESIP-WSR BF. Similarly for the naive BF, we obtain the deterministic equivalent for SINR as,

$$\begin{aligned}\overline{\gamma}_{k,L}^{(N)} &= \frac{\left(1-x_c^{(L,N)}\right) tr\{\mathbf{D}_k^2\left(\mathbf{D}_k+\widetilde{\sigma}_k^2\mathbf{I}\right)^{-1}\}p_k}{\left(\sum_{i=1}^{K}\frac{1}{M}p_i\, tr\{\mathbf{D}_{k,b_i}\}+1\right)}, \\ x_c^{(L,N)} &= \frac{e_c^2}{M_c}\sum_{i=1}^{K}\frac{\beta_i^2\lambda_{i,c}^{(1),2}}{\left(1+\beta_i\lambda_{i,c}^{(1)}e_c\right)^2}.\end{aligned} \tag{7.40}$$

Proof: *The proof is very much involved and is given in Appendix I for various BF and channel estimator combination. In Appendix, we evaluate the actual sum rate using the BFs designed by the optimization of various egodic capacity bounds. For the values of $\overline{r}_k, \overline{r}_{\overline{k}}$ which define the deterministic equivalent of β_k, for the proof, we refer to Theorem 12, where we evaluate the deterministic equivalent of the EWSR.*

In the above SINR expression, the quantities $\left(1-x_c^{(L,Opt)}\right), \left(1-x_c^{(L,N)}\right)$ represents the loss in signal power due to the amount of ZF happening at any SNR, which varies from no ZF at very low SNR to ZF to all the paths ($\sum_{i=1}^{K} L_{k,c}$ of them, case of constant channel estimation error) to which $\mathbf{g}_k$ cause interference or all the user channels (K of them, case of estimation error varying with SNR). The details of this analysis will be dealt in the following sections. Also, note that from (7.39), we can conclude that for the subspace channel estimator the signal power gets reduced by a factor $\frac{\mathbf{tr}\{\mathbf{D}_k\}}{\mathbf{tr}\{\mathbf{D}_k+\widetilde{\sigma}_k^2\mathbf{I}\}}$ compared to the LMMSE. This is attributed to the absence of weighting which is present in the case of LMMSE channel estimator.

7.1.8 EWSMSE BF in the MaMISO Stochastic Geometry Regime

From the definition in Section 7.1.5, by moving desired user channel interference power to the interference power terms, the EWSMSE BF expression can be written as,

$$\begin{aligned}\mathbf{g}_k &\propto \mathbf{F}_k^{-1}\mathbf{C}_k\mathbf{U}_k\widehat{\mathbf{d}}_k\widehat{\mathbf{d}}_k^H\mathbf{U}_k^H\mathbf{C}_k^H\mathbf{g}_k \\ &\overset{(a)}{\propto} \mathbf{F}_k^{-1}\mathbf{C}_k\mathbf{U}_k\widehat{\mathbf{d}}_k, \\ \text{So, } \mathbf{g}_k'' &= \mathbf{F}_k^{-1}\widehat{\mathbf{h}}_k, \text{ where,} \\ \mathbf{F}_k &= \mathbf{\Gamma}_k + \beta_k\mathbf{C}_k\widetilde{\mathbf{d}}_k\mathbf{C}_k^H, \\ \mathbf{\Gamma}_k &= \sum_{i\neq k}\beta_i\mathbf{C}_{i,b_k}\mathbf{W}_{i,b_k}\mathbf{C}_{i,b_k}^H + \mu_{b_k}\mathbf{I}, \\ \widehat{\mathbf{h}}_k &= \mathbf{C}_k\mathbf{U}_k\mathbf{C}_k^H\widehat{\mathbf{h}}_{k,LS} = \mathbf{C}_k\mathbf{U}_k\widehat{\mathbf{d}}_k.\end{aligned} \tag{7.41}$$

(a) follows from $\widehat{\mathbf{d}}_k^H\mathbf{U}_k^H\mathbf{C}_k^H\mathbf{g}_k$ being a scalar. Let us denote $\gamma_{k,L}^{(E)}$ as the SINR for user k under LMMSE channel estimator and EWSMSE BF. Similarly, $\gamma_{k,S}^{(E)}$ indicates the SINR for user k under Subspace channel estimator and EWSMSE BF.

Theorem 11. *In the large system limit, the SINR of an EWSMSE BF with LMMSE and subspace channel estimator converges to a deterministic limit, $\gamma_{k,L}^{(E)} - \overline{\gamma}_{k,L}^{(E)} \xrightarrow[a.s]{M\to\infty} 0, \gamma_{k,S}^{(E)} - \overline{\gamma}_{k,S}^{(E)} \xrightarrow[a.s]{M\to\infty} 0$. Further using the continuous mapping theorem, we can say that the rate of user k converges, $R_k -$*

$$\overline{R}_k \xrightarrow[a.s]{M\to\infty} 0. \tag{7.42}$$

$$\begin{aligned}
\overline{\gamma}_{k,L}^{(E)} &= \gamma_{k,L}^{(E)} - \frac{P_{S_k}}{\Upsilon_{1k}+\Upsilon_{2k}-2\Upsilon_{3k}+1} \xrightarrow[a.s]{M\to\infty} 0, \text{ where,}\\
P_{S_k} &= \frac{\left(1-x_{b_k}^{(L,E)}\right)\left(tr\{\mathbf{U}_k\left(\mathbf{I}-\mathbf{E}_k^{-1}e_{b_k}\right)\mathbf{D}_k\}\right)^2 p_k}{tr\{\mathbf{U}_k^2\left(\mathbf{I}-e_{b_k}\mathbf{E}_k^{-1}\right)^2\left(\mathbf{D}_k+\widetilde{\sigma}_k^2\mathbf{I}\right)\}},\\
\Upsilon_{1k} &= \sum_{i\neq k}\frac{\frac{1}{M_{b_i}}tr\{\mathbf{D}_{k,b_i}\mathbf{B}_{k,b_i}^{-2}\}\,tr\{\mathbf{U}_{i,b_i}^2\left(\mathbf{D}_i+\widetilde{\sigma}_i^2\mathbf{I}\right)\}p_i}{tr\{\mathbf{U}_{i,b_i}^2\left(\mathbf{I}-e_{b_i}\mathbf{E}_i^{-1}\right)^2\left(tr\{\mathbf{D}_i+\widetilde{\sigma}_i^2\mathbf{I}\}\right)\}},\\
\Upsilon_{2k} &= \sum_{i\neq k}\frac{\left(1-x_{b_k}^{(L,E)}\right)\frac{1}{M_{b_i}}tr\{\mathbf{D}_{k,b_i}\mathbf{B}_{k,b_i}^{-2}\}\,tr\{\mathbf{U}_{i,b_i}^2\mathbf{E}_i^{-2}\left(\mathbf{D}_i+\widetilde{\sigma}_i^2\mathbf{I}\right)\}p_i}{tr\{\mathbf{U}_i^2\left(\mathbf{I}-e_{b_i}\mathbf{E}_i^{-1}\right)^2\,tr\{\mathbf{D}_i+\widetilde{\sigma}_i^2\mathbf{I}\}\}},\\
\Upsilon_{3k} &= \sum_{i\neq k}\frac{\left(1-x_{b_k}^{(L,E)}\right)\frac{1}{M_{b_i}}tr\{\mathbf{D}_{k,b_i}\mathbf{B}_{k,b_i}^{-2}\}\,tr\{\mathbf{U}_{i,b_i}^2\mathbf{E}_i^{-1}\left(\mathbf{D}_i+\widetilde{\sigma}_i^2\mathbf{I}\right)\}p_i}{e_{b_i}tr\{\mathbf{U}_i^2\left(\mathbf{I}-e_{b_i}\mathbf{E}_i^{-1}\right)^2\,tr\{\mathbf{D}_i+\widetilde{\sigma}_i^2\mathbf{I}\}\}},\\
\text{with } \mathbf{U}_i &= \left(\mathbf{I}+\widetilde{\sigma}_i^2\mathbf{D}_{i,b_i}^{-1}\right)^{-1} \text{ for LMMSE and}
\end{aligned}$$

$\mathbf{U}_i = \mathbf{I}$ *for Subspace channel estimator.*

Proof: *The proof is given in Appendix IV. We also derive the deterministic equivalents for the ESIP-WSR and naive BFs with LS channel estimate in Appendices V and VI, respectively.*

7.1.9 Deterministic Equivalent of Auxiliary Quantities

In this section, we derive under large system limit, with $M_c, K_c \to \infty$ at a fixed ratio $\frac{K_c}{M_c} < 1, \forall c$, approximations to the scalar quantities involved in the rate expression, which we denote as the deterministic equivalent.

Theorem 12. *In the large system limit, the quantities* $\sigma_k^{(1)} - \overline{\sigma}_k^{(1)} \xrightarrow[a.s]{M_{b_k}\to\infty} 0$, $\sigma_k^{(2)} - \overline{\sigma}_k^{(2)} \xrightarrow[a.s]{M_{b_k}\to\infty} 0$, $r_{\overline{k}} - \overline{r}_{\overline{k}} \xrightarrow[a.s]{M_{b_k}\to\infty} 0$ *and* $r_k - \overline{r}_k \xrightarrow[a.s]{M_{b_k}\to\infty} 0$, *where* $\overline{\sigma}_k^{(1)}, \overline{\sigma}_k^{(2)}, \overline{r}_{\overline{k}}, \overline{r}_k$ *are the deterministic equivalents. Here* $\xrightarrow[a.s]{M_{b_k}\to\infty}$ *denotes almost sure convergence. Further we can show that, since the logarithm is a continuous function, by applying the continuous mapping theorem [96], it follows from the almost sure convergence of* r_k *and* $r_{\overline{k}}$ *that,* $R_k - \overline{R}_k \xrightarrow[a.s]{M_{b_k}\to\infty} 0$, *where* R_k *is the rate of user* k, *with* $\overline{R}_k = \ln(\frac{\overline{r}_k}{\overline{r}_{\overline{k}}})$. *By using similar argument, we state that* $\beta_k - \overline{\beta}_k \xrightarrow[a.s]{M_{b_k}\to\infty} 0$ *and* $\alpha_k - \overline{\alpha}_k \xrightarrow[a.s]{M_{b_k}\to\infty} 0$. *The deterministic limits are obtained as,*

$$\begin{aligned}
\overline{\sigma}_k^{(1)} &= \frac{e_{b_k}^2\lambda_{max}(\mathbf{W}_{k,b_k})}{e'_{b_k}\left(1+\Upsilon_{\overline{k}}\right)},\\
\overline{\sigma}_k^{(2)} &= \frac{1}{M_{b_k}}\sum_{i=1,i\neq k}^{K}\overline{\beta}_i\left[\sum_{r=1}^{L_{i,b_k}}\frac{\zeta_{i,b_k}^{(r)}}{\left(1+\overline{\beta}_i\zeta_{i,b_k}^{(r)}e_{b_k}\right)^2}\right],\\
\overline{r}_{\overline{k}} &= 1+\Upsilon_{\overline{k}},\\
\overline{r}_k &= 1+\Upsilon_{\overline{k}}+p_k\frac{e_{b_k}^2\lambda_{max}(\mathbf{W}_{k,b_k})}{e'_{b_k}},
\end{aligned} \tag{7.43}$$

where,

$$\Upsilon_{\overline{k}} = \sum_{\substack{i=1,\\ i\neq k}}^{K} p_i \frac{1}{M_{b_i}} \left[\sum_{r=1}^{L_{k,b_i}} \frac{\zeta_{k,b_i}^{(r)}}{\left(1+\beta_k \zeta_{k,b_i}^{(r)} e_{b_i}\right)^2} \right], \tag{7.44}$$
$$\overline{\beta}_k = u_k \left(\tfrac{1}{\overline{r}_{\overline{k}}} - \tfrac{1}{\overline{r}_k} \right),\ \overline{\alpha}_k = \tfrac{u_k}{\overline{r}_k},$$

Proof: Main steps leading to this using standard results from random matrix theory [14] are outlined in Appendix I.
All the deterministic equivalents described above depend just on the scalar parameters such as eigenvalues of the channel covariance matrices, transmit powers and channel estimation error variances. Note that the BF computation algorithm based on EWSR still remains iterative in p_i, β_i and μ_{b_i}.

7.1.10 Simplified Sum Rate Expressions with Different BF and Channel Estimators

7.1.10.1 Sum Rate Analysis at any SNR

Even though in general, the true channel eigenvalues may be distinct, it is illustrative to consider an extreme case where the eigenvalues are all equal. In this section, using the results from the Appendix I, we discuss the simplified sum rate expressions for naive, EWSMSE and ESIP-WSR BFs for LMMSE/Subspace/LS channel estimators under multi cell (C cells), with identical parameters, $\widetilde{\sigma}_{k,c}^2 = \widetilde{\sigma}^2$, $L_{k,c} = L$, $\mathbf{D}_{k,c} = \frac{\eta_{k,c}}{L}\mathbf{I}$, $P_c = P, \forall c$ and $M_c = M, \forall k, c$. Number of users in cell c is denoted as $K_c = K/C, \forall c$. For ESIP-WSR BF with LMMSE channel estimate, substituting these values in (7.38), we obtain,

$$\begin{aligned} l\lambda_{k,c}^{(1)} &= \zeta_{k,c} + \lambda_{k,c}^{(2)}, \\ \zeta_{k,c} &= \frac{\eta_{k,c}^2}{L\widetilde{\sigma}^2 + \eta_{k,c}}, \\ \lambda_{k,c}^{(2)} &= \frac{\widetilde{\sigma}^2 \eta_{k,c}}{L\widetilde{\sigma}^2 + \eta_{k,c}}, \end{aligned} \tag{7.45}$$

and rest of the eigenvalues $\lambda_{k,c}^{(r)} = \lambda_{k,c}^{(2)}, \forall r = 2, ..., L$. In the case of naive BF with LMMSE channel estimate, there will only be one eigenvalue and that will be $\zeta_{k,c}$. For the subspace channel estimator, the eigenvalues are $\lambda_{k,c}^{(1)} = \left(\eta_{k,c} + L\widetilde{\sigma}^2\right) + \widetilde{\sigma}^2$, $\lambda_{k,c}^{(r)} = \widetilde{\sigma}^2, \forall r \neq 1$. Similarly for the naive BF with subspace channel estimator, the only one eigenvalue is, $\lambda_{k,c}^{(1)} = \eta_{k,c} + L\widetilde{\sigma}^2$. For ESIP-WSR BF with LS only channel estimate, the only eigenvalue will be $\lambda_{k,c}^{(1)} = \eta_{k,c} + M\widetilde{\sigma}^2$.

Moreover, in order to simplify the analysis, we consider the case of same attenuation for all the channels, $\mathbf{D}_{k,c} = \frac{\eta}{L}\mathbf{I}, \forall k, c$, $\alpha = \frac{KL}{M}$ hence $\beta_i = \beta, \forall i$. In this case, we denote the eigenvalues (which are the same for all the $\mathbf{W}_{k,c}$) are of the form $\lambda_{k,c}^{(1)} = \lambda_\zeta + \lambda_2$ $\lambda_{k,c}^{(r)} = \lambda_2 \, \forall r > 1$, where ζ_1, ζ_2 are defined below. Further we can write the equation for solving e_c from (7.32) as,

$$\begin{aligned} \frac{1}{e_c} &= \frac{K}{M}\frac{\beta\lambda_1}{1+\beta\lambda_1 e_c} + \frac{KL}{M}\frac{\beta\lambda_2}{1+\beta\lambda_2 e_c} + \mu_c, \\ \zeta &= \frac{\eta^2}{L\widetilde{\sigma}^2 + \eta}, \lambda_2 = \frac{\widetilde{\sigma}^2\eta}{L\widetilde{\sigma}^2 + \eta}, \\ \lambda_1 &= \zeta + \lambda_2. \end{aligned} \tag{7.46}$$

We consider below certain special cases for which the implicit equation of e_k can be analytically solved.

Corollary 12.1. *For the naive BF with LMMSE channel estimate (or with LS or Subspace estimator), $\frac{1}{e_c} = \frac{K}{M}\frac{\beta\lambda_1}{1+\beta\lambda_1 e_c} + \mu_c$, after some algebraic manipulations, it can be shown to be the solution of a quadratic equation and the positive e_c can be obtained as,*

$$e_c = \frac{-\left(\mu_c + \beta\lambda_1(\alpha-1)\right) + \sqrt{\left(\mu_c + \beta\lambda_1(\alpha-1)\right)^2 + 4\beta\lambda_1\mu_c}}{2\beta\lambda_1\mu_c}. \tag{7.47}$$

At extreme SNR regions (where $\mu_c \propto 1/P$), it can be deduced that $\lim_{P\to 0} e_c = 0$, $\lim_{P\to\infty} e_c = \infty$. Further by substituting for e_c in (62) *leads to $x_c^{(LS,N)} = x_c^{(L,N)} = x_c^{(S,N)} = \frac{K}{M}$ at high SNR and $x_c^{(LS,N)} = x_c^{(L,N)} = x_c^{(S,N)} = 0$ at low SNR for the naive BFs.*

Corollary 12.2. *For WSR based BF design with perfect CSIT, which represents a special case of ESIP-WSR BF considered in this chapter ($\widetilde{\mathbf{D}} = \mathbf{0}$), the implicit equation for e_c gets simplified as, $\frac{1}{e_c} = \frac{K}{M}\frac{\beta\eta}{1+\beta\eta e_c} + \mu_c$. Note that there is only one eigenvalue corresponding to the true rank one channel vector which is η. Hence we obtain a positive solution by solving the resulting quadratic equation,*

$$e_c = \frac{-\left(\mu_c + \beta\eta\left(\frac{\alpha}{L}-1\right)\right) + \sqrt{\left(\mu_c + \beta\eta\left(\frac{\alpha}{L}-1\right)\right)^2 + 4\beta\eta\mu_c}}{2\beta\eta\mu_c}. \tag{7.48}$$

Again, at extreme SNR regions, it can be deduced that $\lim_{P\to 0} e_c = 0$, $\lim_{P\to\infty} e_c = \infty$. Further substituting these values in (62) *leads to the ZF dimension of K (interfering user channels) and hence the rate expression can be written as $\overline{R} = K\ln\left((1-\frac{K}{M})SNR\eta\frac{CP}{K}\right)$.*

Corollary 12.3. *In the case of CoCSIT, the implicit equation for e_c gets simplified as, $\frac{1}{e_c} = \alpha\frac{\beta\eta}{1+\beta\eta e_c} + \mu_c$ and a positive solution can be obtained as,*

$$e_c = \frac{-\left(\mu_c + \beta\eta(\alpha-1)\right) + \sqrt{\left(\mu_c + \beta\eta(\alpha-1)\right)^2 + 4\beta\eta\mu_c}}{2\beta\eta\mu_c}. \tag{7.49}$$

At extreme SNR regions, it can be shown that $\lim_{P\to 0} e_c = 0$, $\lim_{P\to\infty} e_c = \infty$. Further by substituting for e_c in (62) *leads to $x_c^{(C)} = \frac{KL}{M}$ at high SNR and $x_c^{(C)} = 0$ at low SNR for the naive BFs.*

7.1.10.2 High SNR Analysis ($\widetilde{\sigma}^2 \propto \frac{1}{P}$)

We derive in detail the high SNR analysis simplifications for various BF and channel estimator combination in Appendix K. It is clear from those derivations that the eigenvalues ($\lambda_{k,c}^{(r)}$) will determine the evolution (w.r.t SNR) of the large system approximation values such as e_c or x_c which is defined in (62). The value x_c is also an indication of the ZF dimensions of the various BF design and varies w.r.t the channel estimation quality and the BF optimization along with the particular channel estimate being considered. We define $\rho_{k,c} = \eta_{k,c}p_k$, where $\rho_{k,c}$ is the received SNR at user k from BS c (with $b_k = c$). Substituting these values, we observe that the sum rate expressions at high SNR can be expressed as,

$$\overline{R} = \sum_{k=1}^{K} \ln(1+\omega_k\rho_{k,c}), \tag{7.50}$$

Table 7.1: High SNR Rate Offset for Various BFs ($\mathbf{D}_{k,c} = \frac{\eta_{k,c}}{L}\mathbf{I}$, $\widetilde{\sigma}^2 \propto 1/P$)

ω	naive	EWSMSE	ESIP-WSR
LS	$\frac{\left(1-\frac{K}{M}\right)\frac{\eta_{k,c}}{\eta_{k,c}+\widetilde{\sigma}^2 M}}{\widetilde{\sigma}^2 CP+1}$	$\frac{(1-\frac{K}{M})\frac{\eta_{k,c}}{\eta_{k,c}+\widetilde{\sigma}^2 M}}{\widetilde{\sigma}^2 CP+1}$	$\frac{\left(1-\frac{K}{M}\right)\frac{\eta_{k,c}}{\eta_{k,c}+\widetilde{\sigma}^2 M}}{\widetilde{\sigma}^2 CP+1}$
LMMSE/Subspace	$\left(1-\frac{K}{M}\right)$	$\left(1-\frac{K}{M}\right)$	$\left(1-\frac{K}{M}\right)$

where $\omega = \frac{z}{1+yP}$ represents the rate offset, where z, y varies w.r.t the channel estimator and the type of BF design. For those BF which saturates at high SNR, the saturation level is represented as $\frac{zP}{1+yP} \approx \frac{z}{y}$. The corresponding ω for the 9 different combinations of channel estimator and BFs are depicted in the Table 8.1 below. Also, we assume that the coherence interval T_c is sufficiently large such that the prelog factor $(1-T_c^{-1})$ appearing in the rate expression [122] can be neglected. If we consider the simplified case of identical channel attenuation for all users in the system, $\eta_{k,c} = \eta, \forall k, c$, for which the sum rate simplifies to $\overline{R} = K\ln(1+\omega\rho) \approx K\ln(\omega\rho), \rho = \eta\frac{P}{K_c}$. For the ESIP-WSR BF, the sum rate can be written as, $\overline{R} = K\ln\left((1-\frac{K}{M})SNR\frac{C\eta}{K}\right)$, where we define the Tx $SNR = P$. This in fact tells us how the capacity scales with the system parameters such as M, K, SNR, η and for the ESIP-WSR BF, it can be interpreted that the capacity scales with the rank of the channel under very strong spatial correlation regime. Note that at high SNR, ESIP-WSR BF with subspace channel estimator converges to the performance of the LMMSE estimator, hence we have merged the values of subspace and LMMSE estimator in the tables. Another remark is that the ZF dimension of all the BFs are K. n ESIP-WSR/Naive EWSR BF does project the LS channel estimate to the covariance subspace and the noise part is also reduced to the subspace. However, in the case of LS channel estimate, since there is no projection to the subspace, the estimation error present in the remaining $M-L$ spatial dimensions get multiplied by the Tx power to give a constant rate offset at high SNR. This explains the degraded performance of LS estimates w.r.t LMMSE/Subspace channel estimators.

Corollary 12.4. *From Corollary 12.3 and the deterministic equivalent derived in Appendix VII, we obtain the sum rate at high SNR for the case of CoCSIT as,*

$$\overline{R}_{CoCSIT} = K\ln((1-\tfrac{KL}{M})SNR\tfrac{\eta}{L}\tfrac{C}{K}), \tag{7.51}$$

and this represents a constant rate offset (per-user) of $\ln\frac{(M-K)}{M-KL} + \ln L$ *for the CoCSIT compared to the perfect CSIT at high SNR. This rate offset is attributed due to the lack of perfect channel knowledge. The best CoCSIT can do is to transmit along the dominant eigenvector of the direct channel in the covariance subspace resulting in the factor of* $\ln L$ *reduction in the rate offset. Also, the BF does ZF to the L independent paths of the leakage channels and this results in the signal power reduction of* $\ln\frac{(M-K)}{M-KL}$ *compared to the perfect CSIT.*

7.1.10.3 Sum Rate Analysis at Low SNR ($\widetilde{\sigma}^2 \propto \frac{1}{P}$)

In Appendix J, we provide the detailed derivation of the low SNR analysis of ESIP-WSR, naive and EWSMSE BFs for the various channel estimates. We observe that the sum rate can be written as

Table 7.2: Low SNR Rate Offset for Various BFs ($\mathbf{D}_{k,c} = \frac{\eta_{k,c}}{L}\mathbf{I}_L$)

χ	naive	EWSMSE	ESIP-WSR
LS	$\frac{\eta_{k,c}}{(\eta_{k,c}+\widetilde{\sigma}^2 M)}$	$\frac{\eta_{k,c}}{(\eta_{k,c}+\widetilde{\sigma}^2 M)}$	$\frac{\eta_{k,c}}{(\eta_{k,c}+\widetilde{\sigma}^2 M)}$
LMMSE/Subspace	$\frac{1}{L}$	$\frac{1}{L}$	$\frac{1}{L}$

Table 7.3: Low SNR Rate Offset for Various BFs (with distinct values in $\mathbf{D}_{k,c}$)

χ	naive	EWSMSE	ESIP-WSR
LS	$\frac{\mathrm{tr}\{\mathbf{D}_{k,c}\}}{\mathrm{tr}\{\mathbf{D}_{k,c}+\widetilde{\sigma}^2\mathbf{I}_M\}}$	$\frac{\mathrm{tr}\{\mathbf{D}_{k,c}\}}{\mathrm{tr}\{\mathbf{D}_{k,c}+\widetilde{\sigma}^2\mathbf{I}_M\}}$	$\frac{\mathrm{tr}\{\mathbf{D}_{k,c}\}}{\mathrm{tr}\{\mathbf{D}_{k,c}+\widetilde{\sigma}^2\mathbf{I}_M\}}$
LMMSE	$\frac{\mathrm{tr}\{\mathbf{D}_{k,c}^3\}}{\mathrm{tr}\{\mathbf{D}_{k,c}\}\mathrm{tr}\{\mathbf{D}_{k,c}^2\}}$	1	1
Subspace	$\frac{1}{L}$	$\frac{1}{L}$	$\frac{1}{L}$

follows,

$$\overline{R} = \sum_{c=1}^{C} \ln(1+\chi_c\rho_c) \overset{a}{\approx} \sum_{c=1}^{C} \chi_c\rho_c, \tag{7.52}$$
$$\text{where,} \quad \rho_c = \eta_{k,c}P,$$

where in (a), we made the approximation $\ln(1+x) \approx x$, when $x \ll 1$ and χ represents the SNR offset for various BFs. With $\eta_{k,c} = \eta, \forall c$, the rate becomes $\overline{R} \approx C\chi\rho, \rho = \eta P$. In Table 7.2 and Table 7.3 (distinct eigenvalues in $\mathbf{D}$), we show the χ_c for different BF and channel estimator combination to explain the SNR offset for sub-optimal BFs compared to the ESIP-WSR BF. Note that at low SNR, ILA-WF allocates all the power to the strongest (in terms of channel attenuation) user resulting in the received SNR $\rho = \eta P$ for the corresponding user.

Few remarks which follow from the Table 7.2 are: 1) From Appendix J, the different BF expression with LMMSE/Subspace channel estimators have the expression $\mathbf{g} \propto \mathbf{C}\widetilde{\mathbf{d}}$. This can be interpreted as random beamforming direction in the covariance subspace. This leads to an SNR offset of $1/L$ for the BFs with LMMSE/Subspace channel estimators compared to the case of distinct eigenvalues in $\mathbf{D}$. In the case of distinct diagonal values in $\mathbf{D}$, the BF expression is $\mathbf{g} \propto \mathbf{CD}\widetilde{\mathbf{d}}$, which can be seen as a weighted random beamforming direction in the covariance subspace. 2) For the LS channel estimate, since it does not involve any subspace projection, the estimation error is present along all the M dimensions which explains the reduction in signal power arising from the term $\widetilde{\sigma}^2 M$ in the denominator. 3) For the subspace channel estimate, the BF expression remains the same in both the case of multiple of identity $\mathbf{D}$ and when there is distinct eigenvalues, $\mathbf{g} \propto \mathbf{C}\widetilde{\mathbf{d}}$. This explains the performance loss compared to the LMMSE case when there are distinct eigenvalues in $\mathbf{D}$, wherein which the LMMSE does a weighting for the BF direction.

For the CoCSIT case, using the analysis in Appendix VII, the sum rate at low SNR can be written as, $\overline{R}_{CoCSIT} = \frac{CP\eta}{L}$. This represents an offset of $\frac{1}{L}$ compared to the perfect CSIT case, whose sum rate can be written as $\overline{R}_{iCSIT} = CP\eta$.

7.1.10.4 High SNR Analysis under Constant Channel Estimation Error

In the Appendix L, we derive in detail the high SNR simplifications for the SE under constant channel estimation error regime. Under constant channel estimation error, the ESIP-WSR BF does pathwise zero forcing and hence the reduction in signal power is $(1-\frac{KL}{M})$. With LMMSE channel estimate, since the estimation error is also reduced to the covariance subspace, ZF to the covariance subspace of the interfering channels imply that the interference power gets reduced to zero. Hence, KL spatial dimensions are used to suppress the inter-cell and intra-cell interference. However, for the LS channel estimate, since the estimation error is present in the entire M dimensional space, interference power still remains. For the naive BF, where the estimation error is not considered in the BF design, ZF to all the interfering user channel estimates ($(K-1)\approx K$ of them) does happen and hence the signal power reduction due to ZF is $(1-\frac{K}{M})$. For the naive BF also, the interference power still remains and the sum rate saturates at high SNR. This explains the drastic improvement in performance between ESIP-WSR BF with LMMSE/Subspace channel estimate compared to the ESIP-WSR BF with LS channel estimate and naive BFs.

Corollary 12.5. *For the ESIP-WSR BF with LMMSE channel estimate, the sum rate can be written as,*

$$\overline{R} = K\ln\left((1-\frac{KL}{M})SNR\frac{C\eta^2}{K(\eta+\widetilde{\sigma}^2L)}\right), \tag{7.53}$$

where the Tx $SNR = P$. So we can conclude that the offset with the perfect CSIT case at high SNR is due to the difference in ZF dimensions and a small attenuation factor in the signal power, $\frac{\eta}{\eta+\widetilde{\sigma}^2L}$ which is due to the non-vanishing channel estimation error. Also, for the CoCSIT case, the sum rate can be obtained as,

$$\overline{R}_{CoCSIT} = K\ln\left((1-\frac{KL}{M})SNR\frac{C\eta}{KL}\right). \tag{7.54}$$

This represents a rate offset (per-user) of $\ln\frac{M-K}{M-KL}+\ln L$ w.r.t the perfect CSIT.

Corollary 12.6. *For the finite rate feedback model (which is one instance of the constant channel estimation error regime as discussed initially), [108] shows that under RVQ (random vector quantization), the distortion is upper bounded as $\widetilde{\sigma}^2 < 2^{-\frac{B}{M-1}}$, B being the number of feedback bits. Further comparing Table 8.1 and 7.4, in order to maintain a rate offset no larger than a specified limit, say $\log_2(b)$ (per-user) between WSR with perfect CSIT and ESIP-WSR, we can obtain that it is sufficient to scale the number of bits per-user as, $B=(M-1)\log_2 Lb-(M-1)\log_2(\eta(\frac{M-K}{M-KL}-b))$.*

7.1.11 Simulation Results

In this section, we present the Ergodic Sum Rate Evaluations for BF design for the various channel estimates. Monte Carlo evaluations of ergodic sum rates are done with the following parameters: C, number of cells. K_c, number of (single-antenna) users in cell c and $K=\sum_c K_c$. M, number of transmit antennas in each cell. We consider a path-wise or low rank channel model as in section 7.1.2, with L = number of paths = channel covariance rank. The elements of the eigenvalue matrix $\mathbf{D}$ is generated from an exponential distribution with mean 1. Further, all the entries are scaled such that $\mathrm{tr}\{\mathbf{D}\}=1$. The eigenvectors, $\mathbf{C}$ of user channel covariance matrix are generated as random unitary matrices. We do evaluate the sum rate performance under channel estimation error inversely proportional to SNR and also the case of constant channel estimation error

Table 7.4: High SNR Rate Offset for Various BFs ($\mathbf{D}_{k,c} = \frac{\eta_{k,c}}{L}\mathbf{I}_L$) under Constant Channel Estimation Error

α	naive	EWSMSE	ESIP-WSR
LS	$\frac{(1-\frac{K}{M})\frac{\eta_{k,c}}{\eta_{k,c}+\tilde{\sigma}^2 M}}{\tilde{\sigma}^2 CP+1}$	$\frac{(1-\frac{K}{M})\frac{\eta_{k,c}}{\eta_{k,c}+\tilde{\sigma}^2 M}}{\tilde{\sigma}^2 CP+1}$	$\frac{(1-\frac{K}{M})\frac{\eta_{k,c}}{\eta_{k,c}+\tilde{\sigma}^2 M}}{\tilde{\sigma}^2 CP+1}$
LMMSE/Subspace	$\frac{(1-\frac{K}{M})\frac{\eta_{k,c}}{(\eta_{k,c}+\tilde{\sigma}^2 L)}}{\frac{CP}{KM}\sum_{i\neq k}^{K}\eta_{k,b_i}+1}$	$\frac{(1-\frac{KL}{M})\frac{\eta_{k,c}}{(\eta_{k,c}+\tilde{\sigma}^2 L)}}{\frac{CP}{KM}\sum_{i\neq k}\eta_{k,b_i}+1}$	$(1-\frac{KL}{M})\frac{\eta_{k,c}}{\eta_{k,c}+\tilde{\sigma}^2 L}$

regime. Notations: in the figures, iCSIT refers to the optimal BF design for the instantaneous (or perfect) CSIT case [9]. "LSA" refers to Large System Approximation. In all the figures, we compare the various BF designs such as ESIP-WSR, EWSMSE and naive under different channel estimates. For the multi-cell simulations, we multiply the inter-cell channels by a random scalar factor (< 1) to represent the attenuation in channel power for inter-cell channels from any BS.

7.1.12 Channel Estimation Error $\propto 1/P$

Here, d denotes the scale factor in the LS channel estimation error variance $\tilde{\sigma}^2 = d/SNR$. When $d = 1$, all the BFs with LMMSE/Subspace channel estimates converge to the optimal WSR based BF design with instantaneous CSIT. In Figure 7.2 and Figure 7.3, we also plot the ESIP-WSR BF performance with LMMSE channel estimator comparing to the ESIP-WSR BF performance for the case of large system approximation. It is evident that the deterministic approximations are accurate even for finite M, K. It is evident from the figure that exploiting the channel estimation error covariance information has significant performance gain compared to the sub-optimal methods such as EWSMSE and naive BFs when d deviates from 1. When $d = 1$, the BFs ESIP-WSR, EWSMSE and Naive EWSR converges to the perfect CSIT case at very high SNR. Also, as discussed in Section 7.1.10, the performance of LS only channel estimate is worse compared to other BFs since the estimation error being present in all the M spatial dimensions.

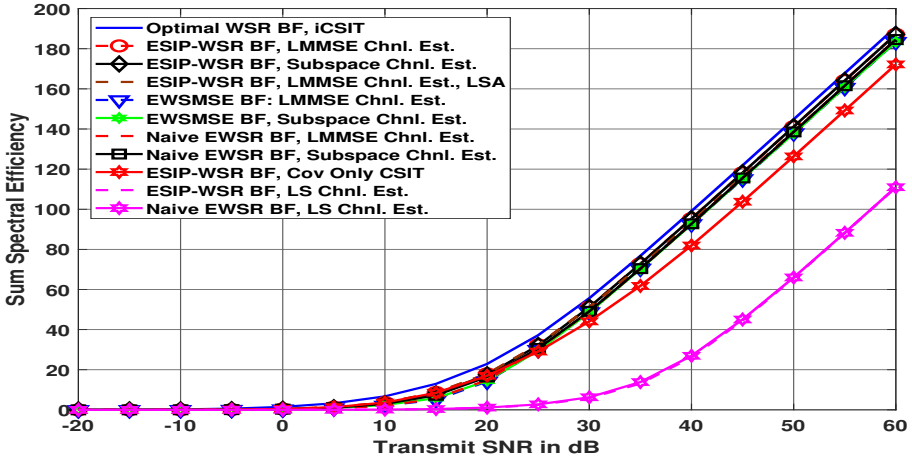


Figure 7.2: EWSR for $C = 1$ cell, $K_1 = K = 20$ users, $M = 64$, $L = 2$, $\tilde{\sigma}^2 = c/SNR, c = 30$.

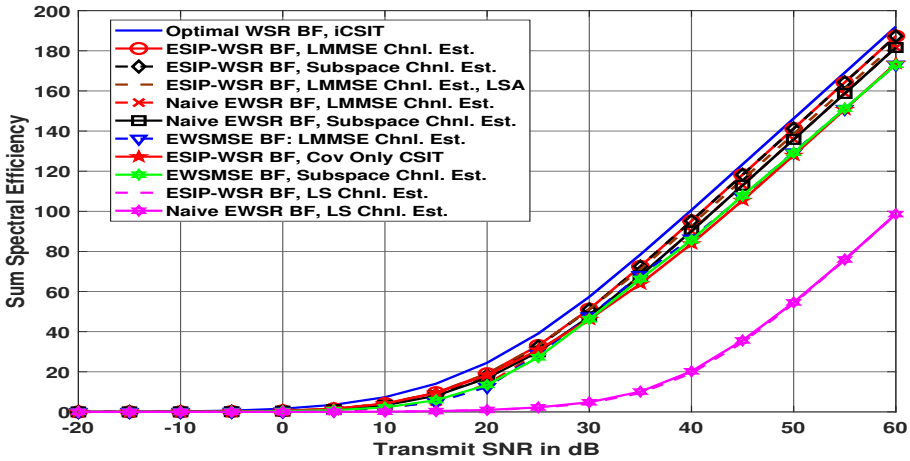


Figure 7.3: EWSR for $C = 2$ cells, $K_1 = K_2 = 10$ users, $M = 64$, $L = 2$, $\widetilde{\sigma}^2 = c/SNR, c = 60$.

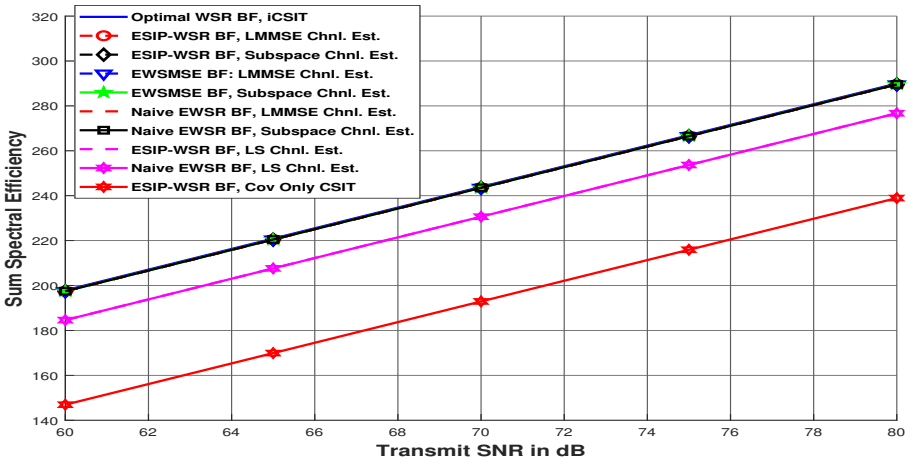


Figure 7.4: EWSR for $C = 1$ cell, $K_1 = K = 20$ users, $M = 64$, $L = 3$, $\widetilde{\sigma}^2 \propto 1/SNR$.

In Figure 7.4, we just plot the high SNR behaviour of the various BF and channel estimator combinations when $d = 1$. It clearly shows the convergence of ESIP-WSR BFs with LMMSE/Subspace and Naive EWSR BFs with LMMSE/Subspace channel estimators to the perfect CSIT sum rate performance. For the BFs with LS channel estimators, from Table 8.1, there is an offset of around 13 bits/sec/Hz which exactly matches the offset seen from the simulations. Similarly for the CoCSIT, there is a sum rate offset of aroung 50 bits/sec/Hz which closely approximates the value predicted by the Corollary 12.4. The huge difference with perfect CSIT is due to the difference in the ZF dimensions for the CoCSIT.

7.1.13 Constant Channel Estimation Error

The constant channel estimate regime looks the most interesting scenario in terms of the superior performance improvement of ESIP-WSR based BF design compared to the very suboptimal schemes such as naive or EWSMSE BFs. The naive and EWSME BFs are observed to saturate at high SNR as seen in Figure 7.6. In the same figure, we also compare the performance of CoCSIT based BF with the ESIP-WSR BF. From the Figure 7.6:a) we deduce that there is a sum rate offset of 17 bits/sec/Hz for the CoCSIT compared to the perfect CSIT which is very close to the rate offset predicted by the large system approximations in Section 7.1.10.4. The BFs with LMMSE and

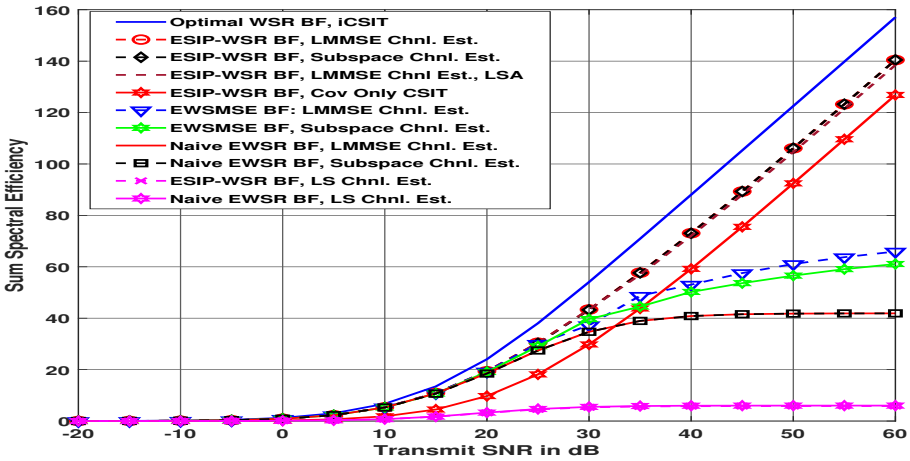


Figure 7.5: EWSR for $C = 1$ cell, $K_1 = K = 15$ users, $M = 100$, $L = 4$, $\widetilde{\sigma}^2 = 0.1$.

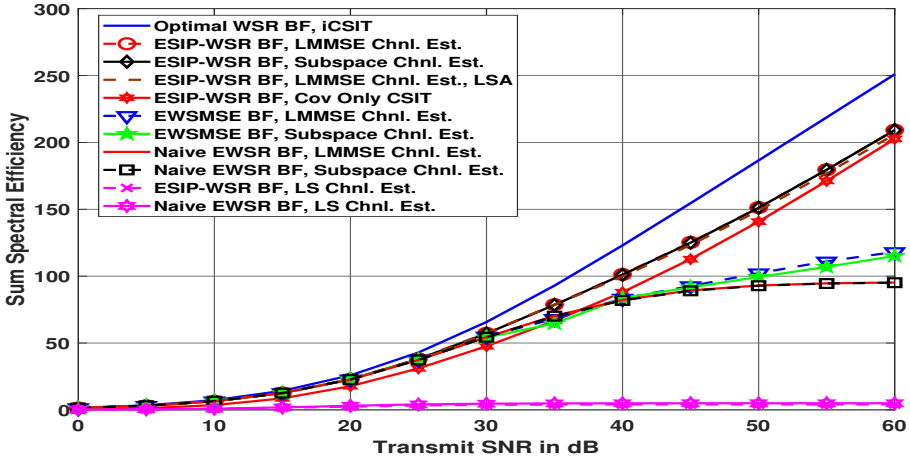


Figure 7.6: EWSR for $C = 4$ cell, $K_i = 7, \forall i$, So, $K = 28$ users, $M = 64$, $L = 2$, $\widetilde{\sigma}^2 = 0.1$.

subspace channel estimators converge to the same performance at high SNR in the simulations which is also analytically proved in the chapter.

7.1.14 Conclusion

Concluding Remarks 6

- This chapter investigated the optimal linear precoder based on partial CSIT in the multi-cell MU-MISO DL. We considered an upper bound of the ergodic capacity to solve the BF design and the tightness of this upper bound in the large antenna limit is also pointed out.
- We introduced a stochastic geometry inspired randomization of the channel covariance eigen spaces of the different users and analyzed the large system behavior. In particular, we focused on a spatial correlation regime where the ratio of the sum of the rank of the channels from a BS to the number of antennas remains a constant. In fact, this condition can capture the strong spatial correlation regimes where the rank of the

Concluding Remarks 6 (cont.)

channel (or the number of multipaths with distinct AoA) scales sublinearly with the number of antennas.

- Moreover, we show the improvement in performance by using an LMMSE channel estimate compared to just having LS estimates, and by furthermore properly exploiting all covariance information. LMMSE channel estimate assumes that the channel covariance matrix is known perfectly at the BS side. Numerical simulations suggest that the large system approximations are accurate even for finite values of M, K. We provided simple and elegant expressions for the sum rate at high and low SNR, providing useful analytical insights into the SNR offsets between different sub-optimal BFs which matches with our simulations. It is also worth noting that we consider two scenarios where the channel estimation error scales differently w.r.t the Tx SNR and the extreme SNR region sum rate expressions derived justify the simulation behaviour.
- However, we remark that few things remains to be done, which is kept as future work. For example, the apparent difference in behaviour at high SNR for EWSMSE and naive EWSR BF under constant channel estimation regime is not captured by the large system simplified results.
- In this chapter, we considered a rather simplified assumption in the channel model that the total number of paths seen by a BS is less than that of the number of Tx antennas. This in turn facilitates the full ZF across the interfering paths, however, it will be more interesting to check the system behaviour when the number of paths exceed the number of Tx antennas.

Part IV

Approximate Bayesian Inference for Sparse Bayesian Learning

Chapter 8

STATIC AND DYNAMIC SPARSE BAYESIAN LEARNING USING MEAN FIELD VARIATIONAL BAYES

8.1 Introduction

Sparse Bayesian Learning (SBL), initially proposed in the Machine Learning literature, is an efficient and well-studied framework for sparse signal recovery. SBL uses hierarchical Bayes with a decorrelated Gaussian prior in which the variance profile is also to be estimated. This is more sparsity inducing than example a Laplacian prior. However, SBL does not scale with problem dimensions due to the computational complexity associated with the matrix inversion in Linear Mimimum Mean Squared Error (LMMSE) estimation. To address this issue, various low complexity approximate Bayesian inference techniques have been introduced for the LMMSE component, including Variational Bayesian (VB) inference, Space Alternating Variational Estimation (SAVE) or Message Passing (MP) algorithms such as Belief Propagation (BP) or Expectation Propagation (EP) or Approximate MP (AMP). These algorithms may converge to the correct LMMSE estimate. In this chapter, we provide a detailed overview of the low complexity approximate Bayesian inference techniques and their superiority (in terms of convergence, computational complexity and robustness w.r.t measurement matrices) compared to the other state of the art techniques.

Sparse signal reconstruction and compressed sensing (CS) has received an enormous amount of attraction in recent years. Few applications include massive multi-input multi-output (MIMO) channel estimation [132], direction of arrival estimation [133], biomagnetic imaging [134], image restoration and echo cancellation. The compressed sensing (CS) problem can be formulated as

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{w}, \tag{8.1}$$

where $\mathbf{y}$ is the observations or data, $\mathbf{A}$ is called the measurement or the sensing matrix which is known and is of dimension $N \times M$ with $N < M$, $\mathbf{x}$ is the M-dimensional sparse signal and $\mathbf{w}$ is the additive noise. $\mathbf{x}$ contains only K non-zero entries, with $K << M$. $\mathbf{w}$ is assumed to be a white Gaussian noise, $\mathbf{w} \sim \mathcal{N}(0, \gamma^{-1}\mathbf{I})$. To address this problem, a variety of algorithms such as the orthogonal matching pursuit [135], the basis pursuit method [136] and the iterative reweighted l_1 and l_2 algorithms [137] exist in the literature. Compared to these algorithms, using Bayesian techniques for sparse signal recovery (SSR) generally achieves the best performance. It is worth mentioning that [138] provides a detailed overview of the various SSR algorithms which fall under l_1 or l_2 norm minimization approaches such as Basis Pursuit, LASSO etc, and SBL

methods. The authors justify the superior recovery performance of SBL compared to the above mentioned conventional methods. In a Bayesian setting, the aim is to calculate the posterior distribution of the parameters given some observations (data) and some a priori knowledge. The SBL algorithm was first introduced by [2] and then proposed for the first time for SSR by [139].

Compared to other state of the art techniques, the critical point about SBL is the hierarchical prior modeling which results in a sparsifying states $\mathbf{x}$. It is also worth mentioning that the Bayesian LASSO [140], uses similar hierarchical modeling which is Gaussian-Exponential prior (equivalent to Laplace prior) and it turns out to be a special case of the Student-t prior in SBL.

In SBL, an estimate of the hyperparameters $\boldsymbol{\alpha}, \gamma$ and sparse signal $\mathbf{x}$ is performed iteratively using evidence maximization. The hyperparameters are estimated first using an evidence maximization, which is referred to as Type II maximum likelihood (ML) method [138]. For a given estimate of $\boldsymbol{\alpha}, \gamma$, the posterior of $\mathbf{x}$ is formulated as $p(\mathbf{x}/\mathbf{y}, \widehat{\boldsymbol{\alpha}}, \widehat{\gamma})$ and the mean of this posterior distribution is used as a point estimate of $\widehat{\mathbf{x}}$. In [141], the authors propose a Fast Marginalized ML (FMML) by alternating maximization of the hyperparameters ξ_i. Both previous approaches allow for a greedy initialization (OMP-like) which improves convergence speed and handles initialization issues. Recently approximate message passing (AMP) [142], generalized AMP and vector AMP [143–145] were introduced to compute the posterior distributions in a message passing (MP) framework and with less complexity. The fundamental idea behind the derivation of AMP is the central limit theorem and Taylor series expansions, which reduces the number of messages to be exchanged in MP. However, so far, the Bayes optimality of these AMP algorithms are shown only for i.i.d. or right orthogonally invariant $\mathbf{A}$, which severely limits the applicability of them.

SBL involves a matrix inversion step at each iteration, which makes it a computationally complex algorithm even for moderately large datasets. An alternative approach to SBL is using a variational approximation for Bayesian inference [146,147]. VB inference tries to find an approximation of the posterior distribution which maximizes the variational lower bound on $\ln p(\mathbf{y})$. [148] introduces a Fast version of SBL by alternatingly maximizing the variational posterior lower bound with respect to single (hyper)parameters.

We also consider the extension of the proposed low complexity algorithms to the dynamic sparse signal case, where the time varyness of the sparse vector is modeled using an autoregressive process of order one (AR(1)). Dynamic autoregressive SBL (DAR-SBL) considered here is a case of joint Kalman filtering (KF) with a linear time-invariant diagonal state-space model, and parameter estimation, which can be considered an instance of nonlinear filtering. In the literature, variations on the KF theme have been derived to handle the joint filtering and parameter estimation problem, such as example the widely used EM-KF algorithm ([149–151]) which uses the famous Expectation Maximization technique (EM), an alternating optimization method of solving ML for the unknown parameters in the AR model. Another well-known variation is the extended KF (EKF) algorithm, which can handle general nonlinear state space models. In this case, the states are extended along with the unknown AR coefficients and hence the new state update equation becomes nonlinear. Another version is the truncated Second-Order EKF (SOEKF) introduced by [152, 153] in which nonlinearities are expanded up to second order, third and higher order statistics being neglected. However, [154] noted that the derivation of SOEKF contains errors due to illogical approximations and a corrected derivation is provided in Henriksen's paper. In ([153, 155]), the Gaussian SOEKF is derived in which fourth-order terms in the Taylor series expansions are retained and approximated by assuming that the underlying joint probability distribution is Gaussian. In [156], Villares at al. introduced the Quadratic Extended KF (QEKF) where they extend the EKF to a new algorithm using quadratic processing and incorporating fourth order statistics of the input signal. The problem of unknown process noise and measurement noise covariance matrices was also tackled in [157]. Here, firstly a statistical

check is done to know whether a particular filter is suboptimal or not. Further, an identification scheme is proposed to obtain asymptotically unbiased and consistent estimates of the unknown variance parameters. The performance of some of these Adaptive KF (AKF) approaches was studied in the literature. In [158], the EM approach was shown to converge to the ML performance. The asymptotic behavior of the EKF for AKF has been treated in [159] where it is proved that no global convergence is guaranteed. The performance analysis of linear and nonlinear KF has also been treated in terms of Cramer Rao Bound (CRB) computations. In [160], the Posterior CRB (PCRB) is developed for the discrete nonlinear KF. Recursive Bayesian CRBs were also developed for continuous and discrete nonlinear filtering for many problems. We can refer to [161] for an overview.

8.1.1 Summary of the Chapter

This chapter of the thesis can be summarized as follows:

- In Section 8.2, the hierarchical prior model for static SBL is introduced. Further, we review the original SBL algorithm proposed by Tipping.
- Followed by that, we give an overview of the existing fast SBL algorithms, by which the readers should get a clarity on the existing methods in the literature and their drawbacks.
- Motivated by the low complexity requirements, we give an overview of our space alternating variational estimation algorithm and the convergence points or guarantees in Section 8.4.2.
- Finally, we extend the SAVE to dynamic sparse states, which integrates hyperparameter estimation also.

8.2 Signal Model-SBL

In Bayesian compressive sensing, a two-layer hierarchical prior is assumed for the $\mathbf{x}$ as in [2]. The hierarchical prior is chosen such that it encourages the sparsity property of $\mathbf{x}$. $\mathbf{x}$ is assumed to have a Gaussian distribution parameterized by $\boldsymbol{\xi} = [\xi_1\ \xi_2\ ...\ \xi_M]$, where ξ_i represents the inverse variance or the precision parameter of x_i.

$$\begin{aligned} p(\mathbf{x}/\boldsymbol{\xi}) &= \prod_{i=1}^{M} p(x_i|\xi_i) \\ &= \prod_{i=1}^{M} \mathscr{CN}(0, \xi_i^{-1}). \end{aligned} \tag{8.2}$$

Further a Gamma prior is considered over $\boldsymbol{\xi}$

$$\begin{aligned} p(\boldsymbol{\xi}) &= \prod_{i=1}^{M} p(\xi_i|a,b) \\ &= \prod_{i=1}^{M} \Gamma^{-1}(a) b^a \xi_i^{a-1} e^{-b\xi_i}. \end{aligned} \tag{8.3}$$

The inverse of noise variance γ is also assumed to have a Gamma prior, $p(\gamma) = \Gamma^{-1}(c) d^c \gamma^{c-1} e^{-d\gamma}$. Now the likelihood distribution can be written as

$$p(\mathbf{y}|\mathbf{x},\gamma) = (2\pi)^{-N} \gamma^N e^{-\gamma \|\mathbf{y}-\mathbf{A}\mathbf{x}\|^2}. \tag{8.4}$$

8.3 SBL using Type-II ML

First, we give a brief overview of the original SBL algorithm to exemplify the motivation behind the low complexity version described in this chapter. Type-I and Type-II ML are two Bayesian approached towards solving the SSR problem. However, Type-II has superior performance as elucidated in [162]. The major factor being that Type-I method seeks the mode of the posterior distribution of $\mathbf{x}$. However, in Type-II, instead of searching for the mode of the true posterior $f_{\mathbf{x}}(\mathbf{x}|\mathbf{y})$, it is approximated as $f_{\mathbf{x}}(\mathbf{x}|\mathbf{y},\widehat{\boldsymbol{\alpha}},\widehat{\gamma})$, where $\widehat{\boldsymbol{\alpha}}$ being obtained by maximizing the true posterior over the subspaces spanned by non zero coefficient indexes. This in turn leading to a better estimate of $\mathbf{x}$, when the true posterior has a skewed peak. For a fixed estimate of the hyperparameters (denoted as $\widehat{\gamma}$, $\widehat{\boldsymbol{\Gamma}}$), the posterior of $\mathbf{x}$ will be Gaussian, i.e.

$$f_{\mathbf{x}}(\mathbf{x}|\mathbf{y},\widehat{\boldsymbol{\alpha}},\widehat{\gamma}) = \mathscr{CN}(\widehat{\mathbf{x}},\boldsymbol{\Sigma}_L), \tag{8.5}$$

leading to the MMSE estimate for $\mathbf{x}$ as follows

$$\begin{aligned} \widehat{\mathbf{x}} &= \widehat{\gamma}(\widehat{\gamma}\mathbf{A}^H\mathbf{A}+\widehat{\boldsymbol{\Gamma}})^{-1}\mathbf{A}^H\mathbf{y}, \\ \boldsymbol{\Sigma}_L &= (\widehat{\gamma}\mathbf{A}^H\mathbf{A}+\widehat{\boldsymbol{\Gamma}})^{-1}. \end{aligned} \tag{8.6}$$

$\boldsymbol{\Sigma}_L$ represents the posterior covariance matrix (with i^{th} diagonal element being σ_i^2) under MMSE estimation. The computational complexity of the above step is $\mathscr{O}(M^3)$, due to the matrix inversion step. Further, the hyperparameters are estimated from the likelihood function by marginalizing over the sparse coefficients $\mathbf{x}$, the marginalized likelihood being denoted as $f_{\mathbf{y}}(\mathbf{y}|\boldsymbol{\alpha},\gamma)$. $\boldsymbol{\alpha},\gamma$ are estimated by maximizing $f_{\mathbf{y}}(\mathbf{y}|\boldsymbol{\alpha},\gamma)$ and this procedure is called as Type-II ML. Type-II ML is solved using EM, which leads to the following updates for the hyperparameters. Note that we represent the point estimates (MMSE) for ξ_i, γ, respectively as $\widehat{\xi}_i,\widehat{\gamma}$. σ_i^2 represents the posterior error variance for x_i.

$$\begin{aligned} \widehat{\xi}_i &= \frac{a+1}{\left(\mathrm{E}(x_i^2)+b\right)}, \\ \text{where} \quad \mathrm{E}(x_i^2) &= \widehat{x}_i^2+\sigma_i^2. \\ \widehat{\gamma} &= \frac{c+N}{\left(\mathrm{E}(\left\|\mathbf{y}-\mathbf{A}\mathbf{x}\right\|^2)+d\right)}, \\ \text{where,} \quad \mathrm{E}(\left\|\mathbf{y}-\mathbf{A}\mathbf{x}\right\|^2) & \\ &= \left\|\mathbf{y}\right\|^2 - 2\mathbf{y}^T\mathbf{A}\widehat{\mathbf{x}} + \mathrm{tr}\left(\mathbf{A}^T\mathbf{A}(\widehat{\mathbf{x}}\widehat{\mathbf{x}}^T+\Sigma)\right), \\ \boldsymbol{\Sigma} &= diag(\sigma_1^2,\sigma_2^2,...,\sigma_M^2), \\ \widehat{\mathbf{x}} &= [\widehat{x}_1,\widehat{x}_2,...,\widehat{x}_M]^T. \end{aligned} \tag{8.7}$$

VB also gives similar result as above, if all the components of $\mathbf{x}$ are considered jointly in the approximate posterior [148]. However, Shutin et. al. note that by computing the stationary points of the variational updates (which is same as that from EM in (8.7)), a fast version of SBL can be constructed. Further analysis of the computed stationary points reveals that SBL with Gaussian priors and noninformative hyperparameters corresponds to pruning components of $\mathbf{x}$ with signal-to-noise-ratio below some threshold. Hence the fast version of SBL in [148] leads to exact sparsity for the estimates $\widehat{\mathbf{x}}$. However, the per-iteration complexity of the resulting algorithm is still high, $\mathscr{O}(L^3)$, where L corresponds to the number of non-zero coefficients of $\mathbf{x}$ retained at any stage of the algorithm. L can be very close to M in the initial iteration stages. Hence, the scope of

the fast SBL by Shutin et. al. is limited as a method of boosting the convergence rate of the original SBL. Furthermore, we note that the same technique can be applied to boost the convergence rate of the low complexity algorithms based on SAVE and BP, which are detailed in the following sections.

Before going further, it is important to also review Type I ML inference method for sparse signal recovery and outlines the difference with Type II described above. Type I is standard MAP estimation (involves integrating out the hyperparameters)

$$\widehat{\mathbf{x}} = \arg\max_{\mathbf{x}} [\log p_{\mathbf{y}}(\mathbf{y}|\mathbf{x}) + p_{\mathbf{x}}(\mathbf{x})], \tag{8.8}$$

while in Type II hyperparameters ($\boldsymbol{\Psi} = \{\boldsymbol{\xi}, \gamma\}$) are estimated using an evidence maximization approach

$$\begin{aligned} \widehat{\boldsymbol{\Psi}} &= \arg\max_{\boldsymbol{\Psi}} p_{\boldsymbol{\Psi}}(\boldsymbol{\Psi}|\mathbf{y}) \\ &= \arg\max_{\boldsymbol{\Psi}} p_{\boldsymbol{\Psi}}(\boldsymbol{\Psi}) \int p_{\mathbf{y}}(\mathbf{y}, \mathbf{x}|\boldsymbol{\Psi})\, d\mathbf{x} \\ &= \arg\max_{\boldsymbol{\Psi}} p_{\boldsymbol{\Psi}}(\boldsymbol{\Psi}) \int p_{\mathbf{y}}(\mathbf{y}|\mathbf{x}, \gamma)\, p_{\mathbf{x}}(\mathbf{x}|\boldsymbol{\xi})\, d\mathbf{x}. \end{aligned} \tag{8.9}$$

Why Type II is better than Type I? In [138], the authors mention that in the evidence maximization framework instead of looking for the mode of the true posterior $p_{\mathbf{x}}(\mathbf{x}|\mathbf{y})$, the true posterior is approximated as $p_{\mathbf{x}}(\mathbf{x}|\mathbf{y}; \widehat{\boldsymbol{\Psi}})$, where $\widehat{\boldsymbol{\Psi}}$ is obtained by maximizing the true posterior mass over the subspaces spanned by the non zero indexes. Type I methods seek the mode of the true posterior and use that as the point estimate of the desired coefficients. Hence, if the true posterior distribution has a skewed peak, then the Type I estimate (Mode) is not a good representative of the whole posterior.

8.3.1 Variational Interpretation of SBL

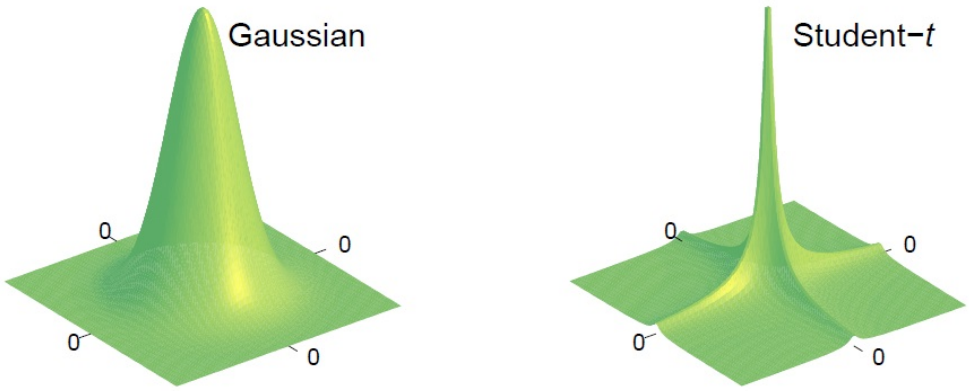


Figure 8.1: Comparing Gaussian and Student-t distributions, source of the figure is [2]. The Gaussian distribution peaks around zero and decays very fast, along all directions, while the student-t have a very sharp peak around zero and falls slowly along the axes. Hence, sparse solutions are favored.

For clarity, we provide here a variational interpretation of the SBL, which tries to address few questions. What exactly is the relationship between the parametrized prior and the presumed

sparse prior (Student-t) derived after the marginalization over the hyperparameters? How does the evidence maximization framework in SBL lead to sparse solution even after using a parametrized Gaussian prior which is not sparsifying? More detailed derivations can also be found in the book [163, Chapter 13]. One popular sparsifying distribution is the Laplacian one.

$$p(\mathbf{x}) = \prod_{i=1}^{M} \frac{\lambda}{2} e^{-\lambda |x_i|}. \tag{8.10}$$

Gaussian likelihood function $p(\mathbf{y}|\mathbf{x},\gamma)$ leads to the MAP estimation identical to the objective function in LASSO [140] (which is with an l_1 norm regularizer function). However, the problem with the Laplacian distribution is that it makes the integral in the computation of $p(\mathbf{y})$ intractable. Indeed, in basis pursuit [136], the cost function involves an l_1 norm regularizer. In a Bayesian sense, this can be interpreted as a Laplacian prior for the sparse vector $\mathbf{x}$. In fact, basis pursuit algorithm is devoid of local minimum and converges always to the global minimum of the cost function. However, one significant shortcoming of basis pursuit is that this converged solution does not necessarily correspond to the sparsest solution for $\mathbf{x}$. A more better sparsifying distribution is the student-t. The probability distribution for student-t can be expressed as follows

$$p(x_i) = \frac{b^a \Gamma(a+\frac{1}{2})}{(2\pi)^{\frac{1}{2}}\Gamma(a)} (b + x_i^2/2)^{-(a+\frac{1}{2})} \tag{8.11}$$

a, b are deterministic parameters which determine the shape of the student-t distribution. In Figure 8.1, we depict a pictorial view of the student-t and Gaussian distributions. From the figure, it is clear that student-t favors coefficients very close to zero, compared to for example, a Gaussian case. Hence, it is a sparsifying distribution. However, with student-t also, it is intractable to compute a closed form expression of the LMMSE estimator for $\mathbf{x}$. To facilitate the estimation in this case, we resort to variational approximation.

$$\begin{aligned} p(\mathbf{y}) &= \int \mathscr{CN}(\mathbf{y}|\mathbf{A}\mathbf{x}, \gamma^{-1}\mathbf{I})\, p(\mathbf{x})\, d\mathbf{x} \\ &\geq \left(\int \mathscr{CN}(\mathbf{y}|\mathbf{A}\mathbf{x}, \gamma^{-1}\mathbf{I}) \mathscr{CN}(\mathbf{0}, \Xi)\, d\mathbf{x} \right) \prod_{i=1}^{M} p(\xi_i) \\ &= \mathscr{CN}(\mathbf{y}|\mathbf{0}, \gamma^{-1}\mathbf{I} + \mathbf{A}\Xi^{-1}\mathbf{A}) \prod_{i=1}^{M} p(\xi_i) = \widehat{p}(\mathbf{y}). \end{aligned} \tag{8.12}$$

Further the posterior of $\mathbf{x}$ becomes

$$p(\mathbf{x}|\mathbf{y},\gamma,\Xi) \approx \widehat{p}(\mathbf{x}|\mathbf{y},\gamma,\Xi) = \frac{\mathscr{CN}(\mathbf{y}|\mathbf{A}\mathbf{x}, \gamma^{-1}\mathbf{I}) \mathscr{CN}(\mathbf{0}, \Xi)}{\widehat{p}(\mathbf{y})} \tag{8.13}$$

Clearly this approximate posterior above is not a bound since normalization w.r.t $\widehat{p}(\mathbf{y})$ has also taken place. In fact, this approximate posterior is a Gaussian, with $\widehat{p}(\mathbf{x}|\mathbf{y},\gamma,\Xi) = \mathscr{CN}(\widehat{\mathbf{x}}, \mathbf{\Sigma})$, where $\widehat{\mathbf{x}}, \mathbf{\Sigma}$ got defined before in (8.7). An alternative interpretation for the hyperparameter estimation using EM algorithm can be given as follows. At each iteration, the EM algorithm maximizes $\mathrm{E}(p(\mathbf{y}|\mathbf{x},\gamma)\widehat{p}(\mathbf{x},\boldsymbol{\xi}))$ (follows from the monotonicity of the logarithm function). This can be equivalently written as the following minimization task

$$\widehat{\boldsymbol{\xi}} = \arg\min_{\boldsymbol{\xi}} \mathrm{E}(p(\mathbf{y}|\mathbf{x},\gamma)|p(\mathbf{x}) - \widehat{p}(\mathbf{x},\boldsymbol{\xi})|). \tag{8.14}$$

Few intuitive interpretations follow from (8.14).

- The hyperparameters $\boldsymbol{\xi}$ are found out such that the it tries to find a best fit of the actual prior for $\mathbf{x}$.
- SBL results in good performance even if the approximation of the prior for $\mathbf{x}$, $p(\mathbf{x})$ is not a good one.
- The major constraint in the approximation of the prior is that the $|p(\mathbf{x}) - \widehat{p}(\mathbf{x}, \boldsymbol{\xi})|$ should be less wherever $p(\mathbf{y}|\mathbf{x}, \gamma)$ is large.
- The approximation of the prior for $\mathbf{x}$ does not matter in regions where the likelihood function $p(\mathbf{y}|\mathbf{x}, \gamma)$ approaches zero.
- Compared to the Laplacian prior (which is the case algorithms like LASSO), SBL provides the flexibility of optimizing extra parameters ($\boldsymbol{\xi}$) to improve the final estimation performance.

8.3.2 Overview of Fast SBL Algorithms

Before discussing further our low complexity VB inference based solutions, we would like to discuss here an overview of the existing state of the art fast SBL algorithms. A first of such algorithms is a fast SBL using Type II ML by Tipping in [141]. This is based on a greedy approach of handling one x_i at a time, plus replacing precisions by their convergence values, leading to pruning of the small x_i components, i.e. explicit sparsity. Fast SBL using VB by Shutin et. al. [148] is another variant inspired from [141]. Shutin uses VB while Tipping is Type II ML as in the original SBL. They do both replace precisions by their convergence values. Shutin also added some extra viewpoints in terms of the pruning condition being interpreted as relating between sparsity properties of SBL and a measure of SNR. Main message of the both being faster convergence compared to original SBL and both do not lead to much reduction in per iteration complexity. BP-SBL [164] uses BP to compute the MMSE estimate of $\mathbf{x}$, while retaining EM for the hyperparameter estimates. In SBL, with fixed hyperparameters, MAP or MMSE estimate (follows from the Gaussian posterior) of $\mathbf{x}$ can be efficiently computed using BP, since all the messages involved are Gaussian (without any approx.). Hyperparameter free sparse estimation [165] does not require hyperparameter tuning compared to SBL. It uses the technique of covariance matching and is equivalent to a weighted version of square root LASSO. In Figure **??**, a comparison of the different fast SBL versions is provided.

8.3.3 Variational Bayes

The computation of the posterior distribution of the parameters is usually intractable. In order to address this issue, in VB framework, the posterior distribution $p(\mathbf{x}, \boldsymbol{\xi}, \gamma|\mathbf{y})$ is approximated by a variational distribution $q(\mathbf{x}, \boldsymbol{\xi}, \gamma)$ that has the factorized form:

$$\begin{aligned} q(\mathbf{x}, \boldsymbol{\xi}, \gamma) &= q_\gamma(\gamma) \prod_{i=1}^{M} q_{x_i}(x_i) \prod_{i=1}^{M} q_{\xi_i}(\xi_i) \\ &= \prod_k q_k(\theta_k). \end{aligned} \tag{8.15}$$

We denote by $\boldsymbol{\theta} = (\mathbf{x}, \boldsymbol{\xi}, \gamma)$ the vector of unknown parameters and θ_k represents each scalar parameter in $\boldsymbol{\theta}$. $q_k(\theta_k)$ represents the approximate posterior marginal of θ_k. Variational Bayes

Table 8.1: Complexity Comparisons-SBL Algorithms.

Algorithm	Complexity per iteration	Convergence (No.) of iterations	Sparsity	Optimization Function	Local Optimum
Type I	$\mathcal{O}(M^3)$		Exact sparsity	Type I ML (Depending upon the prior used type I ML corresponds to LASSO or re-weighted l_1/l_2 min. problems	
Type II SBL	$\mathcal{O}(M^3)$		Exact sparsity (ξ_i converges to ∞	Type II ML solved using EM	
Fast SBL using Type II ML (Tipping'03, focus more on convergence speed)	$\mathcal{O}(L^3), L \leq M$	$\ll L$	Exact sparsity (using an entry dependent thresholding condition which follows from the computation of stationary point of ξ_i)	ξ_i are computed to accelerate convergence	Convergence to a local optimum
Fast SBL using VB by Shutin (focus more on convergence speed)	$\mathcal{O}(L^3), L \leq M$	$\ll L$	Exact sparsity (using a pruning condition similar as in Tipping's	Maximization of ELBO in VB	Convergence to a local optimum of ELBO (MFFE)
Hyperparameter free SBL (Zachariah, Stoica'15)	$\mathcal{O}(M^2)$	$\ll M$	The final objective function is a weighted square root LASSO. So the sum of l_2 norm of (y and $\mathbf{Ax}$) and weighted l_1 norm of $\mathbf{x}$ which promotes sparsity here.	LMMSE estimator for $\mathbf{x}$, with Covariance matching for PDP finally giving rise to an objective function which can be interpreted as weighted square root LASSO.	Convergence to a local optimum
BP-SBL (Tan, Li'10)	$\mathcal{O}(MN)$[1]	$\log(MN)$	Does not give exact sparsity	Posterior of $\mathbf{x}$ computed using BP	Convergence to local optimum of Bethe Free Energy (BFE)
GAMP-SBL (Shoukairi, Schniter, Rao'18)	$\mathcal{O}(MN)$	$\ll M$	Does not give exact sparsity	Using GAMP for posterior of $\mathbf{x}$, EM for hyperparameters	Convergence to local optimum of LSL-BFE
SAVE-SBL (Shoukairi, Schniter, Rao'18)	$\mathcal{O}(MN)$	$\ll M$	Does not give exact sparsity	Maximization of ELBO in VB	Convergence to local optimum of ELBO
Inverse Free SBL (Duan, Yang, Fang, Li'17)	$\mathcal{O}(MN)$	$\ll M$	Does not give exact sparsity	Maximization of an approximate ELBO in VB	Convergence to a local optimum of the approximate ELBO

compute the factors q by minimizing the Kullback-Leibler distance between the true posterior distribution $p(\mathbf{x},\boldsymbol{\xi},\gamma|\mathbf{y})$ and the $q(\mathbf{x},\boldsymbol{\xi},\gamma)$. From [146]

$$KLD_{VB} = KL\big(p(\mathbf{x},\boldsymbol{\xi},\gamma|\mathbf{y})||q(\mathbf{x},\boldsymbol{\xi},\gamma)\big) \tag{8.16}$$

The KL divergence minimization is equivalent to maximizing the evidence lower bound (ELBO) [147]. To elaborate on this, we can write the marginal probability of the observed data as

$$\begin{aligned} \ln p(\mathbf{y}) &= L(q) + KLD_{VB}, \text{ where,} \\ L(q) &= \int q(\boldsymbol{\theta}) \ln \frac{p(\mathbf{y},\boldsymbol{\theta})}{q(\boldsymbol{\theta})} d\boldsymbol{\theta}, \\ KLD_{VB} &= -\int q(\boldsymbol{\theta}) \ln \frac{p(\boldsymbol{\theta}|\mathbf{y})}{q(\boldsymbol{\theta})} d\boldsymbol{\theta}. \end{aligned} \tag{8.17}$$

Since $KLD_{VB} \geq 0$, it implies that $L(q)$ is a lower bound on $\ln p(\mathbf{y})$. Moreover, $\ln p(\mathbf{y})$ is independent of $q(\boldsymbol{\theta})$ and therefore maximizing $L(q)$ is equivalent to minimizing KLD_{VB}. This is called ELBO maximization and doing this in an alternating fashion for each variable in $\boldsymbol{\theta}$ leads to

[1] (Similar complexity as xAMP,see matrix form of the BP-SBL)

$$\begin{aligned} \ln(q_i(\theta_i)) &= < \ln p(\mathbf{y},\boldsymbol{\theta}) >_{k\neq i} + c_i, \\ p(\mathbf{y},\boldsymbol{\theta}) &= p(\mathbf{y}|\mathbf{x},\boldsymbol{\xi},\gamma)\, p(\mathbf{x}|\boldsymbol{\xi})\, p(\boldsymbol{\xi})\, p(\gamma). \end{aligned} \tag{8.18}$$

where $\boldsymbol{\theta} = \{\mathbf{x},\boldsymbol{\xi},\gamma\}$ and θ_i represents each scalar in $\boldsymbol{\theta}$. Here $<>_{k\neq i}$ represents the expectation operator over the distributions $q_k(\theta_k)$ for all $k \neq i$.

8.4 SAVE Sparse Bayesian Learning

In this section, we propose a Space Alternating Variational Estimation (SAVE) algorithm based on alternating optimization between each elements of $\boldsymbol{\theta}$. For SAVE, not any particular structure of $\mathbf{A}$ is assumed, in contrast to AMP which performs poorly when $\mathbf{A}$ is not i.i.d. or sub-Gaussian. The joint distribution can be written as

$$\begin{aligned} \ln p(\mathbf{y},\boldsymbol{\theta}) = N\ln\gamma - \gamma\left\|\mathbf{y}-\mathbf{A}\mathbf{x}\right\|^2 + \sum_{i=1}^{M}\left(\ln\xi_i - \xi_i x_i^2\right) + \sum_{i=1}^{M}\left((a-1)\ln\xi_i + a\ln b - b\xi_i\right) \\ + (c-1)\ln\gamma + c\ln d - d\gamma + \text{constants}, \end{aligned} \tag{8.19}$$

In the following, $c_{x_i}, c'_{x_i}, c_{\xi_i}$ and c_γ represents normalization constants for the respective pdfs.
Update of $q_{x_i}(x_i)$: Using (8.18), $\ln q_{x_i}(x_i)$ turns out to be quadratic in x_i and thus can be represented as a Gaussian distribution as follows

$$\begin{aligned} \ln q_{x_i}(x_i) &= -<\gamma>\left\{ < \left\|\mathbf{y}-\mathbf{A}_{\bar{i}}\mathbf{x}_{\bar{i}}\right\|^2 > - (\mathbf{y}-\mathbf{A}_{\bar{i}}<\mathbf{x}_{\bar{i}}>)^H\mathbf{A}_i x_i - \right. \\ &\quad \left. x_i\mathbf{A}_i^H(\mathbf{y}-\mathbf{A}_{\bar{i}}<\mathbf{x}_{\bar{i}}>) + \|\mathbf{A}_i\|^2 x_i^2\right\} - <\xi_i> x_i^2 + c_{x_i} \\ &= -\frac{1}{\sigma_i^2}\left(x_i - \mu_i\right)^2 + c'_{x_i}. \end{aligned} \tag{8.20}$$

Note that we split $\mathbf{A}\mathbf{x}$ as, $\mathbf{A}\mathbf{x} = \mathbf{A}_i x_i + \mathbf{A}_{\bar{i}}\mathbf{x}_{\bar{i}}$, where $\mathbf{A}_i$ represents the i^{th} column of $\mathbf{A}$, $\mathbf{A}_{\bar{i}}$ represents the matrix with i^{th} column of $\mathbf{A}$ removed, x_i is the i^{th} element of $\mathbf{x}$, and $\mathbf{x}_{\bar{i}}$ is the vector without x_i. The mean and the variance of the resulting Gaussian distribution ($x_i \sim \mathcal{N}(<x_i>, \sigma_i^2)$) becomes

$$\begin{aligned} \sigma_i^2 &= \frac{1}{<\gamma>\|\mathbf{A}_i\|^2 + \xi_i}, \\ <x_i> = \mu_i &= \sigma_i^2\mathbf{A}_i^H\left(\mathbf{y} - \mathbf{A}_{\bar{i}}<\mathbf{x}_{\bar{i}}>\right)<\gamma>, \end{aligned} \tag{8.21}$$

where μ_i represents the point estimate of x_i.
Update of $q_{\xi_i}(\xi_i)$: The variational approximation leads to the following Gamma distribution for the $q_{\xi_i}(\xi_i)$

$$\begin{aligned} \ln q_{\xi_i}(\xi_i) &= a\ln\xi_i - \xi_i\left(<x_i^2> + b\right) + c_{\xi_i}, \\ q_{\xi_i}(\xi_i) &\propto \xi_i^a e^{-\xi_i(<x_i^2>+b)}. \end{aligned} \tag{8.22}$$

The mean of the Gamma distribution is given by

$$\begin{aligned} <\xi_i> &= \frac{a+1}{\left(<x_i^2> + b\right)}, \\ \text{where} \quad <x_i^2> &= \mu_i^2 + \sigma_i^2. \end{aligned} \tag{8.23}$$

Update of $q_\gamma(\gamma)$: Similarly, the Gamma distribution from the VB approximation for the $q_\gamma(\gamma)$ can be written as $q_\gamma(\gamma) \propto \gamma^{c+N-1} e^{-\gamma\left(<\|\mathbf{y}-\mathbf{A}\mathbf{x}\|^2>+d\right)}$. The mean of the Gamma distribution for γ is given by

$$
\begin{aligned}
<\gamma> &= \frac{c+N}{\left(<\left\|\mathbf{y}-\mathbf{A}\mathbf{x}\right\|^2>+d\right)}, \\
\text{where, } <\left\|\mathbf{y}-\mathbf{A}\mathbf{x}\right\|^2> &= \left\|\mathbf{y}\right\|^2 - \mathbf{y}^H\mathbf{A}\boldsymbol{\mu} - \boldsymbol{\mu}^H\mathbf{A}^H\mathbf{y} + \\
&\quad \mathrm{tr}\left(\mathbf{A}^H\mathbf{A}(\boldsymbol{\mu}\boldsymbol{\mu}^H + \Sigma)\right), \\
\Sigma &= diag(\sigma_1^2, \sigma_2^2, ..., \sigma_M^2), \\
\boldsymbol{\mu} &= [\mu_1, \mu_2, ..., \mu_M]^T.
\end{aligned} \tag{8.24}
$$

From (8.21), it can be seen that the estimate of $\mathbf{x} = \boldsymbol{\mu}$ converges to the L-MMSE equalizer, $\widehat{\mathbf{x}} = \boldsymbol{\mu} = (\mathbf{A}^H\mathbf{A} + \frac{1}{<\gamma>}\boldsymbol{\Sigma}^{-1})^{-1}\mathbf{A}^H\mathbf{y}$.

8.4.1 Computational Complexity

For our proposed SAVE, it is evident that we do not need any matrix inversions compared to [148, 166]. Our computational complexity is similar to [167]. Update of all the variable $\mathbf{x}, \boldsymbol{\alpha}, \gamma$ involves simple addition and multiplication operations. We introduce the following variables, $\mathbf{q} = \mathbf{y}^H\mathbf{A}$ and $\mathbf{B} = \mathbf{A}^H\mathbf{A}$. $\mathbf{q}, \mathbf{B}$ and $\left\|\mathbf{y}\right\|^2$ can be precomputed, so only computed once. We also introduce the following notations, $\mathbf{x}_{i-} = [x_1 ... x_{i-1}]^T, \mathbf{x}_{i+} = [x_{i+1} ... x_M]^T$. Also, we represent $\gamma^t = <\gamma>$, $\xi_i^t = <\xi_i>$, $x_i^t = \mu_i$ and $\boldsymbol{\Sigma}^t = \boldsymbol{\Sigma}$ in the following sections, where t represents the iteration stage.

Algorithm 13: SAVE SBL Algorithm

Given: $\mathbf{y}, \mathbf{A}, M, N$.
Initialization: a, b, c, d are taken to be very low, on the order of 10^{-10}. $\xi_i^0 = a/b, \forall i, \gamma^0 = c/d$ and $\sigma_i^{2,0} = \frac{1}{\|\mathbf{A}_i\|^2\gamma^0 + \xi_i^0}, \mathbf{x}^0 = \mathbf{0}$.
At iteration $t+1$,

1. Update $\sigma_i^{2,t+1}$, $x_i^{t+1} = \mu_i, \forall i$ from (8.21) using $\mathbf{x}_{i-}^{t+1}$ and $\mathbf{x}_{i+}^t$.
2. Compute $< x_i^{2,t+1} >$ from (8.23) and update ξ_i^t.
3. Update the noise variance, γ^{t+1} from (8.24).
4. Continue steps $1-4$ till convergence of the algorithm.

8.4.2 Convergence Analysis of SAVE or Mean Field Approximation

For MF or SAVE (space alternating variational estimation) [168], to obtain the free energy $F(q) = U(q) - H(q)$

$$\begin{aligned} U(q) &= -\mathrm{E}_q \ln p(\mathbf{x}|\mathbf{y}) \\ &= \mathrm{E}_q(\mathbf{x}^H\boldsymbol{\Sigma}_L^{-1}\mathbf{x} - \mathbf{y}^H\mathbf{A}\mathbf{x} - \mathbf{x}^H\mathbf{A}^H\mathbf{y}) \\ &= \boldsymbol{\mu}^H\boldsymbol{\Sigma}_L^{-1}\boldsymbol{\mu} - \mathbf{y}^H\mathbf{A}\boldsymbol{\mu} - \boldsymbol{\mu}^H\mathbf{A}^H\mathbf{y} + \sum_i \sigma_i^2(\boldsymbol{\Sigma}_L^{-1})_{i,i} + c_1, \\ H(q) &= -\sum_i \mathrm{E}_{q_i} \ln q_i \\ &= \frac{1}{2}\sum_i \ln \sigma_i^2 + c_2, \end{aligned} \tag{8.25}$$

c_i being constants, independent of $\boldsymbol{\mu}$ and σ_i^2, also $q_i(x_i) = \mathcal{N}(\mu_i, \sigma_i^2), \boldsymbol{\mu} = \widehat{\mathbf{x}} = [\mu_1, .., \mu_M]^T$. Now the MF free energy can be written as

$$F(q) = \boldsymbol{\mu}^H\boldsymbol{\Sigma}_L^{-1}\boldsymbol{\mu} - \mathbf{y}^H\mathbf{A}\boldsymbol{\mu} - \boldsymbol{\mu}^H\mathbf{A}^H\mathbf{y} + \sum_i \sigma_i^2(\boldsymbol{\Sigma}_L^{-1})_{i,i} + \tfrac{1}{2}\sum_i \ln \sigma_i^2 + c. \tag{8.26}$$

It can be noticed that $F(q)$ is a convex function w.r.t $\boldsymbol{\mu}$ and σ_i^2, further optimizing this w.r.t $\boldsymbol{\mu}$ leads to $\boldsymbol{\mu} = \boldsymbol{\Sigma}_L\mathbf{A}^H\mathbf{y}$ and $\sigma_i^2 = \frac{1}{(\Sigma_L^{-1})_{i,i}}$. So we can conclude that the mean converges to LMMSE in the case of SAVE while the variance is not exact. Further, we analyze the convergence conditions. The SAVE iterations for $\boldsymbol{\mu}$ follow

$$\begin{aligned} \text{Let } \mathbf{D} &= \mathrm{diag}\,(\gamma\mathbf{A}^H\mathbf{A} + \boldsymbol{\Gamma}), \\ \mathbf{H} &= \mathrm{offdiag}(\gamma\mathbf{A}^H\mathbf{A}), \\ \mathbf{x}^{(t+1)} &= -\mathbf{D}^{-1}\mathbf{H}\mathbf{x}^{(t)} + \mathbf{D}^{-1}\gamma\mathbf{A}^H\mathbf{y}, \end{aligned} \tag{8.27}$$

Note that MF can also be implemented as message passing in a factor graph. Hence, it is evident from the above expression that the factor graph representation for SAVE corresponds to the case when all the y_i's are treated jointly and all the x_i's at the scalar level. Noting that LMMSE estimate of $\mathbf{x}$ can be written as the solution of $\mathbf{J}\mathbf{x} = \mathbf{b}$, with $\mathbf{J} = \gamma\mathbf{A}^H\mathbf{A} + \boldsymbol{\Gamma}$ and $\mathbf{b} = \gamma\mathbf{A}^H\mathbf{y}$. In fact, SAVE corresponds to the Jacobi iterations [169] for solving this linear system with the splitting of $\mathbf{J} = \mathbf{D} - \mathbf{H}$, which converges to the true value only if $\rho(\mathbf{D}^{-1}\mathbf{H}) < 1$, where ρ represents the spectral radius. Further, we observe that if we rewrite the SAVE iterations as, $x_i^{(t+1)} = \sigma_i^2\mathbf{A}_i^H\left(\mathbf{y} - \mathbf{A}_{i-}\mathbf{x}_{i-}^{(t+1)} - \mathbf{A}_{i+}\mathbf{x}_{i+}^{(t)}\right)\gamma$,, where in the update of x_i at iteration $(t+1)$ we include the updated values of $x_k, k = 1, ..., i-1$. These updated recursions correspond to Gauss-Siedel method [169] for solving the linear system $\mathbf{J}\mathbf{x} = \mathbf{b}$. In Gauss-Siedel version, $\mathbf{J}$ is split as $\mathbf{J} = \mathbf{D} - \mathbf{L} - \mathbf{U}$, where $\mathbf{L}$ is a matrix that represents the lower triangular portion of $\mathbf{H}$ and $\mathbf{U}$ representing the upper triangular portion. Hence for Gauss-Siedel, the SAVE iterations (8.27) can be rewritten as, $\mathbf{x}^{(t+1)} = (\mathbf{D}-\mathbf{L})^{-1}\mathbf{U}\mathbf{x}^{(t)} + (\mathbf{D}-\mathbf{L})^{-1}\gamma\mathbf{A}^H\mathbf{y}$. Certain remarks on the convergence behavior (assuming $\mathbf{A}$ is real) follow as below,

Remarks 7

- From [169], if $\mathbf{J}$ is an $M-$matrix, then Jacobi and Gauss-Siedel iterations for SAVE converge to the true values $\mathbf{x}^* = \mathbf{J}^{-1}\mathbf{b}$, for any arbitrary b. For $\mathbf{J}$ to be an $M-$matrix, it should be nonsingular and $\mathbf{J}^{-1} >= 0$. Moreover the off-diagonal elements, $J_{ij} < 0, \forall i, j, j \neq i$. Also, the diagonal elements of $\mathbf{J}$ represented by $\mathbf{D}$ is nonnegative and nonsingular.

- Another sufficient condition for convergence follows from the diagonal dominance theorem in [169], which says that if $\mathbf{J}$ is strictly or irreducibly diagonally dominant then $\widehat{\mathbf{x}}$ converges to $\mathbf{x}^*$.

- Gauss-Seidel iterations (iterated linear SIC) converges for any $\mathbf{J} = \mathbf{J}^H > 0$! Because these iterations correspond to minimizing $\mathbf{x}^H\mathbf{J}\mathbf{x} - \mathbf{x}^H\mathbf{b} - \mathbf{b}^H\mathbf{x}$ by alternating optimization sweeps over the components of $\mathbf{x}$. Alternating minimization with $\text{diag}(\mathbf{J}) > 0$ is guaranteed to lead to a local minimum and since the cost function is convex ($\mathbf{J} > 0$), it is even the global minimum.

- To further accelerate the convergence, one possibility is to employ the successive over-relaxation method (SOR) [169], in which case, the SAVE iterations gets modified as follows. $\mathbf{x}^{(t+1)} = \mathbf{x}^{(t)} + \omega(\overline{\mathbf{x}}^{(t+1)} - \mathbf{x}^{(t)})$, where $\overline{\mathbf{x}}^{(t+1)}$ corresponds to the Jacobi SAVE iterations (8.27) or the Gauss-Siedel iterations.

- To fix the convergence of SAVE (when $\rho(\mathbf{D}^{-1}\mathbf{H}) > 1$), we can use the diagonal loading method similar to [170]. The modified iterations (with a diagonal loading factor matrix $\mathbf{\Lambda}$) can be written as

$$\begin{aligned}(\mathbf{D}+\mathbf{\Lambda})\mathbf{x}^{(t+1)} &= -(\mathbf{H}-\mathbf{\Lambda})\mathbf{x}^{(t)} + \gamma\mathbf{A}^H\mathbf{y}, \Longrightarrow \\ \mathbf{x}^{(t+1)} &= -(\mathbf{D}+\mathbf{\Lambda})^{-1}(\mathbf{H}-\mathbf{\Lambda})\mathbf{x}^{(t)} + (\mathbf{D}+\mathbf{\Lambda})^{-1}\gamma\mathbf{A}^H\mathbf{y},\end{aligned} \tag{8.28}$$

  The convergence condition gets modified as $\rho((\mathbf{D}+\mathbf{\Lambda})^{-1}(\mathbf{H}-\mathbf{\Lambda})) < 1$. Another point worth noting here is that, if the power delay profile $\mathbf{\Gamma}$ is also estimated using VB as in [168], then we can write $\mathbf{D} = \gamma\,\text{diag}(\mathbf{A}^H\mathbf{A}) + \widehat{\mathbf{\Gamma}}$, where $\widehat{\mathbf{\Gamma}} = \mathbf{\Gamma} + \widetilde{\mathbf{\Gamma}}$. In this case, $\widetilde{\mathbf{\Gamma}}$ may represent an automatic correction factor (diagonal loading) to force convergence of SAVE for cases where $\rho(\mathbf{D}^{-1}\mathbf{H}) > 1$.

8.4.3 Sparsity Analysis with SAVE

In this subsection, we focus on the sparsity analysis of the SAVE iterations described above. Before going into the technical details, we would like to first throw some insights into how sparsification happens in SBL. With a Gamma prior on the precision parameters ξ_i, the marginal pdf of x_i becomes a sparsifying distribution (Student-t), so, in the case of an under determined noiseless system, this will tend to a sparse solution for $\mathbf{x}$. But more generally, in the presence of noise, or if the system is not underdetermined, how does "sparsification" happen? In some techniques of by Tipping [141], or Shutin [148], they set values below a certain threshold to zero. However, such a process of setting to zero whatever is below some threshold may not be optimal according to SBL (i.e it is not inherent in the solution of SBL). Note that in variations of LASSO, there are hard or soft thresholding techniques, which come out automatically of the problem

formulation. Apart from setting stuff to zero, there can be other forms of sparsification. E.g., in Bayesian techniques, the estimate is drawn towards the prior mean. So, if the prior mean is zero, they are drawn to zero, leading to bias. This already happens with Gaussian prior. But it is not clear whether other types of prior (example Student-t) lead to a stronger "shrinking" effect.

We use the approach described in [141, 148], where they compute the stationary point of the precision components ξ_i. The expression for the mean value of ξ_i (for the resulting Gamma posterior from [168]) is, $\widehat{\xi}_i = \frac{a+\frac{1}{2}}{\left(\frac{<x_i^2>}{2}+b\right)}$,where, $< x_i^2 >= \widehat{x}_i^2 + \sigma_i^2$. Further substituting for $\widehat{x}_i^2$ in $\widehat{\xi}_i$

$$\widehat{\xi}_i^{-1} \overset{(a)}{=} \frac{\gamma^2}{(\gamma \mathbf{A}_i^H \mathbf{A}_i + \widehat{\xi}_i)^2}[\mathrm{tr}\{\mathbf{y}\mathbf{y}^H \mathbf{A}_i \mathbf{A}_i^H\} + \mathrm{tr}\{\mathbf{A}_{\bar{i}}^H \mathbf{A}_{\bar{i}} \mathbf{\Sigma}_{\bar{i}} \mathbf{A}_{\bar{i}}^H \mathbf{A}_{\bar{i}}\}] + \frac{1}{\gamma \mathbf{A}_i^H \mathbf{A}_i + \widehat{\xi}_i}, \tag{8.29}$$

We define $c_i = \mathrm{tr}\{\mathbf{y}\mathbf{y}^H \mathbf{A}_i \mathbf{A}_i\}, d_i = \mathrm{tr}\{\mathbf{A}_{\bar{i}}^H \mathbf{A}_{\bar{i}} \mathbf{\Sigma}_{\bar{i}} \mathbf{A}_{\bar{i}}^H \mathbf{A}_i\}$, where $\mathbf{\Sigma}_{\bar{i}}$ is a diagonal matrix with entries $\sigma_n^2, \forall n \neq i$. Also, we made the large system approximation $(M, N \to \infty)$ that $\mathbf{A}_i^H \mathbf{y} \widehat{\mathbf{x}}_{\bar{i}}^H \mathbf{A}_{\bar{i}}^H \mathbf{A}_i \to \mathrm{tr}\{\mathrm{E}(\widehat{\mathbf{x}}_{\bar{i}}^H \mathbf{A}_{\bar{i}}^H \mathbf{A}_i \mathbf{A}_i^H \mathbf{y})\} = 0$. After some algebraic manipulations, solving (8.29) which is of the form $\xi_i^{-1} = \mathscr{F}(\xi_i)$ leads to the following stationary point for ξ_i

$$\widehat{\xi}_i = \begin{cases} \frac{\gamma(\mathbf{A}_i^H \mathbf{A}_i)^2}{\gamma(c_i + d_i) - \mathbf{A}_i^H \mathbf{A}_i}, & \text{if, } \gamma(c_i + d_i) > \mathbf{A}_i^H \mathbf{A}_i \\ \infty, & \text{if, } \gamma(c_i + d_i) \leq \mathbf{A}_i^H \mathbf{A}_i \end{cases} \tag{8.30}$$

The above threshold condition can be intuitively interpreted as follows: $c_i + d_i$ can be interpreted as the signal power in $\mathbf{y}' = \mathbf{y} - \mathbf{A}_{\bar{i}} \mathbf{x}_i$. Hence the threshold above checks whether the signal-to-noise ratio of the residual signal (after the matched filtering by $\mathbf{A}_i$) is greater than 1. As observed in [148], this should further accelerate the convergence of the SAVE iterations.

8.4.4 Simulation results

For the observation model, $\mathbf{y}_t$ is of dimension 100×1 and $\mathbf{x}_t$ is of size 200×1 with 30 non-zero elements. All signals are considered to be real in the simulation. All the elements of $\mathbf{A}_t$ (time varying) are generated i.i.d. from a Gaussian distribution with mean 0 and variance 1. The rows of $\mathbf{A}_t$ are scaled by $\sqrt{30}$ so that the signal part of any scalar observation has a unit variance. Taking the SNR to be 20dB, the variance of each element of $\mathbf{v}_t$ (Gaussian with mean 0) is computed as 0.01. We compare our algorithm with the Fast Inverse-Free SBL (Fast IF SBL) in [167], the G-AMP based SBL in [166] and the fast version of SBL (FV SBL) in [148].

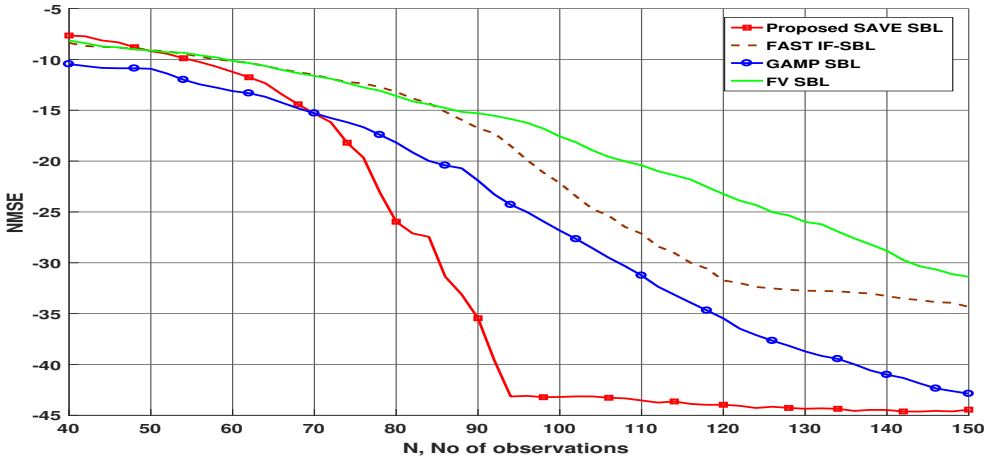


Figure 8.2: NMSE vs the number of observations.

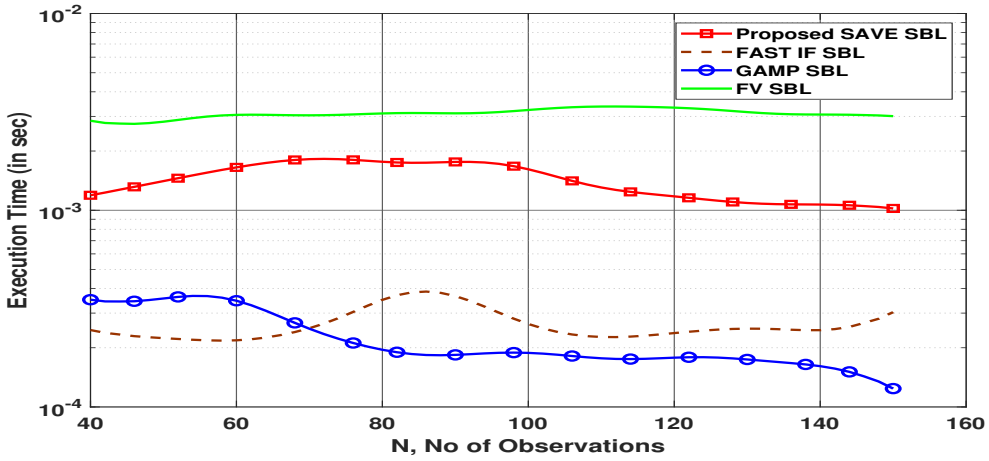


Figure 8.3: Execution time vs the number of observations.

Few remarks follow:

Remarks 8

- The performance of FV-SBL is exactly that of SBL.
- For a sufficient amount of data, SAVE has significantly lower MSE than the other fast algorithms. This is because while a priori, performing component-wise VB for $\mathbf{x}$ whereas joint VB can be done may seem like a bad idea for performance. However, not only the parameters $\mathbf{x}$ but also the hyperparameters ξ and γ need to be estimated simultaneously. The resulting problem appears to be characterized by many local optima. Apparently, the component-wise approach VB appears to allow to avoid a lot of bad local optima, explaining the better performance, apart from lower complexity. At a very low amount of data, suboptimal approaches such as AMP which do not introduce individual hyperparameters per $\mathbf{x}$ component and assume that the x_i behave i.i.d., behave better because of the lower number of hyperparameters to be estimated.

8.4.5 Conclusion

We presented a fast SBL algorithm called SAVE, which uses the variational inference techniques to approximate the posteriors of the data and parameters. SAVE helps to circumvent the matrix inversion operation required in conventional SBL using EM algorithm. We showed that the proposed algorithm has a faster convergence rate and better performance in terms of NMSE than even the state of the art fast SBL solutions.

8.4.6 Open Issues: Reduced Complexity Linear Tx/Rx Computation

An optimal linear Tx/Rx filter in MU MIMO is of the form

$$\mathbf{F} = (\mathbf{A}\mathbf{D}_1\mathbf{A}^H + \lambda\mathbf{I})^{-1}\mathbf{A}\mathbf{D}_2. \tag{8.31}$$

Other sub-optimal beamformers are special cases of this, where, for the R-ZF, $\mathbf{D}_1 = \mathbf{I}$ and ZF $\lambda \to 0$. LMMSE Tx/Rx can also be found by SAVE. Consider the case of a multi-user UL system,

with $\mathbf{A} \in \mathscr{C}^{N\times M}$ SIMO channel, x as the $M \times 1$ transmit signal from all users and $y \in C^{N\times 1}$ is the received signal at the BS. From the convergence analysis of SAVE, it can be seen that the estimate of $\mathbf{x} = \mu$ converges to the L-MMSE equalizer,

$$\begin{aligned} &\sigma_v^2\mu_i + \sigma_i^2\mathbf{A}_i^H\mathbf{A}\mu = \sigma_i^2\mathbf{A}_i^H\mathbf{y} \Rightarrow \\ &\widehat{\mathbf{x}} = \mu = (\mathbf{A}^H\mathbf{A} + \sigma_v^2\mathbf{\Sigma}^{-1})^{-1}\mathbf{A}^H\mathbf{y}. \end{aligned} \tag{8.32}$$

SAVE recursions are similar to PE [171]. However, PE only converges in case of sufficient diagonal dominance of $\mathbf{A}^H\mathbf{A}$, whereas SAVE is guaranteed to converge, employing implicitly varying damping factors (the σ_i^2).

8.5 Dynamic SBL-System Model

In this section, we start looking at the dynamic SBL case. Sparse signal $\mathbf{x}_t$ is modeled using an AR(1) process with a diagonal correlation coefficient matrix $\mathbf{F}$, which can be written as follows

$$\begin{aligned} \text{State Update: } &\mathbf{x}_t = \mathbf{F}\mathbf{x}_{t-1} + \mathbf{w}_t, \\ \text{Observation: } &\mathbf{y}_t = \mathbf{A}^{(t)}\mathbf{x}_t + \mathbf{v}_t, \end{aligned} \tag{8.33}$$

where $\mathbf{x}_t = \left[x_{1,t},\ ...,\ x_{M,t}\right]^T$. Diagonal matrices $\mathbf{F}$ and $\mathbf{\Gamma}$ are defined with its elements, $\mathbf{F}_{i,i} = f_i, f_i \in (-1,1)$ and $\mathbf{\Xi} = \text{diag}(\boldsymbol{\xi}), \boldsymbol{\xi} = [\xi_1, ...\xi_M]$. Here ξ_i represents the inverse variance of $x_{i,t} \sim \mathscr{CN}(0, \frac{1}{\xi_i})$. Further, $\mathbf{w}_t \sim \mathscr{CN}(\mathbf{0}, \mathbf{\Lambda}^{-1})$, where $\mathbf{\Lambda}^{-1} = \mathbf{\Xi}^{-1}(\mathbf{I} - \mathbf{F}\mathbf{F}^H) = \text{diag}(\frac{1}{\lambda_1}, ..., \frac{1}{\lambda_M})$ and $\mathbf{v}_t \sim \mathscr{CN}(\mathbf{0}, \frac{1}{\gamma}\mathbf{I})$. $\mathbf{w}_t$ are the complex Gaussian mutually uncorrelated state innovation sequences. Hence we sparsify the prediction error variance $\mathbf{w}_t$ also, with the same support as $\mathbf{x}_0$ and henceforth enforces the same support set for $\mathbf{x}_t, \forall t$. One remark here is that If we apply SBL now to the prediction error variances of $\mathbf{x}_t$, then trying to sparsify a prediction error variance actually encourage both that the actual variance gets sparse and that the variation gets sparse because a prediction error variance is small if either the quantity variance is small or its variation is small. $\mathbf{v}_t$ is independent of the $\mathbf{w}_t$ process. Although the above signal model seems simple, there are numerous applications such as

- Bayesian adaptive filtering [172] (or wireless channel estimation [150], [173]):
  in this case, $\mathbf{x}_k$ = FIR filter response, and θ contains example the Power Delay Profile (diagonal of a diagonal filter coefficient covariance matrix $\mathbf{P}_0 = \mathbf{P}_k$, and the AR(1) dynamics in example diagonal $\mathbf{F}$ and $\mathbf{Q}$.

  - Position tracking (GPS) (see [174] and references therein):

$$\mathbf{x}_{t+1} = \begin{bmatrix} 1 & \Delta t & \frac{1}{2}\Delta t^2 \\ 0 & 1 & \Delta t \\ 0 & 0 & 1 \end{bmatrix} \cdot \mathbf{x}_t = \begin{bmatrix} x_t + \Delta t \cdot v_t + \frac{1}{2}\Delta t^2 a \\ v_t + \Delta t \cdot a \\ a \end{bmatrix}$$

    the state contains position, velocity and possible acceleration and θ contains acceleration model parameters (example white noise, AR(1))

- Blind Audio Source Separation (BASS) [175]: x_k = source signals, θ: (short+long term) AR parameters, reverb filters

In Bayesian compressive sensing, a two-layer hierarchical prior is assumed for the $\mathbf{x}$ as in [2]. The hierarchical prior is chosen such that it encourages the sparsity property of $\mathbf{x}_t$ or of the innovation sequences $\mathbf{v}_t$. The state update gets represented as

$$p(\mathbf{x}_t/\mathbf{x}_{t-1},\mathbf{F},\mathbf{\Gamma}) = \prod_{i=1}^{M} \mathscr{CN}(f_i x_{i,t-1}, \frac{1}{\lambda_i}). \tag{8.34}$$

For the convenience of analysis, we reparameterize ξ_i in terms of λ_i and assume a Gamma prior for $\mathbf{\Lambda}$, $p(\mathbf{\Lambda}) = \prod_{i=1}^{M} p(\lambda_i/a,b) = \prod_{i=1}^{M} \Gamma^{-1}(a) b^a \lambda_i^{a-1} e^{-b\lambda_i}$. The inverse of noise variance γ is also assumed to have a Gamma prior, $p(\gamma/c,d) = \Gamma^{-1}(c) d^c \gamma_i^{c-1} e^{-d\gamma}$, such that the marginal pdf of $\mathbf{x}_t$ (Student-t distribution) becomes more sparsity inducing than example, a Laplacian prior. The advantage is that the whole machinery of linear MMSE estimation can be exploited, such as example, the Kalman filter. But this is embedded in other layers making things eventually non-Gaussian. Now the likelihood distribution can be written as, $p(\mathbf{y}_t/\mathbf{x}_t,\gamma) = (2\pi)^{-N}\gamma^N e^{-\gamma\|\mathbf{y}_t - \mathbf{A}^{(t)}\mathbf{x}_t\|^2}$. To make these priors non-informative (Jeffrey's prior), we choose them to be small values $a = c = b = d = 10^{-5}$. For the AR(1) coefficients f_k, we do not assume any prior distribution. We define the unknown parameter vector $\boldsymbol{\theta} = \{\mathbf{x}, \mathbf{\Lambda}, \gamma, \mathbf{F}\}$ and θ_i represents each scalar in $\boldsymbol{\theta}$.

8.5.1 Gaussian Posterior Minimizing the KL Divergence

In [176], for any distribution $p(\mathbf{x})$, the Gaussian distribution $q(\mathbf{x}) \sim \mathscr{CN}(\boldsymbol{\mu}, \mathbf{\Sigma})$ which minimizes the Kullback-Leibler divergence, $KL(p||q)$, reduces to matching the mean and covariance

$$\begin{aligned} \boldsymbol{\mu} &= <\mathbf{x}>_{p(x)}, \\ \mathbf{\Sigma} &= <\mathbf{x}\mathbf{x}^H>_{p(x)} - <\mathbf{x}>_{p(x)} <\mathbf{x}>_{p(x)}^H . \end{aligned} \tag{8.35}$$

8.6 SAVE SBL and Kalman Filtering

For the system model in (8.33), a classical method is to estimate the sparse states (for given hyperparameter estimates, possibly using EM) using Kalman filtering (KF). KF is an efficient iterative algorithm whose steps can be summarized as follows,
The prediction step:

$$\begin{aligned} \widehat{\mathbf{x}}_{t|t-1} &= \widehat{\mathbf{F}}\widehat{\mathbf{x}}_{t-1|t-1}, \\ \widehat{\mathbf{P}}_{t|t-1} &= \widehat{\mathbf{F}}\mathbf{P}_{t-1|t-1}\widehat{\mathbf{F}}^H + \frac{1}{\widehat{\lambda}}\mathbf{I}. \end{aligned} \tag{8.36}$$

The measurement step:

$$\begin{aligned} \mathbf{K}_t &= \mathbf{P}_{t|t-1}\mathbf{A}^H(\mathbf{A}\mathbf{P}_{t|t-1}\mathbf{A}^H + \frac{1}{\widehat{\gamma}})^{-1}, \\ \widehat{\mathbf{x}}_{t|t} &= \widehat{\mathbf{x}}_{t|t-1} + \mathbf{K}_t(\mathbf{y}_t - \mathbf{A}\widehat{\mathbf{x}}_{t|t-1}), \\ \mathbf{P}_{t|t} &= (\mathbf{I} - \mathbf{K}_t\mathbf{A})\mathbf{P}_{t|t-1}. \end{aligned} \tag{8.37}$$

Note that due to the matrix inversion operation in the measurement stage, the computational complexity of the original KF does not scale with the problem size. Our novel contribution here is a low complexity KF using VB inference techniques, which we describe in the following sections.

In this section, we propose a Space Alternating Variational Estimation (SAVE) based alternating optimization between each element of $\mathbf{x}_t$ or γ. For SAVE, no particular structure of $\mathbf{A}_t$ is assumed, in contrast to AMP which performs poorly when $\mathbf{A}_t$ is not i.i.d. or is sub-Gaussian. The joint distribution w.r.t the observation of (8.33) can be written as

$$p(\mathbf{y}_t, \mathbf{x}_t, \boldsymbol{\theta}/\mathbf{y}_{1:t-1}) = p(\mathbf{y}_t/\mathbf{x}_t, \boldsymbol{\theta})\, p(\mathbf{x}_t, \boldsymbol{\theta}/\mathbf{y}_{1:t-1}). \tag{8.38}$$

In the following, $c_{x_{k,t}}, c'_{x_{k,t}}, c_{\xi_k}$, c_{λ_k}, c_{x-1}, c_{x_t}, c'_{x_t} and c_γ represents normalization constants for the respective pdfs.

8.6.1 Diagonal AR(1) (DAR(1)) Prediction Stage

In this stage, we compute the prediction about $\mathbf{x}_t$ given the observations till time $t-1$, $\widehat{x}_{k,t|t-1}$. This involves the computation of the posterior $p(\mathbf{x}_t, \boldsymbol{\theta}/\mathbf{y}_{1:t-1})$. The joint distribution for the state space model can be written as

$$\begin{aligned} &\ln p(x_{k,t}, x_{k,t-1}, f_k, \lambda_k | \mathbf{y}_{1:t-1}) = -\lambda_k (x_{k,t} - f_k x_{k,t-1})^H (x_{k,t} - f_k x_{k,t-1}) - \\ &\frac{1}{\sigma^2_{k,t-1|t-1}} |x_{k,t-1} - \widehat{x}_{k,t-1|t-1}|^2 + ((a-1)\ln\lambda_k + a\ln b - b\lambda_k)\,. \end{aligned} \tag{8.39}$$

The prediction about $\mathbf{x}_t$ can be computed from the time update equation of the standard Kalman filter

$$x_{k,t} = \widehat{f}_{k|t-1} x_{k,t-1} + \widetilde{f}_{k|t-1} x_{k,t-1} + w_{k,t}. \tag{8.40}$$

Here we denote $\widehat{f}_{k|t-1}$ as the estimate of f_k given the observations till $t-1$ and $\widetilde{f}_{k|t-1}$ represents the error in the estimation. Similary we can represent $x_{k,t-1} = \widehat{x}_{k,t-1|t-1} + \widetilde{x}_{k,t-1|t-1}$, $\widetilde{x}_{k,t-1|t-1}$ being the estimation error.

$$\begin{aligned} &\widehat{x}_{k,t|t-1} = \widehat{f}_{k|t-1} \widehat{x}_{k,t-1|t-1}, \\ &\widetilde{x}_{k,t|t-1} = \widehat{f}_{k|t-1} \widetilde{x}_{k,t-1|t-1} + \widetilde{f}_{k|t-1} x_{k,t-1} + w_{k,t}, \\ \implies &\sigma^2_{k,t|t-1} \overset{(a)}{=} |\widehat{f}_{k|t-1}|^2 \sigma^2_{k,t-1|t-1} + \sigma^2_{f_k} (|\widehat{x}_{k,t-1|t-1}|^2 + \sigma^2_{k,t-1|t-1}) + \frac{1}{\widehat{\lambda}_k | t-1}, \end{aligned} \tag{8.41}$$

In the variational approximation, we assume that the posterior of f_k and $x_{k,t}$ are independent. (a) in (8.41) follows from this argument. Further the predictive distribution $p(\mathbf{x}_t/\mathbf{y}_{1:t-1})$ can be approximated to be Gaussian distributed (refer to the discussion in section 8.5.1) with mean $\widehat{\mathbf{x}}_{t|t-1} = [\widehat{x}_{1,t|t-1}, ..., \widehat{x}_{M,t|t-1}]^T$ and diagonal error covariance $\widehat{\mathbf{P}}_{t|t-1} = \text{diag}(\sigma^2_{1,t|t-1}, ..., \sigma^2_{M,t|t-1})$. Actually this parametric $q()$ fitting, we only need it for the prediction stage of $\mathbf{x}_t$, all other q's (filtering or smoothing of $\mathbf{x}_t$, all hyperparameters) come out simple due to the choice of conjugate priors. Further the joint distribution in (8.38) can be obtained as

$$\begin{aligned} &\ln p(\mathbf{y}_t, \mathbf{x}_t, \boldsymbol{\theta}/\mathbf{y}_{1:t-1}) = N\ln\gamma - \gamma \left\| \mathbf{y}_t - \mathbf{A}_t \mathbf{x}_t \right\|^2 - M \ln\det(\widehat{\mathbf{P}}_{t|t-1}) - \\ &\left(\mathbf{x}_t - \widehat{\mathbf{x}}_{t|t-1}\right)^H \widehat{\mathbf{P}}^{-1}_{t|t-1} \left(\mathbf{x}_t - \widehat{\mathbf{x}}_{t|t-1}\right) + (c-1)\ln\gamma + c\ln d - d\gamma + \text{constants}, \end{aligned} \tag{8.42}$$

8.6.2 Measurement or Update Stage

Update of $q_{x_{k,t}}(x_{k,t})$: Using (8.18), $\ln q_{x_{k,t}}(x_{k,t})$ turns out to be quadratic in $x_{k,t}$ and thus can be represented as a Gaussian distribution as follows

$$\begin{aligned}\ln q_{x_{k,t}}(x_{k,t}) &= -<\gamma>\Big\{(\mathbf{y}_t-\mathbf{A}_{t,\overline{k}}<\mathbf{x}_{\overline{k},t}>)^H\mathbf{A}_{t,k}x_{k,t}-x_{k,t}^H\mathbf{A}_{t,k}^H(\mathbf{y}_t-\mathbf{A}_{t,\overline{k}}<\mathbf{x}_{\overline{k},t}>)\\ &\quad+\left\|\mathbf{A}_{t,k}\right\|^2\left|x_{k,t}\right|^2\Big\}-\frac{1}{{\sigma^2}_{k,t|t-1}}\Big(\left|x_{k,t}\right|^2-x_{k,t}^H\widehat{x}_{k,t|t-1}-x_{k,t}\widehat{x}_{k,t|t-1}^H\Big)+c_{x_{k,t}}\\ &=-\frac{1}{\sigma^2_{k,t|t}}\left|x_{k,t}-\widehat{x}_{k,t|t}\right|^2+c'_{x_{k,t}}.\end{aligned}\tag{8.43}$$

Note that we split $\mathbf{A}_t\mathbf{x}_t$ as, $\mathbf{A}_t\mathbf{x}_t=\mathbf{A}_{t,k}x_{k,t}+\mathbf{A}_{t,\overline{k}}\mathbf{x}_{\overline{k},t}$, where $\mathbf{A}_{t,k}$ represents the k^{th} column of $\mathbf{A}_t$, $\mathbf{A}_{t,\overline{k}}$ represents the matrix with k^{th} column of $\mathbf{A}_t$ removed. Clearly, the mean and the variance of the resulting Gaussian distribution becomes

$$\begin{aligned}\sigma^{-2,(i)}_{k,t|t}&=<\gamma>\left\|\mathbf{A}_{t,k}\right\|^2+\sigma^{-2,(i-1)}_{k,t|t},\\ <x^{(i)}_{k,t|t}>&=\sigma^{2,(i)}_{k,t|t}\left(\mathbf{A}^H_{t,k}\left(\mathbf{y}_t-\mathbf{A}_{t,\overline{k}}<\mathbf{x}^{(i-1)}_{\overline{k},t}>\right)<\gamma>+\frac{\widehat{x}_{k,t|t-1}}{\sigma^2_{k,t|t-1}}\right),\end{aligned}\tag{8.44}$$

where i represents the iteration stage with $\lim_{i\to\infty}<x^{(i)}_{k,t|t}>=\widehat{x}_{k,t|t}$ represents the point estimate of $x_{k,t}$. However, in (8.44) the computation of $<x^{(i)}_{k,t|t}>$ requires the knowledge of $<\mathbf{x}^{(i)}_{\overline{k},t}>$. So we need to perform enough iterations between the components of $<x_{k,t|t}>$ till convergence. Moreover, we initialize $<x^{(0)}_{k,t|t}>$ by $\widehat{x}_{k,t|t-1}$and $\sigma^{-2,(0)}_{k,t}=\sigma^{-2}_{k,t|t-1}$, which is obtained in the prediction stage. One remark is that forcing a Gaussian posterior q with diagonal covariance matrix on the original Kalman measurement equations gives the same result as SAVE. Note that the derivations in [177] for VB-KF are not correct as it does not have the correct variance expressions that vary with iteration! For the convenience of the derivations in the following sections, we define $\widehat{\mathbf{P}}_{t|t}=diag(\sigma^2_{1,t|t},...,\sigma^2_{M,t|t})$, $\widehat{\mathbf{x}}_{t|t}=[\widehat{x}_{1,t|t},...,\widehat{x}_{M,t|t}]^T$.

8.6.3 Fixed Lag Smoothing

Kalman filtering in the EM-KF is not enough to adapt the hyperparameters, instead we need atleast a lag 1 smoothing [178]. Motivated by this result, we propose fixed lag smoothing with delay 1 for SAVE-KF. We rewrite the state space model as follows

$$\begin{aligned}\mathbf{y}_t&=\mathbf{A}_t\mathbf{F}\mathbf{x}_{t-1}+\underbrace{\mathbf{A}_t\mathbf{w}_{t-1}+\mathbf{v}_t}_{\widetilde{\mathbf{v}}_t},\\ p(\mathbf{y}_t,\mathbf{x}_{t-1},\boldsymbol{\theta}/\mathbf{y}_{1:t-1})&=p(\mathbf{y}_t/\mathbf{x}_{t-1},\boldsymbol{\theta})p(\mathbf{x}_{t-1},\boldsymbol{\theta}/\mathbf{y}_{1:t-1}),\end{aligned}\tag{8.45}$$

where $\widetilde{\mathbf{v}}_t\sim\mathscr{CN}(\mathbf{0},\widetilde{\mathbf{R}}_t)$, $\widetilde{\mathbf{R}}_t=\mathbf{A}_t\boldsymbol{\Lambda}\mathbf{A}_t^H+\frac{1}{\gamma}\mathbf{I}$. The posterior distribution $p(\mathbf{x}_{t-1}/\mathbf{y}_{1:t-1})$ is approximated using variational approximation as $q(\mathbf{x}_{t-1}/\mathbf{y}_{1:t-1})$ with mean and covariance as $\widehat{\mathbf{x}}_{t-1|t-1}$ and $\widehat{\mathbf{P}}_{t-1|t-1}$.

$$\begin{aligned}&\ln p(\mathbf{y}_t,\mathbf{x}_{t-1},\boldsymbol{\theta}/\mathbf{y}_{1:t-1})=\tfrac{-1}{2}\ln\det\widetilde{\mathbf{R}}_t-\\ &(\mathbf{y}_t-\mathbf{A}_{t,k}f_kx_{k,t-1}-\mathbf{A}_{t,\overline{k}}\mathbf{F}_{\overline{k}}\mathbf{x}_{\overline{k},t-1})^H\widetilde{\mathbf{R}}_t^{-1}(\mathbf{y}_t-\mathbf{A}_{t,k}f_kx_{k,t-1}-\mathbf{A}_{t,\overline{k}}\mathbf{F}_{\overline{k}}\mathbf{x}_{\overline{k},t-1})\\ &-\tfrac{1}{2}\det(\widehat{\mathbf{P}}_{t-1|t-1})-\left(\mathbf{x}_{t-1}-\widehat{\mathbf{x}}_{t-1|t-1}\right)^H\widehat{\mathbf{P}}^{-1}_{t-1|t-1}\left(\mathbf{x}_{t-1}-\widehat{\mathbf{x}}_{t-1|t-1}\right)+c_{x-1},\end{aligned}\tag{8.46}$$

where $\mathbf{F}_{\overline{k}}$ represents $\mathbf{F}$ with k^{th} column and row removed.
Prediction of $\mathbf{x}_{t-1}$: Using (8.18), $\ln q_{\mathbf{x}_{t-1}}(\mathbf{x}_{t-1}/\mathbf{y}_{1:t})$ turns out to be quadratic in $\mathbf{x}_{t-1}$ and thus can

be represented as a Gaussian distribution with mean and covariance as $\widehat{\mathbf{x}}_{t-1|t}$ and $\widehat{\mathbf{P}}_{t-1|t}$ respectively

$$\begin{aligned}
\sigma^{-2,(i)}_{k,t-1|t} &= (\widehat{f}^2_{k|t} + \sigma^2_{f_k|t})\mathbf{A}^H_{t,k}\widetilde{\mathbf{R}}^{-1}_t\mathbf{A}_{t,k} + \sigma^{-2,(i-1)}_{k,t-1|t}, \\
\widehat{\mathbf{P}}_{t-1|t} &= \text{diag}(\sigma^2_{1,t-1|t}, ..., \sigma^2_{M,t-1|t}), \\
< x^{(i)}_{k,t-1|t} > &= \sigma^{2,(i)}_{k,t-1|t}(\widehat{f}^H_{k|t}\mathbf{A}^H_{t,k}\widetilde{\mathbf{R}}^{-1}_t(\mathbf{y}_t - \mathbf{A}_{t,\bar{k}}\mathbf{F}_{\bar{k}|t} < \mathbf{x}^{(i-1)}_{\bar{k},t-1|t} >) + \frac{\widehat{x}_{k,t-1|t-1}}{\sigma^2_{k,t-1|t-1}}).
\end{aligned} \tag{8.47}$$

Note that, in the algorithm implementation as shown in Algorithm 1 below, we introduce an iterative procedure (with i denoting the stage number) for the smoothing updates unlike [177] where there is no iteration for the covariance part. Note that we initialize the mean and variance in (8.47) from the converged values from the filtering stage.

8.6.4 Estimation of Hyperparameters

Update of $q_\gamma(\gamma)$: The Gamma distribution from the VB approximation for the $q_\gamma(\gamma)$ can be written as

$$\begin{aligned}
\ln q_\gamma(\gamma) &= (c - 1 + N)\ln\gamma - \gamma\left(< \left\|\mathbf{y}_t - \mathbf{A}_t\mathbf{x}_t\right\|^2 > + d\right) + c_\gamma, \\
q_\gamma(\gamma) &\propto \gamma^{c+N-1}e^{-\gamma\left(< \left\|\mathbf{y}_t - \mathbf{A}_t\mathbf{x}_t\right\|^2 > + d\right)}.
\end{aligned} \tag{8.48}$$

The mean of the Gamma distribution for γ is given by

$$\begin{aligned}
< \gamma > = \widehat{\gamma}_t &= \frac{c + \frac{N}{2}}{(\zeta_t + d)}, \\
\zeta_t &= \beta\zeta_{t-1} + (1 - \beta) < \left\|\mathbf{y}_t - \mathbf{A}_t\mathbf{x}_t\right\|^2 >, \text{ where,} \\
< \left\|\mathbf{y}_t - \mathbf{A}_t\mathbf{x}_t\right\|^2 > &= \left\|\mathbf{y}_t\right\|^2 - 2\Re(\mathbf{y}^H_t\mathbf{A}_t\widehat{\mathbf{x}}_{t|t}) + \text{tr}\left(\mathbf{A}^H_t\mathbf{A}_t(\widehat{\mathbf{x}}_{t|t}\widehat{\mathbf{x}}^H_{t|t} + \widehat{\mathbf{P}}_{t|t})\right),
\end{aligned} \tag{8.49}$$

where we introduced temporal averaging also and β denotes the weighting coefficients which are less than one.

Update of $q_{f_k}(f_k)$: Using variational approximation we get a quadratic expression for $\ln q(f_k|\mathbf{y}_{1:t}) \sim \text{E}_{q(\mathbf{x}_t,\mathbf{x}_{t-1},\mathbf{\Lambda}/\mathbf{y}_{1:t})}\ln p(\mathbf{x}_t, \mathbf{x}_{t-1}, \mathbf{\Lambda}, \mathbf{y}_{1:t})$. Finally we write the mean and variance of the resulting Gaussian distribution as

$$\begin{aligned}
\sigma^2_{f_k|t} &= \frac{1}{\lambda_k < x^2_{k,t-1} >_{|t}}, \\
\widehat{f}_{k|t} &= \frac{< x_{k,t|t}x^H_{k,t-1|t} >_{|t}}{< x^2_{k,t-1} >_{|t}}
\end{aligned} \tag{8.50}$$

Here $<>_{|t}$ represents the temporal average given the observations till time t. We introduce temporal averaging here to approximate terms of the form $< x_{k,t|t}x^H_{k,t-1|t} >$. This is done using the orthogonality property of LMMSE. So $< x_{k,t|t}x^H_{k,t-1|t} > = < \widehat{x}_{k,t|t}\widehat{x}^H_{k,t-1|t} > + < \widetilde{x}_{k,t|t}\widetilde{x}^H_{k,t-1|t} >$. The Kalman filter (in linear state-space models and Gaussian noise) provides instantaneous $\widehat{x}_{k,t|t}, \widehat{x}^H_{k,t-1|t}$ and $\sigma^2_{k,t|t}, \sigma^2_{k,t-1|t}$. This explains why we do temporal averaging (sample average replacing statistical average). We define $\widehat{\mathbf{P}}_{\mathbf{F}|t} = \text{diag}(\sigma^2_{f_1|t}, ..., \sigma^2_{f_M|t})$. Also we define the following

covariance matrices, $R_t^{m,n} =< \mathbf{x}_{t-n}\mathbf{x}_{t-m}^H >_{|t}$ and ξ_t represents the temporal weighting coefficient which is less than one [178]

$$\begin{aligned}
&\mathbf{R}_t^{0,0} = (1-\xi_t)\,\mathbf{R}_{t-1}^{0,0} + \xi_t\,(\widehat{\mathbf{x}}_{t|t}\widehat{\mathbf{x}}_{t|t}^H + \widehat{\mathbf{P}}_{t|t}),\\
&\mathbf{R}_t^{1,0} = (\mathbf{R}_t^{0,1})^H = (1-\xi_t)\,\mathbf{R}_{t-1}^{1,0} + \xi_t\,\mathbf{F}(\widehat{\mathbf{x}}_{t-1|t}\widehat{\mathbf{x}}_{t-1|t}^H + \widehat{\mathbf{P}}_{t-1|t}),\\
&\mathbf{R}_t^{1,1} = (1-\xi_t)\,\mathbf{R}_{t-1}^{1,1} + \xi_t(\widehat{\mathbf{x}}_{t-1|t}\widehat{\mathbf{x}}_{t-1|t}^H + \widehat{\mathbf{P}}_{t-1|t}).
\end{aligned} \tag{8.51}$$

In (8.51), if we do not use lag-1 smoothing, at time 0, combining with (8.50), it is clear that $\widehat{\mathbf{F}} = \mathbf{R}^{0,1}(\mathbf{R}^{0,0})^{-1} = \mathbf{F}$. In other words, we would need the knowledge of the true $\mathbf{F}$ to estimate it. This indeed is a sufficient condition to show that filtering is not sufficient to estimate the hyperparameters. Further, we denote the $(i,j)^{th}$ element of $\mathbf{R}_t^{m,n}$ as $\mathbf{R}_t^{m,n}(i,j)$.
Update of $q_{\lambda_k}(\lambda_k)$: Using variational approximation
$\ln q(\lambda_k|\mathbf{y}_{1:t}) \sim \mathrm{E}_{q(\mathbf{x}_t,\mathbf{x}_{t-1},f_k/\mathbf{y}_{1:t})} \ln p(\mathbf{x}_t,\mathbf{\Lambda},f_k|\mathbf{y}_{1:t})$, leading to

$$\begin{aligned}
&\ln\lambda_k - \lambda_k(< \left|x_{k,t} - f_k x_{k,t-1}\right|^2 > + b) + (a-1)\ln\lambda_k + c_{\lambda_k},\\
&q_{\lambda_k}(\lambda_k) \propto \lambda_k^a e^{-\lambda_k\left(<|x_{k,t}-f_k x_{k,t-1}|^2>+b\right)}.
\end{aligned} \tag{8.52}$$

The resulting gamma distribution is parameterized just by one quantity, the mean value, which gets used in the prediction stage and can be written as

$$< \lambda_k >= \frac{(a+1)}{(<|x_{k,t}-f_k x_{k,t-1}|^2>_{|t}+b)}. \tag{8.53}$$

The temporal average $< \left|x_{k,t} - f_k x_{k,t-1}\right|^2 >_{|t}$ can be written as

$$< \left|x_{k,t} - f_k x_{k,t-1}\right|^2 >_{|t} = \mathbf{R}_t^{0,0}(k,k) - 2\Re\{\widehat{f}_{k|t}^H \mathbf{R}_t^{1,0}(k,k)\} + (|\widehat{f}_{k|t}|^2 + \sigma^2_{f_k|t})\mathbf{R}_t^{1,1}(k,k). \tag{8.54}$$

In Algorithm 1, we describe the GSAVE-KF algorithm in detail.

8.7 VB-KF for Diagonal AR(1) (DAR(1))

In this section, we treat the components of the state $\mathbf{x}_t$ jointly, with all the hyperparameters λ_k, f_k, γ assumed to be independent in the q's. So the expressions for the estimates of the hyperparameters can be shown to be the same as in the previous section on SAVE-KF.

8.7.1 DAR(1) Prediction Stage

The prediction about $\mathbf{x}_t$ can be computed from the time update equation of the standard Kalman filter, $\mathbf{x}_t = \widehat{\mathbf{F}}_{|t-1}\mathbf{x}_{t-1|t-1} + \widetilde{\mathbf{F}}_{|t-1}\mathbf{x}_{t-1|t-1} + \mathbf{v}_t$, $\mathbf{F} = \widehat{\mathbf{F}}_{|t-1} + \widehat{\mathbf{F}}_{|t-1}$, where $\widehat{\mathbf{F}}_{|t-1} = \mathrm{diag}\,(\widehat{f}_{1|t-1},...,\widehat{f}_{M|t-1})$. We also define $\widehat{\mathbf{\Lambda}}_{|t-1} = \mathrm{diag}\,(\frac{1}{\widehat{\lambda}_{1|t-1}},...,\frac{1}{\widehat{\lambda}_{M|t-1}})$. Substituting $\mathbf{x}_{t-1|t-1} = \widehat{\mathbf{x}}_{t-1|t-1} + \widetilde{\mathbf{x}}_{t-1|t-1}$

$$\begin{aligned}
&\widehat{\mathbf{x}}_{t|t-1} = \widehat{\mathbf{F}}_{|t-1}\widehat{\mathbf{x}}_{t-1|t-1},\\
&\widetilde{\mathbf{x}}_{t|t-1} = \widehat{\mathbf{F}}_{|t-1}\widetilde{\mathbf{x}}_{t-1|t-1} + \widetilde{\mathbf{F}}_{|t-1}\mathbf{x}_{t-1|t-1} + \mathbf{w}_t, \Longrightarrow\\
&\widehat{\mathbf{P}}_{t|t-1} = \widehat{F}_{|t-1}\widehat{\mathbf{P}}_{t-1|t-1}\widehat{F}_{|t-1}^H + \widehat{\mathbf{P}}_{\mathbf{F}|t-1}\,\mathrm{diag}\,(\widehat{\mathbf{x}}_{t-1|t-1}\widehat{\mathbf{x}}_{t-1|t-1}^H + \widehat{\mathbf{P}}_{t-1|t-1}) + \widehat{\mathbf{\Lambda}}_{|t-1}.
\end{aligned} \tag{8.55}$$

Algorithm 14: The GSAVE-KF Algorithm

Given: $\mathbf{A}_t, \mathbf{y}_t, N, M, \lambda_{k|0} = a/b\ \forall k, \gamma_0 = c/d, \sigma^2_{k,0|0} = 0, \widehat{x}_{k,0|0} = 0\ \forall k, t > 0.$

Prediction Stage

$\sigma^2_{k,t|t-1} = (|\widehat{f}_{k|t-1}|^2 + \sigma^2_{f_{k|t-1}})\sigma^2_{k,t-1|t-1} + \frac{1}{\widehat{\lambda}_{k|t-1}},\ \widehat{x}_{k,t|t-1} = \widehat{f}_{k|t-1}\widehat{x}_{k,t-1|t-1},$

Update Stage

Initialization: $\sigma^{2,(0)}_{k,t|t} = \sigma^{2,(0)}_{k,t|t-1}, \widehat{x}^{(0)}_{t,\overline{k}|t} = \widehat{x}_{t,\overline{k}|t-1}$

for $i = 1, ...$until convergence

$\sigma^{2,(i)}_{k,t|t} = \sigma^{2,(i-1)}_{k,t|t}(\sigma^{2,(i-1)}_{k,t|t}\widehat{\gamma}_{t-1}\left\|\mathbf{A}_{t,k}\right\|^2 + 1)^{-1}$, Kalman Gain: $\mathbf{K}_{k,t} = \sigma^{2,(i)}_{k,t|t}\mathbf{A}^H_{t,k}\widehat{\gamma}_{t-1}$,

$\widehat{x}^{(i)}_{k,t|t} = \frac{\sigma^{2,(i)}_{k,t|t}}{\sigma^2_{k,t|t-1}}\widehat{x}_{k,t|t-1} + \mathbf{K}_{k,t}\left(\mathbf{y}_t - \mathbf{A}_{t,\overline{k}}\widehat{\mathbf{x}}^{(i-1)}_{t,\overline{k}|t}\right),$

end for

Smoothing Stage

Initialization: $\widehat{\mathbf{P}}^{(0)}_{t-1|t} = \widehat{\mathbf{P}}_{t-1|t-1}, \widehat{\mathbf{x}}^{(0)}_{t-1|t} = \widehat{\mathbf{x}}_{t-1|t-1}$

for $i = 1, ...,$ until convergence

$\widehat{\mathbf{P}}^{-(i)}_{t-1|t} = (\widehat{\mathbf{F}}^H_{|t}\mathbf{A}^H_t\widetilde{\mathbf{R}}^{-1}_t\mathbf{A}_t\widehat{\mathbf{F}}_{|t} + \text{diag}\,(\mathbf{A}^H_t\widetilde{\mathbf{R}}^{-1}_t\mathbf{A}_t)\widehat{\mathbf{P}}_{\mathbf{F}|t} + \widehat{\mathbf{P}}^{-(i-1)}_{t-1|t}),$

$\widehat{\mathbf{x}}^{(i)}_{t-1|t} = \widehat{\mathbf{P}}^{(i)}_{t-1|t}(\widehat{\mathbf{P}}^{-1}_{t-1|t-1}\widehat{\mathbf{x}}^{(i-1)}_{t-1|t} + \widehat{\mathbf{F}}^H\mathbf{A}^H_t\widetilde{\mathbf{R}}^{-1}_t\mathbf{y}_t).$

end for

Estimation of Hyperparameters

Compute $\zeta_t, \mathbf{R}^{m,n}_t$ from (8.49), (8.51).

$\sigma^2_{f_k|t} = \frac{1}{\lambda_k \mathbf{R}^{1,1}_t(k,k)},\ \widehat{f}_{k|t} = \frac{\mathbf{R}^{1,0}_t(k,k)}{\mathbf{R}^{1,1}_t(k,k)}.$

$\widehat{\gamma}_t = \frac{c+\frac{N}{2}}{(\zeta_t + d)},\ \widehat{\lambda}_{k|t} = \frac{a+1}{(\mathbf{R}^{0,0}_t(k,k) - 2\Re\{\widehat{f}^H_{k|t}\mathbf{R}^{1,0}_t(k,k)\} + (|\widehat{f}_{k|t}|^2 + \sigma^2_{f_k|t})\mathbf{R}^{1,1}_t(k,k) + b)}.$

8.7.2 Measurement or Update Stage

Using (8.18),

$$\begin{aligned}\ln q_{\mathbf{x}_t}(\mathbf{x}_t) &= -<\gamma>\left\{-\mathbf{y}^H_t\mathbf{A}_t\mathbf{x}_t - \mathbf{x}^H_t\mathbf{A}^H_t\mathbf{y}_t + \mathbf{x}^H_t\mathbf{A}^H_t\mathbf{A}_t\mathbf{x}_t\right\} - \mathbf{x}^H_t\widehat{\mathbf{P}}^{-1}_{t|t-1}\mathbf{x}_t \\ &\quad + \mathbf{x}^H_t\widehat{\mathbf{P}}^{-1}_{t|t-1}\widehat{\mathbf{x}}_{t|t-1} + \widehat{\mathbf{x}}^H_{t|t-1}\widehat{\mathbf{P}}^{-1}_{t|t-1}\mathbf{x}_t + c_{x_t} \\ &= -(\mathbf{x}_t - \widehat{\mathbf{x}}_{t|t})^H\widehat{\mathbf{P}}^{-1}_{t|t}(\mathbf{x}_t - \widehat{\mathbf{x}}_{t|t}) + c'_{x_t},\end{aligned} \tag{8.56}$$

where the mean and variance are written as

$$\begin{aligned}\widehat{\mathbf{P}}^{-1}_{t|t} &= <\gamma>\mathbf{A}^H_t\mathbf{A}_t + \widehat{\mathbf{P}}^{-1}_{t|t-1}, \\ \widehat{\mathbf{x}}_{t|t} &= \widehat{\mathbf{P}}_{t|t}(<\gamma>\mathbf{A}^H_t\mathbf{y}_t + \widehat{\mathbf{P}}^{-1}_{t|t-1}\widehat{\mathbf{x}}_{t|t-1}).\end{aligned} \tag{8.57}$$

8.7.3 Fixed Lag Smoothing

The posterior distribution $p(\mathbf{x}_{t-1}/\mathbf{y}_{1:t-1})$ is approximated using variational approximation as $q(\mathbf{x}_{t-1}/\mathbf{y}_{1:t-1})$ with mean and covariance as $\widehat{\mathbf{x}}_{t-1|t-1}$ and $\widehat{\mathbf{P}}_{t-1|t-1}$.

(8.58)
$$\ln p(\mathbf{y}_t, \mathbf{x}_{t-1}, \boldsymbol{\theta}/\mathbf{y}_{1:t-1}) = \frac{-1}{2}\ln\det\widetilde{\mathbf{R}}_t - (\mathbf{y}_t - \mathbf{A}_t\mathbf{F}\mathbf{x}_{t-1})\widetilde{\mathbf{R}}_t^{-1}(\mathbf{y}_t - \mathbf{A}_t\mathbf{F}\mathbf{x}_{t-1}) - \frac{1}{2}\det(\widehat{\mathbf{P}}_{t-1|t-1}) - \left(\mathbf{x}_{t-1} - \widehat{\mathbf{x}}_{t-1|t-1}\right)^H \widehat{\mathbf{P}}_{t-1|t-1}^{-1}\left(\mathbf{x}_{t-1} - \widehat{\mathbf{x}}_{t-1|t-1}\right) + c_{x-1},$$

Prediction of $\mathbf{x}_{t-1}$: Using (8.18), $\ln q_{\mathbf{x}_{t-1}}(\mathbf{x}_{t-1}/\mathbf{y}_{1:t})$ turns out to be quadratic in $\mathbf{x}_{t-1}$ and thus can be represented as a Gaussian distribution with mean and covariance as $\widehat{\mathbf{x}}_{t-1|t}$ and $\widehat{\mathbf{P}}_{t-1|t}$ respectively

(8.59)
$$\begin{aligned}\widehat{\mathbf{P}}_{t-1|t}^{-(i)} &= (\widehat{\mathbf{F}}_{|t}^H\mathbf{A}_t^H\widetilde{\mathbf{R}}_t^{-1}\mathbf{A}_t\widehat{\mathbf{F}}_{|t} + \mathrm{diag}\,(\mathbf{A}_t^H\widetilde{\mathbf{R}}_t^{-1}\mathbf{A}_t)\widehat{\mathbf{P}}_{\mathbf{F}|t} + \widehat{\mathbf{P}}_{t-1|t}^{-(i-1)})^{-1},\\ \widehat{\mathbf{x}}_{t-1|t}^{(i)} &= \widehat{\mathbf{P}}_{t-1|t}^{(i)}(\widehat{\mathbf{P}}_{t-1|t-1}^{-1}\widehat{\mathbf{x}}_{t-1|t}^{(i-1)} + \widehat{\mathbf{F}}_{|t}^H\mathbf{A}_t^H\widetilde{\mathbf{R}}_t^{-1}\mathbf{y}_t).\end{aligned}$$

8.7.4 Simulation Results

For the observation model, $\mathbf{y}_t$ is of dimension 100×1 and $\mathbf{x}_t$ is of size 200×1 with 30 non-zero elements. All signals are considered to be real in the simulation. All the elements of $\mathbf{A}_t$ (time varying) are generated i.i.d. from a Gaussian distribution with mean 0 and variance 1. The rows of $\mathbf{A}_t$ are scaled by $\sqrt{30}$ so that the signal part of any scalar observation has unit variance. Taking the SNR to be 20dB, the variance of each element of $\mathbf{v}_t$ (Gaussian with mean 0) is computed as 0.01.

Consider the state update, $\mathbf{x}_t = \mathbf{F}\mathbf{x}_{t-1} + \mathbf{w}_t$. To generate $\mathbf{x}_0$, the first 30 elements are chosen as Gaussian (mean 0 and variance 1) and then the remaining elements of the vector $\mathbf{x}_0$ are put to zero. Then the elements of $\mathbf{x}_0$ are randomly permuted to distribute the 30 non-zero elements across the whole vector. The diagonal elements of $\mathbf{F}$ are chosen uniformly in $[0.9, 1)$. Then the covariance of $\mathbf{w}_t$ can be computed as $\boldsymbol{\Lambda}(\mathbf{I} - \mathbf{F}\mathbf{F}^H)$. Note that $\boldsymbol{\Lambda}$ contains the variances of the elements of $\mathbf{x}_t$ (including $t = 0$), where for the non-zero elements of $\mathbf{x}_0$ the variance is 1 and for the zero elements it is 0. In Fig. 8.4, the blue curve corresponds to the case of a standard Kalman Filter with known state-space model parameters. The red curve corresponds to GSAVE-KF with again all these hyperparameters known. The green curve corresponds to the case of GSAVE-KF with all the hyperparameters also estimated with lag-1 smoothing. Further, we show that filtering for AR(1) coefficients (black curve) does not converge to the basic KF. NMSE is the normalized mean squared error at time t computed as $||\mathbf{x}_t - \widehat{\mathbf{x}}_t||^2$, averaged over 100 different realizations of $\mathbf{A}_t$, $\mathbf{F}$, and of course the noise realizations. The simulations show that in the scenario considered, GSAVE-KF exhibits hardly any MSE degradation over the more complex standard Kalman Filter in steady-state, but takes time to reach steady-state. Adding the estimation of the parameters leads to further slight degradations in steady-state and transient.

Concluding Remarks on GSAVE-KF and Joint VB-KF 9

- We presented a fast SBL algorithm called GSAVE-KF, which uses the variational inference techniques to approximate the posteriors of the data and parameters and track a time varying sparse signal.

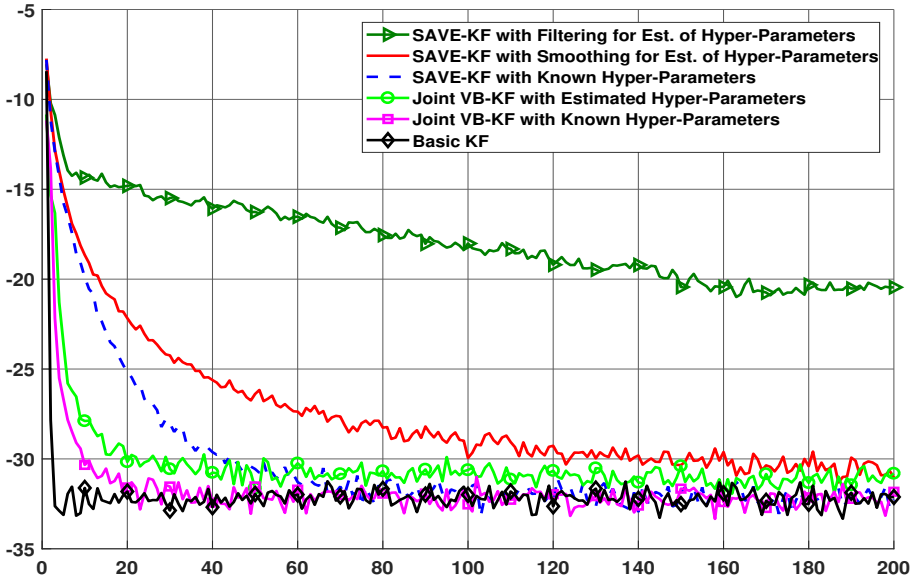


Figure 8.4: NMSE as a function of time (i.e. number of measurements or iteration index).

Concluding Remarks on GSAVE-KF and Joint VB-KF 9 (cont.)

- GSAVE-KF helps to circumvent the matrix inversion operation required in conventional SBL using the EM algorithm.
- We showed that in spite of the significantly reduced computational complexity, the proposed algorithm with estimation of the unknown model parameters has similar steady-state performance compared to the standard Kalman filter, at the price of a significantly increased transient.
- Joint VB-KF has better steady state performance compared to SAVE-KF. Moreover, with estimated hyperparameters, both versions of the proposed algorithm slightly degrades in performance compared to the case when hyperparameters are known.
- One open issue worth exploring would be to compare the MSE performance when hyperparameters are estimated using Type-II ML (which corresponds to empirical Bayes).

Chapter 9

SPARSE BAYESIAN LEARNING USING MESSAGE PASSING ALGORITHMS

In the previous chapter, we looked at mean field (MF) approximation based low complexity solutions for SBL. However, we observed that the predicted posterior variance by SAVE is incorrect compared to the LMMSE posterior covariance. Hence, it makes sense to look at other better variational approximations than MF and this in turn inspired us to look at other alternatives which are the focus of this chapter. In [164], they introduce a BP based SBL algorithm which is more computationally efficient than the original algorithm. The authors use BP to infer the posterior pdf of $\mathbf{x}$ and the hyperparameters are estimated using the EM algorithm. The authors in [179] propose a message passing (MP) approach for inferring the posteriors combining BP and mean field (MF) approximations. MF is a special case of Variational Bayes (VB) in which the partitioning of variables is pushed to the scalar granularity. The advantages of the MF approach are that it always admits a convergent implementation while BP yields a good approximation of the posterior marginals if the factor graph has no cycles. The authors show that the MP fixed-point equations for a combination of BP and the MF approximation correspond to stationary points of one single constrained region-based free energy approximation and provide a clear rule stating how to couple the messages propagating in the BP and MF part. Hence, it is advantageous to apply BP and the MF approximation on the same factor graph in such a combination that their respective virtues can be exploited while overcoming their drawbacks (complexity for BP, potential suboptimality for MF). However, [6] does not treat at all the topic of how to split nodes between BP and MF. We also note that the approximate message passing algorithms [142, 143] suffer from the limitation that the large system limits assume i.i.d. Gaussian or right rotationally invariant $\mathbf{A}^{(t)}$, and the algorithms may exhibit convergence problems.

Another point worth mentioning here is that, combining BP on one hand and MF (or other variations such as EM) on the other hand has a long history since example [180]. Now, in [180] and in all other papers since then on BP-MF combinations, including the example in [181], the application is joint iterative detection and channel estimation with invariably BP being applied to the detection part (with discrete variables) and MF or EM to the estimation part (continuous variables).

9.0.1 Summary of this Chapter

- SBL Space Alternating Variational Estimation (SAVE) provides (largely) underestimated variance estimates. AMP style algorithms may provide more accurate variance information. The existing State Evolution analysis of xAMP (variants of AMP) may show conver-

gence of the (sum) MSE to the MMSE value. In this chapter, xAMP refers to AMP or its variants. But we are interested also in the MSE of the individual components.

- We propose new low complexity SBL algorithms for the static and dynamic cases based on message passing algorithms, with joint hyperparameter estimation.
- Building on the framework of [179], we combine BP and MF approximations in such a way as to optimize the message passing framework, unlike most of the existing applications of the framework, which apply BP and MF to the variable subsets with discrete and continuous distributions resp.
- Using Fisher Information Matrix (FIM) analysis, we propose an optimal partitioning of the unknown parameters in the factor graph such that we can combine BP and (EP) VB in an efficient way, with low complexity and no suboptimality in terms of Laplace approximation (FIM).
- Various new algorithms in this chapter are an application of these parameter partitioning and BP/VB split guidelines. For both a static (classic) compressed sensing model or a dynamic case with autoregressive evolution of the unknown x (corresponding to a classical linear state-space model apart from sparsity considerations). We furthermore show in Lemma 1, in another application of the FIM analysis, that identifiability of the hyperparameters (state space model parameters) requires smoothing (filtering is not sufficient). Although (regardless of sparsity) KF with joint parameter estimation has been the subject of many approaches over decades, this smoothing requirement has never been pointed out or certainly not been analyzed before.
- However, we note that our BP or SAVE based algorithms may not be robust to general $\mathbf{A}$ matrices, similar to SBL solutions based on xAMP algorithms. This leads to the motivation behind the derivation of generalized SwAMP (GSwAMP) SBL in §9.8, which indeed is an extension of the SwAMP to more general priors [182]. In our numerical simulations, we did not observe any divergence for GSwAMP for deviation from i.i.d $\mathbf{A}$, such as ill-conditioned or rank deficient or non-zero mean cases.

9.1 Approximate Inference Cost Functions: An Overview

Maximum likelihood (ML) can be interpreted as the minimization of KLD of $p_{\mathbf{y}}(\mathbf{y}|\theta)$ to empirical distribution of $\mathbf{y}$ ($p_{\mathbf{y}}(y)=\delta(\mathbf{y}-y)$)

$$\begin{aligned}\boldsymbol{\theta}_{min,KL} &= \arg\min_{\boldsymbol{\theta}} D_{KL}(p_{\mathbf{y}}(\mathbf{y})||p_{\mathbf{y}}(\mathbf{y}|\theta)) \\ &= \arg\max_{\boldsymbol{\theta}} \ln(p_{\mathbf{y}}(y|\boldsymbol{\theta})) = \boldsymbol{\theta}_{MLE}.\end{aligned} \tag{9.1}$$

As noted before, VB minimizes KLD of factored approximate posterior ($q(\boldsymbol{\theta}) = \prod_i q_{\boldsymbol{\theta}_i}(\boldsymbol{\theta}_i)$).

$$KLD_{VB} = D_{KL}\big(q(\boldsymbol{\theta})||p(\boldsymbol{\theta}|\mathbf{y})\big). \tag{9.2}$$

However, it can be shown that VB can be reformulated as the minimization of Variational Free Energy (VFE) ($U(q)$ = Average System Energy, $H(q)$ = Entropy), which is defined as follows. As-

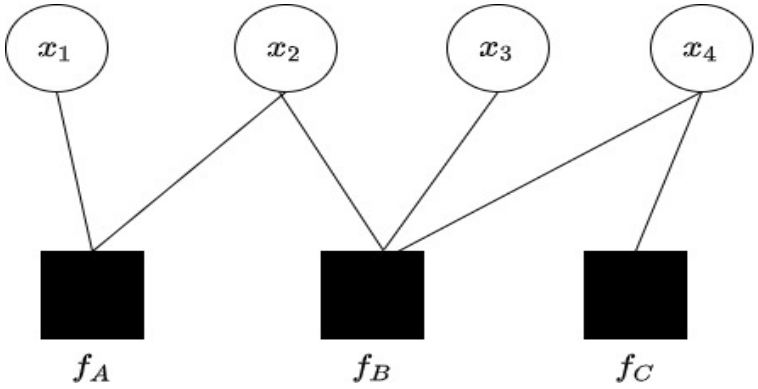


Figure 9.1: A small factor graph representing the posterior $p(x_1, x_2, x_3, x_4) = \frac{1}{Z} f_A(x_1, x_2) f_B(x_2, x_3, x_4) f_C(x_4)$.

sume actual posterior $p(\boldsymbol{\theta}|\mathbf{y}) = \frac{p(\boldsymbol{\theta},\mathbf{y})}{p(\mathbf{y})} = \frac{\prod_a p_a(\boldsymbol{\theta}_a)}{Z}$ and $F_H = -\ln Z$ (Helmholtz Free Energy or log-partition function).

$$
\begin{aligned}
F(q(\boldsymbol{\theta})) &= D_{KL}(q(\boldsymbol{\theta})||p(\boldsymbol{\theta}|\mathbf{y})) + F_H \\
&= \underbrace{-\sum_{\boldsymbol{\theta}} q(\boldsymbol{\theta}) \sum_a ln p_a(\boldsymbol{\theta}_a)}_{U(q)} + \underbrace{\sum_{\boldsymbol{\theta}} q(\boldsymbol{\theta}) \ln q(\boldsymbol{\theta})}_{-H(q)} \\
&= D_{KL}(q(\boldsymbol{\theta})||\prod_a p_a(\boldsymbol{\theta}_a)).
\end{aligned} \tag{9.3}
$$

From (9.3), it is clear that $F(q) \geq F_H$, with equality only if $q(\boldsymbol{\theta}) = p(\boldsymbol{\theta}|\mathbf{y})$. It may not be feasible to directly minimize the VFE at all times due to the complexity associated with the actual posterior factorization. A more practical approach would be upper bound F_H by minimizing $F(q)$ over a restricted class of probability distributions leading to Kikuchi, BP or MF approximations. BP minimizes Bethe Free Energy (BFE) while Mean Field (MF) minimizes MFFE (MF Free Energy). BP converges to exact posterior when the factor graph is a tree. For MF (VB pushed to scalar level), $q(\boldsymbol{\theta}) = \prod_{i=1}^{M} q_{\boldsymbol{\theta}_i}(\boldsymbol{\theta}_i)$. In general, it can be conjectured that the following inequality holds

$$MFFE \geq BFE \geq VFE.$$

Note that in [183], the authors consider region based Free Energy approximations (RFE). The intuitive idea behind a RFE approximation is to break up the factor graph into a set of large regions that include every factor and variable node, and say that the overall free energy is the sum of the free energies of all the regions. BP is a special case of this. Expectation Propagation (EP) can be derived using BFE under moment matching constraints.

9.1.1 Region Based Free Energy

A region R of a factor graph to be a set $\mathcal{V}_R$ of variable nodes and set $\mathcal{A}_R$ of factor nodes, such that $a \in \mathcal{A}_R \implies$ all variable nodes connected to a are in $\mathcal{V}_R$. θ_R is defined as the set of all variable nodes belonging to the region R. Region energy is defined as $E_R(\boldsymbol{\theta}_R) = -\sum_{a \in \mathcal{A}_R} \ln p_a(\boldsymbol{\theta}_a)$. Region

based free energy can be expresed using region entropy and region average energy

$$\begin{aligned} U_R(q_R) &= \sum_{\boldsymbol{\theta}_R} q_R(\boldsymbol{\theta}_R) E_R(\boldsymbol{\theta}_R), \\ H_R(q_R) &= \sum_{\boldsymbol{\theta}_R} q_R(\boldsymbol{\theta}_R) \ln q_R(\boldsymbol{\theta}_R). \\ \text{and } \; F_R(q_R) &= U_R(q_R) - H_R(q_R). \end{aligned} \tag{9.4}$$

Region-based free energy using region-based entropy and region-based average energy

$$\begin{aligned} U_{\mathscr{R}}(\{q_R\}) &= \sum_{R\in\mathscr{R}} c_R U_R(q_R), \; H_{\mathscr{R}}(\{q_R\}) \\ &= \sum_{R\in\mathscr{R}} c_R H_R(q_R). \\ \text{and } \; F_{\mathscr{R}}(\{q_R\}) &= U_{\mathscr{R}}(\{q_R\}) - H_{\mathscr{R}}(\{q_R\}). \end{aligned} \tag{9.5}$$

The intuitive idea is to break up the factor graph into a set of large regions that include every factor and variable node, and say that the overall VFE is the sum of the VFEs of all the regions. If some of the large regions overlap, then we will have erred by counting the free energy contributed by some nodes two or more times, so we then need to subtract out the free energies of these overlap regions in such a way that each factor and variable node is counted exactly once (weight c_R takes care of this). In BP, each factor node (and its neighbouring variable nodes) form one set of regions. Another set of regions which contain only one variable node.

9.1.2 Combined BP/MF Approximation

The fixed points of the standard BP algorithm are shown to be the stationary points of the Bethe free energy (BFE) [179]. However, for the MF approximation in VB, the approximate posteriors are shown to be converging to a local minimum of the MF free energy which is an approximation of the BFE. Moreover, we observe in [168, 184] that for estimation of the signals from interference corrupted observations, MF is a poor choice since it does not give the accurate posterior variance (posterior variance of x_i is observed to be independent of the error variances of other $x_l, l \neq i$). Assume that the posterior be represented as, $p(\boldsymbol{\theta}) = \frac{1}{Z} \prod_{a\in\mathscr{A}_{BP}} f_a(\boldsymbol{\theta}_a) \prod_{b\in\mathscr{A}_{MF}} f_b(\boldsymbol{\theta}_b)$, where $\mathscr{A}_{BP}, \mathscr{A}_{MF}$ represent the set of nodes belonging to the BP part and MF part respectively with $\mathscr{A}_{BP} \cap \mathscr{A}_{MF} = \emptyset$. Z represents the normalization variable. Throughout the chapter, the vector $\boldsymbol{\theta}_i$ represents a subset of $\boldsymbol{\theta}$ and θ_i represents a scalar parameter in $\boldsymbol{\theta}$. $\mathscr{N}(i), \mathscr{N}(a)$ represent the number of neighbouring nodes of any variable node i or factor node a. $\mathscr{N}_{BP}(i)$ represents the number of neighbouring nodes of i which belong to the BP part, similarly $\mathscr{N}_{MF}(i)$ is defined. Also, we define $\mathscr{I}_{MF} = \bigcup_{a\in\mathscr{A}_{MF}} \mathscr{N}(a), \mathscr{I}_{BP} = \bigcup_{a\in\mathscr{A}_{BP}} \mathscr{N}(a)$. The resulting free energy obtained by the combination of BP and MF are written as below (Note that we use an abuse of notation and let $q_i(\theta_i)$ represents the belief about θ_i (the approximate posterior))

$$\begin{aligned} F_{BP,MF} = &\sum_{a\in\mathscr{A}_{BP}} \sum_{\boldsymbol{\theta}_a} q_a(\boldsymbol{\theta}_a) \ln \frac{q_a(\boldsymbol{\theta}_a)}{f_a(\boldsymbol{\theta}_a)} - \sum_{a\in\mathscr{A}_{MF}} \sum_{\mathbf{x}_a} \prod_{i\in\mathscr{N}(a)} q_i(\theta_i) \ln f_a(\boldsymbol{\theta}_a) \\ &- \sum_{i\in\mathscr{I}} (|\mathscr{N}_{BP}(i)| - 1) \sum_{\theta_i} q_i(\theta_i) \ln q_i(\theta_i). \end{aligned} \tag{9.6}$$

The beliefs have to satisfy the following normalization and marginalization constraints

$$\begin{aligned}
&\sum_{\theta_i} q_i(\theta_i) = 1, \forall\, i \in \mathscr{I}_{MF} \setminus \mathscr{I}_{BP}, \\
&\sum_{\boldsymbol{\theta}_a} q_a(\boldsymbol{\theta}_a) = 1, \forall\, a \in \mathscr{A}_{BP}, \\
&q_i(\theta_i) = \sum_{\boldsymbol{\theta}_a \setminus \theta_i} q_a(\boldsymbol{\theta}_a),\ \forall\, a \in \mathscr{A}_{BP}, i \in \mathscr{N}(a).
\end{aligned} \tag{9.7}$$

Let $m_{a\to i}$ represents the message passed from any factor node a to variable node i and $n_{i\to a}$ represents the message passed from any variable node i to factor node a. The fixed point equations corresponding to the constrained optimization of (9.6) can be written as follows [179]

$$\begin{aligned}
q_i(\theta_i) &= z_i \prod_{a \in \mathscr{N}_{BP}(i)} m^{BP}_{a\to i}(\theta_i) \prod_{a \in \mathscr{N}_{MF}(i)} m^{MF}_{a\to i}(\theta_i), \\
n_{i\to a}(\theta_i) &= \prod_{a \in \mathscr{N}_{BP}(i)\setminus a} m_{a\to i}(\theta_i) \prod_{a \in \mathscr{N}_{MF}(i)} m_{a\to i}(\theta_i), \\
m^{MF}_{a\to i}(\theta_i) &= \exp(< \ln f_a(\theta_a) >_{\prod_{j \in \mathscr{N}(a)\setminus i} n_{j\to a}(\theta_j)}), \\
m^{BP}_{a\to i}(\theta_i) &= \Big(\int \prod_{j \in \mathscr{N}(a)\setminus i} n_{j\to a}(\theta_j) f_a(\theta_a) \prod_{j \neq i} d\theta_j\Big),
\end{aligned} \tag{9.8}$$

where $<>_q$ represents the expectation w.r.t distribution q.

9.1.2.1 What do the MP Expressions Indicate?

BP-MF combo can be written as the alternating optimization of Lagrangian [183]:

$$\mathscr{L} = F_{BP,MF} + \sum_a \gamma_a [\sum_{\boldsymbol{\theta}_a} q_a(\boldsymbol{\theta}_a) - 1] + \sum_i \gamma_i [\sum_{\theta_i} q_i(\theta_i) - 1] + \sum_i \sum_{a \in \mathscr{N}(i)} \sum_{\theta_i} \lambda_{ai}(\theta_i)[q_i(\theta_i) - \sum_{\boldsymbol{\theta}_a \setminus \theta_i} q_a(\boldsymbol{\theta}_a)]. \tag{9.9}$$

At any iteration or convergence, the following fixed points can be obtained.

$$\begin{aligned}
q_a(\boldsymbol{\theta}_a) &= p_a(\boldsymbol{\theta}_a)\Big(\prod_{i \in \mathscr{N}(a)} q_i(\theta_i) \exp\left[-\lambda_{ai}(\theta_i)\right]\Big) \exp\left[\gamma_a - 1\right] \\
&= \frac{1}{z_a} p_a(\boldsymbol{\theta}_a) \prod_{i \in \mathscr{N}(a)} \underbrace{\frac{q_i(\theta_i)}{m_{a\to i}(\theta_i)}}_{n_{i\to a}(\theta_i)},\ a \in \mathscr{A}_{BP}
\end{aligned} \tag{9.10}$$

$$q_i(\theta_i) = \underbrace{\exp\left[|\mathscr{N}_{BP}(i)| - 1 + \mathrm{I}_{\mathscr{I}_{MF}\setminus\mathscr{I}_{BP}}(i)\,\gamma_i\right]}_{1/z_i} \prod_{a \in \mathscr{N}_{MF}(i)} \underbrace{exp(< \ln p_a(\boldsymbol{\theta}_a) >_{q_j(\theta_j), j \in \mathscr{N}(a)\setminus i})}_{m^{MF}_{a\to i}(\theta_i)} \prod_{a \in \mathscr{N}_{BP}(i)} \underbrace{\exp(\lambda_{ai}(\theta_i))}_{m^{BP}_{a\to i}(\theta_i)}.$$

where $\mathrm{I}_{\mathscr{A}}(i)$ = indicator function for $i \in \mathscr{A}$. Applying the marginalization constraint $q_i(\theta_i) = \sum_{\boldsymbol{\theta}_a \setminus \theta_i} q_a(\boldsymbol{\theta}_a),\ \forall\, a \in \mathscr{A}_{BP}$ leads to the expression for $m^{BP}_{a\to i}(\theta_i)$ as in (9.8). The Lagrange multipliers λ_{ai} are indeed the log of the BP messages and γ_a, γ_i lead to the normalization constants z_a, z_i for the beliefs $q_a(\boldsymbol{\theta}_a)$, $q_i(\theta_i)$, respectively.

$$\lambda_{ai}(\boldsymbol{\theta}_i) = \ln m^{BP}_{a\to i}(\theta_i).$$

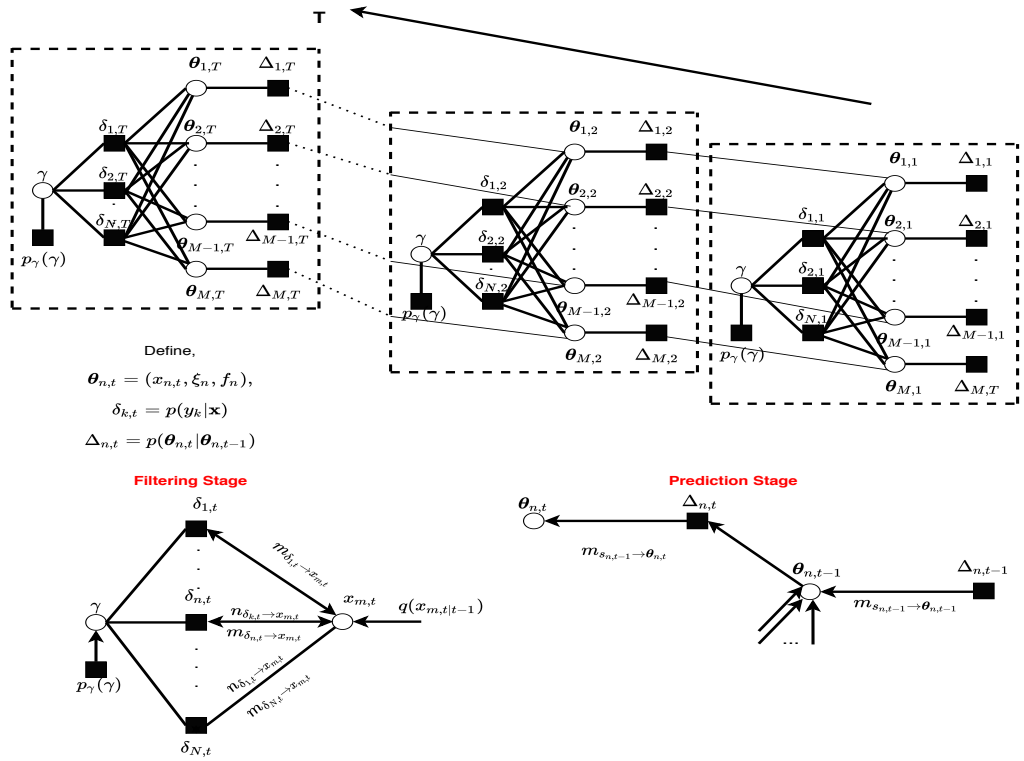


Figure 9.2: Factor Graph for the dynamic SBL. Note that messages from the smoothing stage are not shown here.

9.2 Dynamic SBL System Model

Sparse signal $\mathbf{x}_t$ is modeled using an AR(1) process with a diagonal correlation coefficient matrix $\mathbf{F}$, which can be written as follows

$$\begin{aligned} \text{State Update: } \mathbf{x}_t &= \mathbf{F}\mathbf{x}_{t-1} + \mathbf{w}_t, \\ \text{Observation: } \mathbf{y}_t &= \mathbf{A}^{(t)}\mathbf{x}_t + \mathbf{v}_t, \end{aligned} \tag{9.11}$$

where $\mathbf{x}_t = \left[x_{1,t},\ ...,\ x_{M,t}\right]^T$. Diagonal matrices $\mathbf{F}$ and Ξ are defined with its elements, $\mathbf{F}_{i,i} = f_i, f_i \in (-1,1)$ and $\Xi = \text{diag}(\boldsymbol{\xi}), \boldsymbol{\xi} = [\xi_1, ...\xi_M]$. Here ξ_i represents the inverse variance of $x_{i,t} \sim \mathscr{CN}(0, \frac{1}{\xi_i})$. Further, $\mathbf{w}_t \sim \mathscr{CN}(\mathbf{0}, \boldsymbol{\Lambda}^{-1})$, where $\boldsymbol{\Lambda}^{-1} = \Xi^{-1}(\mathbf{I} - \mathbf{F}\mathbf{F}^H) = \text{diag}(\frac{1}{\lambda_1}, ..., \frac{1}{\lambda_M})$ and $\mathbf{v}_t \sim \mathscr{CN}(\mathbf{0}, \frac{1}{\gamma}\mathbf{I})$. $\mathbf{w}_t$ are the complex Gaussian mutually uncorrelated state innovation sequences. Hence we sparsify the prediction error variance $\mathbf{w}_t$ also, with the same support as $\mathbf{x}_0$ and henceforth enforces the same support set for $\mathbf{x}_t, \forall\, t$. $\mathbf{v}_t$ is independent of the $\mathbf{w}_t$ process. Although the above signal model seems simple, there are numerous applications such as 1) Bayesian adaptive filtering [185], 2) Wireless channel estimation: multipath parameter estimation as in [186]. In this case, $\mathbf{x}_t$ = FIR filter response, and Ξ represents example the power delay profile.

In Bayesian compressive sensing, a two-layer hierarchical prior is assumed for the $\mathbf{x}$ as in [2]. The hierarchical prior is chosen such that it encourages the sparsity property of $\mathbf{x}_t$ or of the innovation sequences $\mathbf{v}_t$. The state update gets represented as

$$p(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{F}, \Xi) = \prod_{i=1}^{M} \mathscr{CN}(f_i x_{i,t-1}, \frac{1}{\xi_i}). \tag{9.12}$$

For the convenience of analysis, we reparameterize ξ_i in terms of λ_i and assume a Gamma prior for $\mathbf{\Lambda}$, $p(\mathbf{\Lambda}) = \prod_{i=1}^{M} p(\lambda_i|a,b) = \prod_{i=1}^{M} \Gamma^{-1}(a) b^a \lambda_i^{a-1} e^{-b\lambda_i}$. The inverse of noise variance γ is also assumed to have a Gamma prior, $p(\gamma|c,d) = \Gamma^{-1}(c) d^c \gamma_i^{c-1} e^{-d\gamma}$, such that the marginal pdf of $\mathbf{x}_t$ (student-t distribution) becomes more sparsity inducing than example, a Laplacian prior. The advantage is that the whole machinery of linear MMSE estimation can be exploited, such as example, the Kalman filter. But this is embedded in other layers making things eventually non-Gaussian. Now the likelihood distribution can be written as

$$p(\mathbf{y}_t|\mathbf{x}_t,\gamma) = (2\pi)^{-N} \gamma^N e^{-\gamma||\mathbf{y}_t - \mathbf{A}^{(t)}\mathbf{x}_t||^2}. \tag{9.13}$$

To make these priors non-informative (Jeffrey's prior), we choose them to be small values $a = c = b = d = 10^{-5}$. For the AR(1) coefficients f_k, we do not assume any prior distribution. We define the unknown parameter vector $\boldsymbol{\theta} = \{\mathbf{x}, \mathbf{\Lambda}, \gamma, \mathbf{F}\}$ and θ_i represents each scalar in $\boldsymbol{\theta}$.

9.2.1 BP-MF based Static SBL

The figure 9.5 represents the factor graph (note that static case is a special case with the state update nodes being not present), where it is divided into two disjoint subsets $\mathscr{A}_{BP} = f_{\delta_{n,t}} \forall n, l, t$ and $\mathscr{A}_{MF}$ represents rest of the factor or variable nodes. To combine BP and MF, we introduce the new variables $h_{n,t} = \mathbf{A}^{(t)}_{n,:}\mathbf{x}_t$, $s_{l,t} = f_l x_{l,t-1}$ and the hard constraint factor nodes

$$\begin{aligned} f_{\delta_{n,t}} &= \delta(h_{n,t} - \mathbf{A}^{(t)}_{n,:}\mathbf{x}_t), \forall n \in [1:N], t, f_{\Delta_{l,t}} \\ &= \delta(s_{l,t} - f_l x_{l,t-1}), \forall l \in [1:M], t. \end{aligned} \tag{9.14}$$

For the static case, the system model will be $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{v}$, so $f_l = 0, \lambda_l = \xi_l, \forall l$. We omit subscript t for simplicity. The message $m_{f_{\delta_n} \to x_l}$ from the hard factor f_{δ_n} to variable node x_l is computed by the BP rule with the incoming messages to the node, $n_{h_n \to f_{\delta_n}}(h_n) = m_{f_{y_n} \to h_n}(h_n)$ and $n_{x_{l'} \to f_{\delta_n}}(x_{l'}), \forall l' \neq l$, later defined in (9.19). So

$$m_{f_{\delta_n \to x_l}}(x_l) = \int f_{\delta_n} n_{h_n \to f_{\delta_n}}(h_n) \prod_{l' \neq l} n_{x_{l'} \to f_{\delta_n}}(x_{l'}) \prod_{l' \neq l} dx_{l'}. \tag{9.15}$$

For notational brevity, we denote subscript (l,n) or (n,l) to represent the messages passed from l to n or viceversa. All the messages (beliefs or continuous pdfs) passed between them can be shown to be Gaussian [164] and thus it suffices to represent them by the mean and variance of the beliefs. With the hard constraints, the equivalent observation model can be written as

$$\begin{aligned} y_n - \textstyle\sum_{l' \neq l}^{M} A_{n,l'} \widehat{x}_{l',n} &= A_{n,l} x_l + \textstyle\sum_{l' \neq l}^{M} A_{n,l'} \widetilde{x}_{l',n} + \upsilon_n, \\ \text{where, } \widetilde{x}_{l',n} &\sim \mathscr{CN}(0, \nu_{l',n}), \text{ and } m_{f_{\delta_n \to x_l}} \propto \mathscr{CN}(\widehat{x}_{n,l}, \nu_{n,l}), \end{aligned} \tag{9.16}$$

We obtain the message, $m_{f_{\delta_n \to x_l}}(x_l) \sim \mathscr{N}(\widehat{x}_{n,l}, \nu_{n,l})$, where the mean and variance of the resulting posterior can be represented as

$$\begin{aligned} \widehat{x}_{n,l} &= A_{n,l}^{-1}(y_n - p_n + A_{n,l}\widehat{x}_{l,n}), \\ p_n &= \textstyle\sum_{l'=1}^{M} A_{n,l'} \widehat{x}_{l',n}, \\ \nu_{n,l} &= |A_{n,l}|^{-2}(\widehat{\gamma}^{-1} + \nu_n - |A_{n,l}|^2 \nu_{l,n}), \\ \nu_n &= \textstyle\sum_{l'=1}^{M} |A_{n,l'}|^2 \nu_{l',n}. \end{aligned} \tag{9.17}$$

We define

$$
\begin{aligned}
d_l &= (\sum_{n=1}^{N} \nu_{n,l}^{-1})^{-1}, \\
r_l &= d_l(\sum_{n=1}^{N} \frac{\widehat{x}_{n,l}}{\nu_{n,l}}).
\end{aligned}
\tag{9.18}
$$

Given the messages, $m_{f_{\delta_n} \to x_l}(x_l)$, the belief $q(x_l)$ can be obtained as $(f_{\lambda_i}(\lambda_i) = p(\lambda_k|a,b))$, $q(x_l) \propto f_{\lambda_i}(\lambda_i) \prod_{n=1}^{N} m_{f_{\delta_n} \to x_l} \propto \mathscr{CN}(\widehat{x}_l, \sigma_l^2)$

$$
\begin{aligned}
\text{where } \sigma_l^{-2} &= \lambda_l + d_l^{-1}, \\
\widehat{x}_l &= \frac{r_l}{1 + d_l \sigma_l^{-2}}.
\end{aligned}
\tag{9.19}
$$

One remark here is that compared to our previous work using VB [168], combining BP and MF gives a more accurate approximation of the error variance as shown in (9.19), where σ_l^2 incorporates the effect of all $\sigma_{l'}^2, l' \neq l$. Since the factor node $f_{\delta_n} \in \mathscr{A}_{BP}$, the message $n_{x_l \to f_{\delta_n}}(x_l)$ from variable node x_l to f_{δ_n} is updated by the BP rule as follows

$$
\begin{aligned}
n_{x_{l,t} \to f_{\delta_n}}(x_l) &= \frac{q(x_l)}{m_{f_{\delta_n \to x_l}}(x_l)} \propto \mathscr{CN}(\widehat{x}_{l,n}, \nu_{l,n}), \\
\text{where, } \nu_{l,n}^{-1} &= (\sigma_l^{-2} - \nu_{n,l}^{-1}), \\
\widehat{x}_{l,n} &= \nu_{l,n}(\frac{\widehat{x}_l}{\sigma_l^2} - \frac{\widehat{x}_{n,l}}{\nu_{n,l}}).
\end{aligned}
\tag{9.20}
$$

9.2.2 Dynamic BP-MF-EP based SBL

The joint distribution of all the observations and parameters can be written as, $p(\mathbf{y}_t, \boldsymbol{\theta}|\mathbf{y}_{1:t-1}) = p(\mathbf{y}_t|\boldsymbol{\theta})\, p(\boldsymbol{\theta}|\mathbf{y}_{1:t-1})$, where $p(\boldsymbol{\theta}|\mathbf{y}_{1:t-1})$ denotes the predictive distribution. Similar as in KF, first we compute the posterior distribution of θ_i given the observations till $(t-1)$, which is called as the prediction stage. Since the correlation coefficient matrix $\mathbf{F}$ is diagonal, all the $x_{i,t}$ are decoupled in the state update model and we exploit this fact to predict the states and the hyperparameters in the state update model using MF.

9.2.2.1 Diagonal AR(1) (DAR(1)) Prediction Stage

Assuming that the belief $q(\gamma)$ at time t, of noise precision γ is known, the message $m_{f_{y_{n,t}} \to h_{n,t}}(h_{n,t})$ from the factor node $f_{y_{n,t}} \in \mathscr{A}_{MF}$ is calculated using the MF rule $m_{f_{y_{n,t}} \to h_{n,t}}(h_{n,t}) = < \exp\left(\ln f_{y_{n,t}}(h_{n,t}, \gamma)\right) >_{q(\gamma)}$, which becomes, $m_{f_{y_{n,t}} \to h_{n,t}}(h_{n,t}) \propto \mathscr{CN}(y_{n,t}, \widehat{\gamma}_t^{-1})$. Here $\widehat{\gamma}_t = < \gamma >_{q(\gamma)}$. For more detailed derivation, we refer to our paper [187]. With the hard constraints $f_{\Delta_{l,t}}$, the equivalent state space model can be re-written as, The joint distribution for the state space model can be $x_{l,t} = \widehat{f}_{l|t-1} x_{l,t-1} + \widetilde{f}_{l|t-1} x_{l,t-1} + w_{l,t}$, with $w_{l,t} \sim \mathscr{CN}(0, \widehat{\lambda}_{k|t}^{-1})$. Here we denote $\widehat{f}_{l|t-1}$ as the estimate of f_l given the observations till $t-1$ and $\widetilde{f}_{l|t-1}$ represents the error in the estimation. Similarly we can represent $x_{l,t-1} = \widehat{x}_{l,t-1|t-1} + \widetilde{x}_{l,t-1|t-1}$, $\widetilde{x}_{l,t-1|t-1}$ being the estimation error. Now the mean and variance of the message passed from $f_{\Delta_{l,t}}$ to the variable node $x_{l,t}$ can be computed as

$$
\begin{aligned}
\widehat{x}_{l,t|t-1} &= \widehat{f}_{l|t-1} \widehat{x}_{l,t-1|t-1}, \\
\sigma_{l,t|t-1}^2 &= |\widehat{f}_{l|t-1}|^2 \sigma_{l,t-1|t-1}^2 + \sigma_{f_l|t-1}^2 (|\widehat{x}_{l,t-1|t-1}|^2 + \sigma_{l,t-1|t-1}^2) + \widehat{\lambda}_{l|t-1}^{-1}.
\end{aligned}
\tag{9.21}
$$

$m_{f_{\Delta_{l,t}} \to x_{l,t}}(x_{l,t})$ is not a tractable distribution and thus using EP [188], we project it into the class of Gaussian distribution (ϕ), where the projection operator can be represented as $\text{Proj}_{\phi}[p] = \arg\min_{q \in \phi} KL(p||q)$. This leads to moment matching (approximated $q \in \mathscr{CN}(\mu, \nu)$ has the same mean and variance as p). So we approximate,
$m_{f_{\Delta_{l,t}} \to x_{l,t}}(x_{l,t}) = q(x_{l,t|t-1}) \approx \mathscr{CN}(\widehat{x}_{l,t|t-1}, \sigma^2_{l,t|t-1})$.

9.2.2.2 Measurement Update Stage

In the measurement update stage, the posterior for $\mathbf{x}_t$ is inferred using BP as in Section 9.2.1.and we represent the messages by $\widehat{x}^{(t)}_{n,l}, \nu^{(t)}_{n,l}$ and the beliefs by $\widehat{x}_{l,t|t}, \sigma^2_{l,t|t}$. In the measurement stage, the prior for $x_{k,t}$ gets replaced by the belief from the prediction stage and thus the term r_l need to be rewritten as, $r_{l,t}$=$d_{l,t}(\sum_{n=1}^{N} \frac{\widehat{x}^{(t)}_{n,l}}{\nu^{(t)}_{n,l}} + \frac{\widehat{x}_{l,t|t-1}}{\sigma^2_{l,t|t-1}})$.

$$
\begin{aligned}
m_{f_{\delta_{n,t}} \to x_{l,t}} &\propto \mathscr{CN}(\widehat{x}^{(t)}_{n,l}, \nu^{(t)}_{n,l}), \text{where,} \\
\widehat{x}^{(t)}_{n,l} &= \frac{y_{n,t} - p_{n,t} + A^{(t)}_{n,l} \widehat{x}^{(t)}_{l,n}}{A^{(t)}_{n,l}}, \\
\nu^{(t)}_{n,l} &= \frac{\widehat{\gamma}_t^{-1} + \nu_{n,t} - |A^{(t)}_{n,l}|^2 \nu^{(t)}_{l,n}}{|A^{(t)}_{n,l}|^2}, \\
p_{n,t} &= \sum_{l=1}^{M} A^{t}_{n,l} \widehat{x}^{(t)}_{l,n}, \\
\nu_{n,t} &= \sum_{l=1}^{M} |A^{(t)}_{n,l}|^2 \nu^{(t)}_{l,n}.
\end{aligned}
\tag{9.22}
$$

Given the messages, $m_{f_{\Delta_{l,t}} \to x_{l,t}}(x_{l,t}) \propto \mathscr{CN}(\widehat{x}_{l,t|t-1}, \sigma^2_{l,t|t-1})$, which comes from the prediction stage, the belief $q(x_{l,t|t}) \propto \mathscr{CN}(\widehat{x}_{l,t|t}, \sigma^2_{l,t|t})$, (the posterior marginal of $x_{l,t}$ given the observations till t) can be obtained as

$$
\begin{aligned}
\sigma^{-2}_{l,t|t} &= \sigma^{-2}_{l,t|t-1} + \sum_{n=1}^{N} \nu^{-(t)}_{n,l}, \\
d_{l,t} &= (\sum_{n=1}^{N} \nu^{-(t)}_{n,l})^{-1}. \\
\widehat{x}_{l,t|t} &= \frac{r_{l,t}}{1 + d_{l,t} \sigma^{-2}_{l,t|t-1}}, \\
r_{l,t} &= d_{l,t} \left(\sum_{n=1}^{N} \frac{\widehat{x}^{(t)}_{n,l}}{\nu^{(t)}_{n,l}} + \frac{\widehat{x}_{l,t|t-1}}{\sigma^2_{l,t|t-1}} \right).
\end{aligned}
\tag{9.23}
$$

Further, we can obtain the message from $x_{l,t}$ to $\delta_{n,t}$ as follows

$$
\begin{aligned}
n_{x_{l,t} \to f_{\delta_{n,t}}}(x_{l,t}) &\propto \mathscr{CN}(\widehat{x}^{(t)}_{l,n}, \nu^{(t)}_{l,n}), \\
\nu^{-(t)}_{l,n} &= \left(\sigma^{-2}_{l,t|t} - \nu^{-(t)}_{n,l} \right), \\
\widehat{x}^{(t)}_{l,n} &= \nu^{(t)}_{l,n} \left(\frac{\widehat{x}_{l,t|t}}{\sigma^2_{l,t|t}} - \frac{\widehat{x}^{(t)}_{n,l}}{\nu^{(t)}_{n,l}} \right).
\end{aligned}
\tag{9.24}
$$

9.2.2.3 Lag-1 Smoothing Stage

We show in Lemma 6 that KF is not enough to adapt the hyperparameters, instead we need to do at least a lag 1 smoothing (i.e. the computation of $\hat{x}_{k,t-1|t}, \sigma^2_{k,t-1|t}$ through BP). All the hyperparameters λ_l, f_l, γ belong to $\mathscr{A}_{MF}$. Note that the notations $\widehat{f}_{k|t}, \widehat{\lambda}_{k|t}, \widehat{\gamma}_t$ refers to mean of the posteriors (which is equal to the LMMSE point estimates) for the respective hyperparameters at time t and $\sigma^2_{f_k|t}$ represents the posterior variance of f_k at time t. For the smoothing stage, we use BP with Gaussian Markov Random Fields (GMRF) based factorization. GMRF refers to the representation of BP [183], when the underlying Gaussian distribution is expressed in terms of pairwise connections between scalar variables $x_{i,t}$. Substituting the state update equation into the observation model (9.11), we obtain the system model for the smoothing stage as follows

$$\begin{aligned} \mathbf{y}_t &= \mathbf{A}^{(t)}\mathbf{F}\mathbf{x}_{t-1} + \widetilde{\mathbf{v}}_t, \text{ where} \\ \widetilde{\mathbf{v}}_t &= \mathbf{A}^{(t)}\mathbf{w}_t + \mathbf{v}_t, \end{aligned} \tag{9.25}$$

where $\widetilde{\mathbf{v}}_t \sim \mathscr{CN}(0, \widetilde{\mathbf{R}}_t)$ with $\widetilde{\mathbf{R}}_t = \mathbf{A}^{(t)}\mathbf{\Lambda}^{-1}\mathbf{A}^{(t)\,H} + \frac{1}{\gamma}\mathbf{I}$. The joint distribution can be factorized as, $p(\mathbf{y}_t, \boldsymbol{\theta}|\mathbf{y}_{1:t-1}) = p(\mathbf{y}_t|\boldsymbol{\theta})\, p(\mathbf{x}_{t-1}|\mathbf{y}_{1:t-1})\, p(\mathbf{F}, \mathbf{\Lambda}, \gamma|\mathbf{y}_{1:t-1})$.

$$\begin{aligned} \ln p(\mathbf{y}_t, \boldsymbol{\theta}|\mathbf{y}_{1:t-1}) = \frac{-1}{2}\ln\det\widetilde{\mathbf{R}}_t - |f_i|^2|x_i|^2\mathbf{A}_i^{(t)\,H}\widetilde{\mathbf{R}}_t^{-1}\mathbf{A}_i^{(t)} \\ + 2\Re(f_i^H x_i^H \mathbf{A}_i^{(t)\,H}\widetilde{\mathbf{R}}_t^{-1}(\mathbf{y}_t - \mathbf{A}_{\bar{i}}^{(t)}\mathbf{F}_{\bar{i}}\mathbf{x}_{\bar{i},t})) + c_f, \end{aligned} \tag{9.26}$$

where c_f being the terms independent of f_i, $\mathbf{A}_{\bar{i}}^{(t)}, \mathbf{x}_{\bar{i},t}$ represents the matrix $\mathbf{A}^{(t)}$ or the vector $\mathbf{x}_t$ with i^{th} column or element removed. Note that we propose to compute $\widetilde{\mathbf{R}}_t$ by substituting the point estimates of $\mathbf{\Lambda}, \gamma$. We also define $\widehat{\mathbf{F}}_{\bar{i}|t} = \text{diag}(\widehat{f}_{j|t}, j \neq i)$ with i^{th} element removed. Further applying the MF rule from (9.8), we write the mean and variance of the resulting Gaussian distribution as

$$\begin{aligned} \sigma^{-2}_{f_i|t} &= (|\widehat{x}_{i,t-1|t}|^2 + \sigma^2_{i,t-1|t})\mathbf{A}_i^{(t)\,H}\widetilde{\mathbf{R}}_t^{-1}\mathbf{A}_i^{(t)}, \\ \widehat{f}_{i|t} &= \sigma^2_{f_i|t}\widehat{x}^H_{i,t-1|t}\mathbf{A}_i^{(t)\,H}\widetilde{\mathbf{R}}_t^{-1}(\mathbf{y}_t - \mathbf{A}_{\bar{i}}^{(t)}\widehat{\mathbf{F}}_{\bar{i}|t}\widehat{\mathbf{x}}_{\bar{i},t-1|t}). \end{aligned} \tag{9.27}$$

The entire algorithm (a combination of BP, MF and EP, we call it as Combined BP-MF-EP DAR-SBL) is described in Algorithm 15. Also we remark that for the estimation of λ_k, γ, we follow the same approach as in our paper [187] and we refer to it for more details. One remark here is that another version called as Combined Vector BP-MF-EP DAR-SBL follows immediately from the derivations for Algorithm 15, where all the components of $\mathbf{x}_t$ are considered jointly in the factor graph. Even though the performance will be higher (as observed in the simulations) for the vector case, it comes at the cost of a higher complexity due to the matrix inversion involved. Note that in Algorithm 15, we introduce temporal averaging for certain quantities (represented by $<>_{|t}$) in hyperparameter estimates and β being the temporal weighting coefficient which is less than one, see [187] for more details.

9.3 Optimal Partitioning of BP and MF nodes

In this section, we show that the partitioning of BP and MF nodes can be characterized through the computation of FIM = $\text{E}(\frac{\partial \ln p(\mathbf{y},\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial \ln p(\mathbf{y},\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}^H)$. For our analysis, we will allude briefly to an extended concept of Cramer-Rao bound (CRB), the mismatched CRB ($mCRB$) [189] of VB ($mCRB_{VB}$), which is a version of the CRB under model misspecification, and corresponds to the Laplace approximation covariance. Let CRB corresponds to the proper Bayesian CRB and $mCRB_{BP}$ refers to the $mCRB$ for the BP.

Algorithm 15: Combined BP-MF-EP DAR-SBL

Initialization $\widehat{f}_{l|0}, \widehat{\lambda}_{l|0} = \frac{a}{b}, \widehat{\gamma}_0 = \frac{c}{d}, \widehat{x}_{l,0|0} = 0, \sigma^2_{l,0|0} = 0, \forall l$. Define $\mathbf{\Sigma}_{t-1|t-1} = \text{diag}(\sigma^2_{l,t|t-1})$.
for $t = 1 : T$ do
Prediction Stage:

1. Compute $\widehat{x}_{l,t|t-1}, \sigma^2_{l,t|t-1}$ from (9.21).

Filtering Stage:

1. Compute $\widehat{x}^{(t)}_{n,l}, v^{(t)}_{n,l}$ from (9.22) and update $\widehat{x}_{l,t|t}, \sigma^{-2}_{l,t|t}$ from (9.19).

2. Compute $v^{(t)}_{l,n}, \widehat{x}^{(t)}_{l,n}$ from (9.24). 3. Continue steps 1) to 2) until convergence.

Smoothing Stage:
Initialization: $\mathbf{\Sigma}^{(0)}_{t-1|t} = \mathbf{\Sigma}_{t-1|t-1}, \widehat{\mathbf{x}}^{(0)}_{t-1|t} = \widehat{\mathbf{x}}_{t-1|t-1}$. Define
$\mathbf{B}^{(t)} = \mathbf{F}^H \mathbf{A}^{(t)H} \widetilde{\mathbf{R}}_t^{-1} \mathbf{A}^{(t)} \mathbf{F} + \mathbf{\Sigma}_{t-1|t-1}, \mathbf{h}_t = \mathbf{F}^H \mathbf{A}^{(t)H} \widetilde{\mathbf{R}}_t^{-1} \mathbf{y}_t$.

1. $P_{i,j} = \frac{-B^{(t)2}_{i,j}}{B^{(t)}_{i,i} + \sum\limits_{k \in \mathcal{N}(i) \setminus j} P_{k,i}}, \mu_{i,j} = (h_{i,t} + \sum\limits_{k \in \mathcal{N}(i) \setminus j} P_{k,i} \mu_{k,i}), \forall i, j.$

2. $\sigma^{-2}_{i,t-1|t} = B^{(t)}_{i,i} + \sum\limits_{k \in \mathcal{N}(i)} P_{k,i}, \widehat{x}_{i,t-1|t} = \sigma^2_{i,t-1|t}(h_{i,t} + \sum\limits_{k \in \mathcal{N}(i)} P_{k,i} \mu_{k,i})$

Estimation of hyperparameters (Define: $x'_{k,t} = x_{k,t} - f_k x_{k,t-1}, \zeta_t = \beta \zeta_{t-1} + (1 - \beta) < \left\| \mathbf{y}_t - \mathbf{A}^{(t)} \mathbf{x}_t \right\|^2 >$**) :**

1. Compute $\widehat{f}_{l|t}, \sigma^2_{f_{l|t}}$ from (9.27), $\widehat{\gamma}_t = \frac{c+N}{(\zeta_t + d)}$ and $\lambda_{l|t} = \frac{(a+1)}{(< \left| x'_{k,t} \right|^2 >_{|t} + b)}$.

Theorem 13. *If the parameter partitioning in VB is such that the different parameter blocks are decoupled at the level of Fisher Information Matrix, then VB is not suboptimal in terms of (mismatched) Cramer-Rao Bound. If a finer partitioning granularity is used (such as up to scalar level as in MF), then VB becomes quite suboptimal, which can be alleviated by using BP instead.*

$$\begin{aligned} &mCRB_{BP} = blkdiag(CRB) = blkdiag(FIM^{-1}), \\ &mCRB_{VB} = (blkdiag(FIM))^{-1}, \\ &\qquad So, \\ &mCRB_{BP} = mCRB_{VB} \; if FIM = blkdiag(FIM). \end{aligned} \tag{9.28}$$

Proof: We briefly outline the proof here. Laplace approximation refers to the evaluation of marginal likelihood or free energy using Laplace's method [190]. This is equivalent to a Gaussian approximation of the posterior $q(\boldsymbol{\theta}_i|\mathbf{y})$ around a maximum a posteriori (MAP) estimate ($\boldsymbol{\theta}^{(0)}_i$), motivated by the fact that in the asymptotic limit (large amount of data or high SNR), the posterior approaches a Gaussian around the MAP point. Under the Laplace approximation, the belief becomes $q(\boldsymbol{\theta}_i) = \mathcal{CN}(\boldsymbol{\theta}^{(0)}_i, \mathbf{\Sigma}^{(0)}_i)$. Further we evaluate the free energy [183] (F denotes the free

energy and $L = \ln p(\mathbf{y}, \boldsymbol{\theta})$)

$$
\begin{aligned}
F &= L(\boldsymbol{\theta}^{(0)}) + \frac{1}{2}\sum_{i=1}^{M}(G_i + \ln\det\boldsymbol{\Sigma}_i^{(0)} + k_i\ln(2\pi e)),\\
\ln q_i(\boldsymbol{\theta}_i) &= L(\boldsymbol{\theta}_i, \boldsymbol{\theta}_{\tilde{i}}) + \frac{1}{2}\sum_{j=1, j\neq i}^{M} G_i,\\
G_i &= \mathrm{tr}\{\boldsymbol{\Sigma}_i \frac{\partial}{\partial\boldsymbol{\theta}_i}(\frac{\partial L}{\partial\boldsymbol{\theta}_i})^H\}.
\end{aligned}
\tag{9.29}
$$

Here k_i refers to the number of scalars in $\boldsymbol{\theta}_i$ and $\ln(2\pi e)$ is the entropy of a Gaussian random variable. Now by differentiating $\ln q_i(\boldsymbol{\theta}_i)$ w.r.t the posterior covariance, we obtain the approximate covariances as

$$
\begin{aligned}
\boldsymbol{\Sigma}_i &= -(\frac{\partial}{\partial\boldsymbol{\theta}_i}(\frac{L(\boldsymbol{\theta}^{(0)})}{\partial\boldsymbol{\theta}_i})^H)^{-1}\\
&= (\mathrm{blkdiag}(FIM))^{-1}
\end{aligned}
\tag{9.30}
$$

The posterior covariance in (9.30) is computed by evaluating the Hessian at the variational mode or maximum a posteriori (MAP) point. This variational mode can be obtained as $\boldsymbol{\theta}_i^{(0)} = \max_{\boldsymbol{\theta}_i}\ln q(\boldsymbol{\theta}_i)$. In the Laplace approximation, all pdfs are Gaussian with CRB (portions) as covariance and LMMSE estimates as means. So in the too fine partitioning case, the VB partitioning is applied to the FIM, taking a too fine blockdiagonal part, and since that partitioning is finer than the blockdiagonal FIM structure, then the inverse of the too fine blockdiagonal part of the FIM does not give the correct CRB. So $mCRB_{VB} \neq CRB$. So the nodes in the factor graph are decided based on the partitioning of the blocks in the FIM block diagonal structure, such that the $mCRB_{VB} = CRB$. Here ends the proof.

9.3.1 Optimal Partitioning for Static SBL:

We define $\mathbf{J}_{\boldsymbol{\theta}_i\boldsymbol{\theta}_j} = \mathrm{E}(\frac{\partial\ln p(\mathbf{y},\boldsymbol{\theta})}{\partial\boldsymbol{\theta}_i}\frac{\partial\ln p(\mathbf{y},\boldsymbol{\theta})}{\partial\boldsymbol{\theta}_j}^H)$, which represents the part of the FIM which shows the correlation of $\boldsymbol{\theta}_i, \boldsymbol{\theta}_j$. For brevity of notation, we denote $\mathbf{J}_{\boldsymbol{\theta}_i\boldsymbol{\theta}_i} = \mathbf{J}_{\boldsymbol{\theta}_i}$. First we consider the static case when $f_l = 0, \forall l$. We omit the index t for simplicity. $f_{\xi_i}(\xi_i) = p(\xi_i|a,b), \xi_i = \lambda_i$ represents the prior distribution of the precision parameter ξ_i which is chosen as Gamma.

$$
FIM = \mathbf{J}_s = \begin{bmatrix} \mathbf{A}^H\mathbf{A} + \mathbf{I} & & \mathbf{0}_M \\ \mathbf{0}_M & \mathbf{J}_{\boldsymbol{\xi}\boldsymbol{\xi}} & \mathbf{0}_M \\ \mathbf{0}_M & \mathbf{0}_M & \mathbf{J}_{\gamma\gamma} \end{bmatrix}
\tag{9.31}
$$

The non block diagonal elements of the FIM are crosscorrelation as follows,

$$
\begin{aligned}
\mathbf{J}_{\gamma\mathbf{x}} &= \mathrm{E}(\frac{\partial\ln p(\mathbf{y},\mathbf{x},\gamma,\Xi)}{\partial\gamma}\frac{\ln p(\mathbf{y},\mathbf{x},\gamma,\Xi)}{\partial\mathbf{x}}^H)\\
&= (N/\gamma - \mathbf{v}^H\mathbf{v})(\gamma\mathbf{A}^H\mathbf{v} - \Xi x) = 0.
\end{aligned}
\tag{9.32}
$$

Similarly the crosscorrelation between $\mathbf{x}$ and Ξ will be zero and also for Ξ and γ. The cross correlations are zero because of zero mean circularly symmetric complex Gaussian variables because 3rd order moments of zero mean $\mathbf{v}$ and $\mathbf{x}$ are zero. Thus the resulting FIM will be block diagonal. In this block diagonal structure, the crosscorrelation matrix $\mathbf{J}_{\mathbf{xx}} = \mathbf{A}^H\mathbf{A} + \mathbf{I}$ will be full and thus requires the estimation of $\mathbf{x}$ using BP, while scalar factors which are decoupled γ, ξ_i can be estimated using MF. This explains the optimality of our BP-MF partitioning as shown in the Figure 9.5.

9.3.2 Optimal Partitioning for DAR-SBL:

In this section we formulate the optimal partitioning between VB and BP for the dynamic SBL case. Here we need to consider the FIMs recursively, i.e. FIM of the time update stage followed by the measurement stage. For the time update stage, we abbreviate $p(x_{k,t}, x_{k,t-1}, f_k, \lambda_k|\mathbf{y}_{1:t-1}) = p$ for convenience here.

$$\ln p = \ln \lambda_k - \lambda_k |x_{k,t} - f_k x_{k,t-1}|^2 - \sigma^{-2}_{k,t-1|t-1} |x_{k,t-1} - \widehat{x}_{k,t-1|t-1}|^2 + \sum_{k=1}^{M} \ln q_{\lambda_k}(\lambda_k). \tag{9.33}$$

The measurement FIM (9.31) is the prior FIM for the next time update. Thus it follows that BP is needed for the inference of $\mathbf{x}_t$ and MF for γ. One remark here is that the prior $\mathbf{x}_t$ covariance for the measurement update is the inverse FIM of the time update and is diagonal here.

Lemma 6. *The AR(1) model parameters require (at least lag* 1*) smoothing for identifiability.*

Proof: Considering, with augmented state $\boldsymbol{\theta}_t = [\mathbf{x}_t; \mathbf{f}; \text{diag}(\boldsymbol{\Lambda}); \gamma]$ ($3M+1$ dimensional), we obtain the FIM, $\mathbf{J}_t =$
blkdiag$(\mathbf{J}_{\mathbf{x},t}, \mathbf{J}_{\mathbf{F},t}, \mathbf{J}_{\boldsymbol{\Lambda},t}, \mathbf{J}_{\gamma,t})$. In [191], Tichavský et al. derived an elegant recursive approach to calculate the FIM recursions for a general discrete-time nonlinear filtering problem. Based on a similar derivation, we arrive at the following recursions for the sequence $\mathbf{J}_{\boldsymbol{\theta}_i,t}$ of posterior information submatrices for estimating $\boldsymbol{\theta}_i$

$$\begin{aligned}
\mathbf{J}_{\mathbf{x},t} &= \boldsymbol{\Lambda} + \gamma \mathbf{A}^{(t)\,H} \mathbf{A}^{(t)} + \boldsymbol{\Lambda}\mathbf{F}(\mathbf{F}\boldsymbol{\Lambda}\mathbf{F}^H + \mathbf{J}_{x,t-1})^{-1}\boldsymbol{\Lambda}\mathbf{F}^H, \\
\mathbf{J}_{\mathbf{F},t} &= \mathbf{J}_{\mathbf{F},t} + \mathbf{D} - \mathbf{J}_{\mathbf{F}\mathbf{x},t}(\mathbf{F}\boldsymbol{\Lambda}\mathbf{F}^H + \mathbf{J}_{\mathbf{F},t-1})^{-1}\mathbf{J}^T_{\mathbf{x}\mathbf{F},t} \\
&\text{with } \mathbf{D} = (\mathbf{I} - \mathbf{F}\mathbf{F}^H)^{-1}, \\
\mathbf{J}_{\mathbf{x}\mathbf{F},t} &= \mathbf{F}\boldsymbol{\Lambda}[\mathbf{J}_{\mathbf{x},t} + \mathbf{F}\boldsymbol{\Lambda}\mathbf{F}]^{-1}\mathbf{J}_{\mathbf{x}\mathbf{F}}, \\
\mathbf{J}_{\boldsymbol{\Lambda},t} &= \mathbf{D} - \mathbf{D}(\mathbf{D} + \mathbf{J}_{\boldsymbol{\Lambda},t-1})^{-1}\mathbf{D} \\
&\text{with } \mathbf{D} = \boldsymbol{\Lambda}^{-2},\ \mathbf{J}_{\gamma,t} = N/\gamma^2.
\end{aligned} \tag{9.34}$$

Note that $if\, \mathbf{J}_{\mathbf{x}\mathbf{F},-1} = 0,$ then $\mathbf{J}_{\mathbf{x}\mathbf{F},t} = 0, \forall\, t \geq 0$. FIM recursions show that filtering may be enough for the estimation of AR(1) parameters. However, closely looking at the expressions for $\widehat{f}_{k|t}$ derived in our work [187, eq. (24-25)] shows that $\widehat{f}_{k|t} = f_k$. This implies that we need to know the true f_k to estimate it, in the joint estimation framework. Further to prove the unidentifiability, we use the concept of global identifiability provided in [192].

$$\begin{aligned}
p(\mathbf{f}|\mathbf{x}_t, \mathbf{y}_t) &= p(\mathbf{y}_t|\mathbf{x}_t)\, p(\mathbf{x}_t|\mathbf{f})\, p(\mathbf{f}) / p(\mathbf{y}_t, \mathbf{x}_t) \\
&= p(\mathbf{x}_t|\mathbf{f})\, p(\mathbf{f}) / \int p(\mathbf{x}_t|\mathbf{f})\, p(\mathbf{f})\, d\mathbf{f} \\
&= p(\mathbf{f}|\mathbf{x}_t)
\end{aligned} \tag{9.35}$$

The above expression (9.35) suggests that posterior of $\mathbf{f}$ given $\mathbf{x}_t$ does not depend on $\mathbf{y}_t$ or in other words the observations does not provide any extra information about $\mathbf{f}$ other than the prior $p(\mathbf{f}|\mathbf{x}_t)$ and hence $\mathbf{f}$ is globally not identifiable. This proves the Lemma. (9.35) also shows that $\mathbf{f}, \mathbf{x}_t$ are coupled in the estimation unlike the decoupling property shown by the FIM analysis.

Few remarks follows: Th *mCRB* analysis in Theorem 13 indicates that the $\mathbf{x}$ part needs to be treated jointly, motivating joint VB or BP. We conjecture that whatever local identifiability analysis indicates as necessitating joint treatment for optimality requires indeed joint treatment. But local analysis may not capture all dependencies. The local analysis (recursive CRB) shows that

filtering would be sufficient for local identifiability of $\mathbf{f}$ and that the f_i and the x_i are decoupled. However, global identifiability analysis reveals that filtering is not enough for identifiability of $\mathbf{f}$ and that the estimation of x_i and f_i is coupled. The gap between local and global analysis may perhaps be reflected in the observation that the hyperparameters could be estimated (in what corresponds to filtering) by Type-II Maximum Likelihood (ML) [162] (ie ML for hyperparameters, with the random parameters x integrated out). Such Type-II ML approach for hyperparameter estimation in the dynamic problem considered here will be investigated further in future work.

Corollary 13.1. *For the smoothing stage* (9.25)*, an optimal partitioning is to apply BP for estimation of the sparse vector,* $\widehat{\mathbf{x}}_{t-1|t}$ *and MF for the correlation coefficient* $\mathbf{F}$.

Proof: The FIM recursions for smoothing stage can be obtained as (detailed derivation is skipped due to space constraints)

$$\mathbf{J}_t = \text{blkdiag}(\mathbf{J}_{\mathbf{x},t}, \mathbf{J}_{\mathbf{F},t}, \mathbf{J}_{p,t}), \tag{9.36}$$

where $\mathbf{J}_{p,t}$ representing the information submatrix for the precision parameters $\mathbf{\Lambda}, \gamma$. We obtain $\mathbf{J}_{\mathbf{x},t} =$
$\mathbf{F}^T \mathbf{A}^{(t)\,H} \widetilde{\mathbf{R}}_t^{-1} \mathbf{A}^{(t)} \mathbf{F} + \mathbf{\Lambda} - \mathbf{\Lambda}\mathbf{F}(\mathbf{F}\mathbf{\Lambda}\mathbf{F}^H + \mathbf{J}_{\mathbf{x},t-1})^{-1}\mathbf{\Lambda}\mathbf{F}^H$, which is a full matrix.

$$\begin{aligned}
\mathbf{J}_{\mathbf{F},t} &= \mathbf{J}_{\mathbf{F},t-1} + \mathbf{\Xi}\,\text{diag}\,(\mathbf{A}^{(t)\,H}\widetilde{\mathbf{R}}_t^{-1}\mathbf{A}^{(t)}) + \mathbf{D} - \mathbf{J}_{\mathbf{F}\mathbf{x},t}(\mathbf{D} + \mathbf{J}_{\mathbf{F},t-1})^{-1}\mathbf{J}_{\mathbf{x}\mathbf{F},t}, \\
&\text{with}\, \mathbf{D} = (\mathbf{I} - \mathbf{F}\mathbf{F}^H)^{-1}, \\
\mathbf{J}_{\mathbf{x}\mathbf{F},t} &= \mathbf{\Lambda}\mathbf{F}[\mathbf{J}_{\mathbf{x},t} + \mathbf{F}\mathbf{\Lambda}\mathbf{F}]^{-1}\mathbf{J}_{\mathbf{x}\mathbf{F},t}, \\
\mathbf{J}_{p,t} &= \begin{bmatrix} \mathbf{J}_{\mathbf{\Lambda},t} & \mathbf{J}_{\mathbf{\Lambda}\gamma,t} \\ \mathbf{J}_{\mathbf{\Lambda}\gamma,t} & \mathbf{J}_{\gamma\gamma} \end{bmatrix} \\
&\text{where,}\ \mathbf{J}_{\gamma\gamma} = \frac{1}{\gamma^4}\text{tr}\{\widetilde{\mathbf{R}}_t^{-1}\widetilde{\mathbf{R}}_t^{-1}\}, \\
\mathbf{J}_{\mathbf{\Lambda},t} &= \mathbf{C}_{\mathbf{\Lambda},t} + \mathbf{D} - \mathbf{D}(\mathbf{D} + \mathbf{J}_{\mathbf{\Lambda},t-1})^{-1}\mathbf{D} \\
&\text{with}\, \mathbf{D} = \mathbf{\Lambda}^{-2}, \\
(\mathbf{C}_{\mathbf{\Lambda},t})_{i,j} &= \frac{1}{\lambda_i^2\lambda_j^2}\text{tr}\{\widetilde{\mathbf{R}}_t^{-1}\mathbf{A}_i^{(t)}\mathbf{A}_i^{(t)\,H}\mathbf{A}_j^{(t)}\mathbf{A}_j^{(t)\,H}\widetilde{\mathbf{R}}_t^{-1}\}, \\
\mathbf{J}_{\mathbf{\Lambda}\gamma,t} &= \mathbf{c}_{\mathbf{\Lambda}\gamma,t}, \\
(\mathbf{c}_{\mathbf{\Lambda}\gamma,t})_i &= \frac{1}{\lambda_i^2\gamma^2}\text{tr}\{\widetilde{\mathbf{R}}_t^{-1}\mathbf{A}_i^{(t)}\mathbf{A}_i^{(t)\,H}\widetilde{\mathbf{R}}_t^{-1}\}.
\end{aligned} \tag{9.37}$$

$(\mathbf{c}_{\mathbf{\Lambda}\gamma,t})_i$ represents the i^{th} element of the vector $\mathbf{c}_{\mathbf{\Lambda}\gamma,t}$. Here also, if $\mathbf{J}_{\mathbf{x}\mathbf{F},-1} = \mathbf{0}$, then $\mathbf{J}_{\mathbf{x}\mathbf{F},t} = \mathbf{0}, \forall\, t$. Thus the FIM for $\mathbf{x}_t$ is full and it follows from Theorem 13 that optimal partitioning is to apply BP for $\mathbf{x}_t$ and MF for the correlation coefficient $\mathbf{F}$ (since $\mathbf{J}_{\mathbf{F},t}$ is diagonal and also positive definite at any time instant t) in the smoothing stage. Here ends the proof.

9.4 Simulation Results

For the observation model, the parameters chosen are $N = 100, M = 200, K = 30$. All signals are considered to be real in the simulation. All the elements of $\mathbf{A}^{(t)}$ (time varying) are generated i.i.d. from a Gaussian distribution with mean 0 and variance 1. The rows of $\mathbf{A}^{(t)}$ are scaled by $\sqrt{30}$ so that the signal part of any scalar observation has unit variance. Taking the SNR to be 20dB, the variance of each element of $\mathbf{v}_t$ (Gaussian with mean 0) is computed as 0.01.

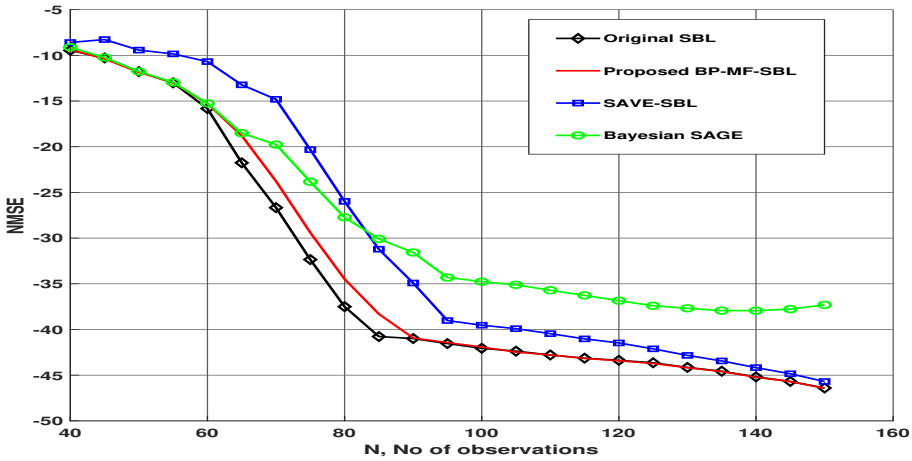


Figure 9.3: Static SBL: NMSE as a function of N.

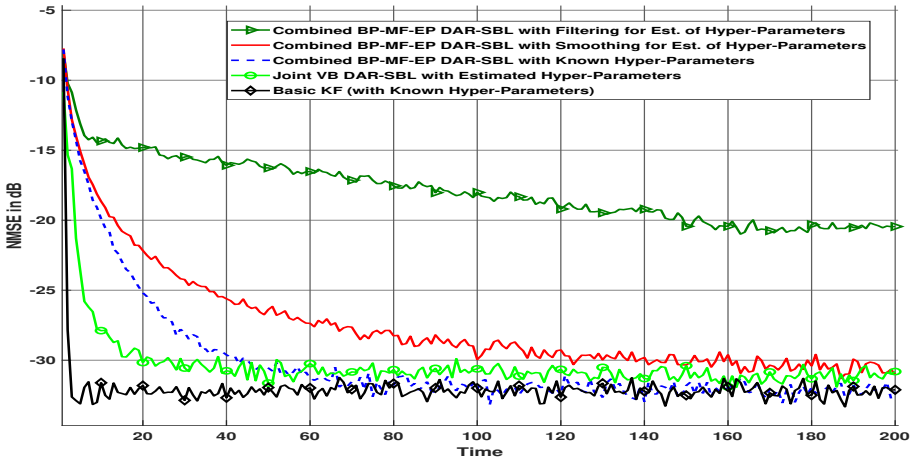


Figure 9.4: DAR-SBL: NMSE as a function of time.

Consider the state update, $\mathbf{x}_t = \mathbf{F}\mathbf{x}_{t-1} + \mathbf{w}_t$. To generate $\mathbf{x}_0$, the first 30 elements are chosen as Gaussian (mean 0 and variance 1) and then the remaining elements of the vector $\mathbf{x}_0$ are put to zero. Then the elements of $\mathbf{x}_0$ are randomly permuted to distribute the 30 non-zero elements across the whole vector. The diagonal elements of $\mathbf{F}$ are chosen uniformly in $[0.9, 1)$. Then the covariance of $\mathbf{w}_t$ can be computed as $\Xi(\mathbf{I} - \mathbf{F}\mathbf{F}^H)$. Note that Ξ contains the variances of the elements of $\mathbf{x}_t$ (including $t = 0$), where for the non-zero elements of $\mathbf{x}_0$ the variance is 1. Following observations can be made from the simulations. In Figure 9.3, for SBL with estimated hyperparameters, there is substantial improvement in normalized MSE (NMSE) by using BP instead of MF method for estimating $\mathbf{x}$. Bayesian SAGE (Space Alternating Generalized EM) corresponds to the application of [186] to SBL. In Figure 9.4, we evaluate the performance of the proposed BP-MF-EP DAR SBL and show that the parameter estimation benefits from BP. Also we show that there is a drastic improvement in performance with lag-1 smoothing for hyperparameter estimation compared to just using filtering. The gap in performance compared to the basic KF, when hyper parameters are also estimated is attributed to the estimation error in hyperparameters.

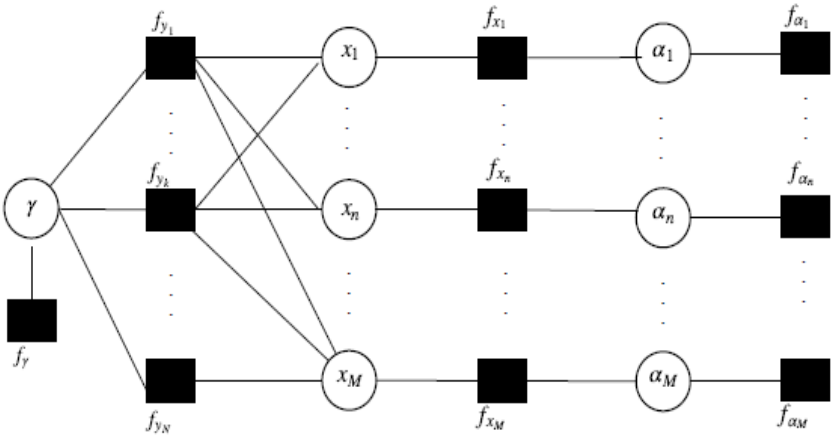


Figure 9.5: Factor Graph for the static SBL.

9.4.1 Conclusions

We presented a fast SBL algorithm called BP-MF-EP DAR-SBL, which uses a combination of BP, MF and EP techniques to approximate the posteriors of the data and parameters and track a time varying sparse signal. BP-MF-EP DAR-SBL helps to circumvent the matrix inversion operation required in the original SBL algorithm. We propose for the first time in the literature an optimal way to select the partitioning of BP and MF nodes with CRB as a performance evaluation criteria. Future work include extension of the combined BP-MF framework for Kronecker structured dictionary learning [125].

9.5 Posterior Variance Prediction: Large System Analysis for SBL using BP

We first review the BP messages being passed between the variable nodes and factor nodes corresponding to the factor graph in Figure 9.5. All the messages (beliefs or continuous pdfs) passed between them are all Gaussian [164]. So in message passing (MP), it suffices to represent them by two parameters, which are the mean and variance of the beliefs. Also, for the first instance, we assume that all the hyperparameters are known. We remark that the estimation of hyperparameters can be done using VB as in [168]. Below, indices m,n is used for representing variable nodes and i,k is used for representing factor nodes. We represent $S_{n,k}$ as the inverse variance (precision) of the message passed from variable node n (corresponding to x_n) to factor node k (corresponds to y_k) and $M_{n,k}$ be the mean of the message passed from n to k, total NM of them. Similarly $S_{k,n}, M_{k,n}$ for messages from k to n. Let $A_{k,n}$ represents the $(k,n)^{th}$ element of $\mathbf{A}$. We restrict to the case of real variables here. We start with the message passing expressions derived in [164].

$$
\begin{aligned}
S_{n,k} &= \xi_n + \sum_{i\neq k} S_{i,n}, \\
M_{n,k} &= S_{n,k}^{-1} \sum_{i\neq k} S_{i,n} M_{i,n}. \\
S_{k,n} &= A_{k,n}^2 \left(\frac{1}{\gamma} + \sum_{m\neq n} A_{k,m}^2 S_{m,k}^{-1}\right)^{-1}, \\
M_{k,n} &= A_{k,n}^{-1} \left(y_k - \sum_{m\neq n} A_{k,m} M_{m,k}\right).
\end{aligned}
\tag{9.38}
$$

Interpretation of $m_{n\to k}(x_n)$ (as Bayesian information combining): First, define the matrix $\mathbf{S}$ with entries $\sigma_{k,n}^{-2}$. At variable node n, we have

$$\widehat{\mathbf{x}}_n = \begin{bmatrix} M_{1,n} \\ \vdots \\ M_{N,n} \end{bmatrix} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} x_n + \mathcal{N}(\mathbf{0}, \mathrm{diag}\,(\mathbf{S}_{:,n})^{-1}) \tag{9.39}$$

with prior $\mathcal{N}(0, \xi_n^{-1})$.

Interpretation of $m_{k\to n}(x_n)$ (as Interference Cancellation): Substituting $x_m = M_{m,k} + \widetilde{x}_{m,k}$ ("extrinsic" information from variables $m \neq n$ for measurement k) in $y_k = \sum_m A_{k,m} x_m + v_k$ leads to the 1-1 measurement

$$\begin{aligned} &(y_k - \sum_{m\neq n} A_{k,m} M_{m,k}) = A_{k,n} x_n + (v_n + \sum_{m\neq n} A_{k,m} \widetilde{x}_{m,k}), \\ &\text{with total "noise" } v_n + \sum_{m\neq n} A_{k,m} \widetilde{x}_{m,k} \text{ of variance } \gamma^{-1} + \sum_{m\neq n} A_{k,m}^2 S_{m,k}^{-1}. \end{aligned} \tag{9.40}$$

So the (deterministic) estimate and variance from this measurement by itself are $M_{k,n} = A_{k,n}^{-1}(y_k - \sum_{m\neq n} A_{k,m} M_{m,k})$ and $S_{k,n} = A_{k,n}^2 (\frac{1}{\gamma} + \sum_{m\neq n} A_{k,m}^2 S_{m,k}^{-1})^{-1}$. This is like Bayesian SAGE! Except BSAGE did not split into messages going each way. The reason why this seemingly approximate approach, in which all correlations between all x_n and all measurements y_k are ignored, works is with i.i.d $\mathbf{A}$ indeed all x_n and all y_k get completely decorrelated, once one starts considering $\mathbf{A}$ as random. Note that instead of BP, if we use MF for the estimation of $\mathbf{x}$, the expressions above would remain the same except $S_{k,n}$ which gets written as $S_{k,n} = A_{k,n}^2 \gamma$. This can be interpreted as, MF does not take into account the error variances in other $x_m, m \neq n$ while passing the belief about x_n from any factor node y_k and hence it is suboptimal. Further, substituting $S_{n,k}$ in $S_{k,n}$

$$S_{k,n} = A_{k,n}^2 (\frac{1}{\gamma} + \sum_{m\neq n} A_{k,m}^2 (\xi_m + \sum_{i\neq k} S_{i,m})^{-1})^{-1}, \tag{9.41}$$

so this is now only in terms of the message variances in the direction k to n. Finally, the belief (estimates) computed for each x_n is

$$\begin{aligned} \sigma_n^2 &= (\xi_n + \sum_i S_{i,n})^{-1}, \\ \mu_n &= \sigma_n^2 (\sum_i S_{i,n} M_{i,n}). \end{aligned} \tag{9.42}$$

Further we simplify the messages and beliefs using the results from random matrix theory, for the simplest case of i.i.d $\mathbf{A}$ in the large system regime where $M, N \to \infty$ at a fixed ratio $\frac{N}{M} > 0$ (represented in short as $\xrightarrow[a.s]{M\to\infty}$. For the large system analysis, we use Theorem 1 and Lemma 4 from [14]. We briefly summarize the Lemma's here. Lemma 4 in Appendix VI of [14] states that $\mathbf{x}_N^H \mathbf{A}_N \mathbf{x}_N \xrightarrow{N\to\infty} (1/N) tr \mathbf{A}_N$ when the elements of $\mathbf{x}_N$ are i.i.d with variance $1/N$ and independent of $\mathbf{A}_N$, and similarly when $\mathbf{y}_N$ is independent of $\mathbf{x}_N$, that $\mathbf{x}_N^H \mathbf{A}_N \mathbf{y}_N \xrightarrow{N\to\infty} 0$. Theorem 1 from [14] implies that any terms of the form $\frac{1}{N}\mathrm{tr}\{(\mathbf{A}_N - z\mathbf{I}_N)^{-1}\}$, where $\mathbf{A}_N$ is the summation of independent rank one matrices with covariance matrix $\mathbf{\Theta}_i$ is equal to the unique positive solution of $e_j = \frac{1}{N}\mathrm{tr}\{(\sum_{i=1}^{K} \frac{\mathbf{\Theta}_i}{1+e_i} - z\mathbf{I}_N)^{-1}\}$. Under the LSL simplifications using these results, we arrive at the following theorem,

Theorem 14. *In the LSL, under i.i.d entries in* **A***, the predicted (by BP or xAMP algorithms) per component MSE (or the posterior variance* σ_n^2*) converges exactly to the Bayes optimal values (i.e. the diagonal elements of the posterior covariance matrix for LMMSE). This result being applicable for AMP (GAMP also under i.i.d* **A***), since the derivation of AMP follows from BP under the LSL.*

Proof: In the large system limit, we can approximate (neglecting terms of $\mathcal{O}(A_{i,j}^2)$) $S_{n,k} = \xi_n + \sum_i S_{i,n} = S_n$, independent of k. Further we define $\mathbf{S} = \text{diag}(S_n)$. Considering the term $S_{k,n} = A_{k,n}^2(\frac{1}{\gamma} + \sum_{m\neq n} A_{k,m}^2 S_{m,k}^{-1})^{-1}$, in the large system it can be approximated by

$$S_{k,n} = A_{k,n}^2(\frac{1}{\gamma} + \mathbf{A}_{k,:}\mathbf{S}^{-1}\mathbf{A}_{k,:}^T)^{-1}. \tag{9.43}$$

Also

$$\mathbf{A}_{k,:}\mathbf{S}^{-1}\mathbf{A}_{k,:}^T \xrightarrow[a.s]{M\to\infty} \frac{1}{M}\text{tr}\{\mathbf{S}^{-1}\} = \tau'_{BP}. \tag{9.44}$$

From (9.42), it follows that $MSE = \tau_{BP} = \text{tr}\{\mathbf{S}^{-1}\}$. $\mathbf{A}_{k,:}$ represents the k^{th} row of $\mathbf{A}$. Further we obtain,

$$\begin{aligned} S_n &= \xi_n + (\frac{1}{\gamma} + \tau'_{BP})^{-1}\sum_i A_{i,n}^2, \\ \sum_i A_{i,n}^2 &\xrightarrow[a.s]{M\to\infty} 1, \\ \text{thus}\quad S_n &= \xi_n + (\frac{1}{\gamma} + \tau'_{BP})^{-1}. \end{aligned} \tag{9.45}$$

Finally we can conclude that, τ'_{BP} can be obtained as the unique positive solution of the following fixed point equation

$$\tau'_{BP} = \sum_{n=1}^{M}(\xi_n + (\tfrac{1}{\gamma} + \tau'_{BP})^{-1})^{-1}. \tag{9.46}$$

Next step is to simplify the expression for LMMSE posterior covariance in the large system limit using similar techniques as above. The posterior covariance can be written as

$$\begin{aligned} \mathbf{\Sigma}_L &= \mathbf{\Xi}^{-1} - \mathbf{\Xi}^{-1}\mathbf{A}^T(\mathbf{A}\mathbf{\Xi}^{-1}\mathbf{A}^T + \frac{1}{\gamma})^{-1}\mathbf{A}\mathbf{\Xi}^{-1}, \\ \mathbf{A}^T(\mathbf{A}\mathbf{\Xi}^{-1}\mathbf{A}^T + \frac{1}{\gamma})^{-1}\mathbf{A} &\xrightarrow[(a)]{M\to\infty} \mathbf{D}, \\ \mathbf{D}_{i,i} &= \frac{e}{1+\frac{e}{\xi_i}}, \end{aligned} \tag{9.47}$$

where (a) follows from Theorem 1 in [14] and e is defined as the unique positive solution of the

following fixed point equation

$$
\begin{aligned}
e &= (\frac{1}{N}\sum_{i=1}^{M}\frac{\xi_i^{-1}}{1+\frac{e}{\xi_i}}+\frac{1}{\gamma})^{-1},\\
\mathrm{tr}\{\mathbf{\Sigma}_L\} = MSE &= \sum_{i=1}^{M}\frac{\xi_i^{-1}e}{1+\frac{e}{\xi_i}},\\
\text{From } e, \frac{1}{e}-\frac{1}{\gamma} &= \frac{1}{N}\sum_{i=1}^{M}\frac{\xi_i^{-1}}{1+\frac{e}{\xi_i}}\\
&= \frac{1}{N}MSE = \frac{\tau}{N} = \tau',\\
\frac{1}{e} &= \frac{1}{\gamma}+\tau',\\
\tau' &= \frac{1}{N}\sum_{i=1}^{M}\frac{\xi_i^{-1}(\frac{1}{\gamma}+\tau')}{\frac{1}{\gamma}+\tau'+\frac{1}{\xi_i}}\\
&= \frac{1}{N}\sum_{i=1}^{M}\frac{1}{\xi_i+(\frac{1}{\gamma}+\tau')^{-1}}.
\end{aligned}
\tag{9.48}
$$

Comparing (9.46) and (9.48), it can be observed that the MSE under BP, τ_{BP} and the MMSE τ can be obtained as a unique positive solution of the same fixed point equation. This implies that in the large system limit, under i.i.d $\mathbf{A}$, if BP converges, the MSE of SBL (assuming the hyperparameters are fixed or known) converges to the exact MMSE. Moreover, it can be observed from (9.48) that, the per component MSE predicted by BP matches the diagonal elements of the LMMSE covariance, which has never been pointed out before in the literature. Here ends the proof.

One remark here is that the above large system analysis based on [14] can be applied to more general measurement matrices case, with rows of $\mathbf{A}$ being restricted to have different covariance matrices, i.e. $\mathrm{E}(\mathbf{A}_{i,:}^T\mathbf{A}_{i,:}) = \mathbf{\Theta}_i$. Certain remarks comparing the existing convergence conditions for BP is as follows. In [193], Jian Du et al. shows that depending on the underlying graphical structure (GMRF or factor graph based factorization) GaBP may exhibit diffferent convergence properties. They prove that the convergence condition for the mean provided based on the factor graph representation encompasses much larger class of models than those given by the GMRF based walk-summable condition [194]. Further they show that GaBP always converges if the factor graph is a union of single loop and a forest. Moreover, they also analyze the convergence of the inverse of the message variances (message information matrix) and analytically show that with arbitrary positive semidefinite matrix initialization, the message information matrix converges to a uniques positive definite matrix. So we can conclude that for BP there is a decoupling between the dynamics of the variance updates and that of the mean updates. And that we know that the mean converges to the LMMSE estimate under certain conditions. But it is to be mentioned that the convergence conditions and convergence values for the variance are more tricky, still requires rigorous analysis to characterize its behaviour, which is the main motivation behind this section.

9.5.1 Iterations in Matrix Form

Let us denote $d(\mathbf{A})$ as the vector with entries as the diagonal elements of $\mathbf{A}$. $\mathbf{B}$ is defined as the matrix with entries as $A_{i,j}^2$. Let $\mathbf{L}$(of size $M \times N$), $\mathbf{S}$,$\mathbf{M}$ (of size $N \times M$) be the matrix with entries $S_{n,k}M_{n,k}$, $S_{k,n}$ and $M_{k,n}$, respectively. Defining $\mathbf{T}$ to be a matrix of size $M \times N$, with entries as the

inverse variance of the Gaussian messages transmitted from the variable nodes, $S_{n,k}$, we obtain

$$\begin{aligned}\mathbf{T} &= (d(\boldsymbol{\Xi}) + \mathbf{S}^T\mathbf{1}_N)\otimes\mathbf{1}_N^T - \mathbf{S}^T,\\ \mathbf{L} &= d(\mathbf{S}^T\mathbf{M})\otimes\mathbf{1}_N^T - (\mathbf{S}\circ\mathbf{M})^T,\\ \mathbf{L}' &= \mathbf{T}^{-1}\circ\mathbf{L}.\end{aligned} \tag{9.49}$$

We denote any matrix $\mathbf{A}_{inv}$ as a matrix with entries as the element wise inverse of the matrix $\mathbf{A}$. Similarly, for the messages at the factor nodes, define $\mathbf{C}$ to be the matrix with entries $A_{k,n}^2 S_{k,n}^{-1}$

$$\begin{aligned}\mathbf{C} &= \left(\frac{1}{\gamma}\mathbf{1}_N + d(\mathbf{B}\mathbf{T}_{inv})\right)\otimes\mathbf{1}_M^T - \mathbf{B}\circ\mathbf{T}_{inv}^T,\\ \mathbf{S} &= \mathbf{C}_{inv}\circ\mathbf{B},\\ \mathbf{V} &= (\mathbf{y} - d(\mathbf{A}\mathbf{L}'))\otimes\mathbf{1}_M^T + \mathbf{A}\circ\mathbf{L}'^T,\\ \mathbf{M} &= \mathbf{A}_{inv}\circ\mathbf{V},\end{aligned} \tag{9.50}$$

where $\mathbf{V}$ being the matrix with entries $A_{k,n}M_{k,n}$. The computational complexity of all the matrix operations above is $\mathcal{O}(MN)$, since the number of computations in the Hadamard product or Kronecker products in the above expressions is only MN. Assuming the number of iterations required to converge is N_{it}, the total complexity of the BP algorithm can be written as $N_{it}\mathcal{O}(MN)$.

9.5.2 Convergence Analysis of BP

In this subsection, we consider the convergence analysis of the mean and variance of the messages passed in BP. For the ease of analysis, we consider a simplified case, where we neglect terms of the order $\mathcal{O}(A_{i,j}^2)$ under the large system limit $M,N\to\infty$. Hence the precisions of the posteriors passed $A_{k,n}^{-1}S_{k,n}, S_{n,k}$ in (9.38) can be approximated as $S_n = \xi_n + \sum_i S_{i,k}$ and $A_{k,n}^{-2}S_{k,n} = (\frac{1}{\gamma} + \sum_m A_{k,m}^2 S_{m,k}^{-1})^{-1} \triangleq S_k$. In fact, $S_n, A_{k,n}^{-2}S_{k,n}$ represent the precision variables in the input and output stages of the GAMP algorithm derived in [195, Algorithm 1]. Using theorem 1 in [195], we can show that for any non-negative matrix $B >= 0$, S_n, S_k converge to a positive value. However, we remark that it remains to be understood to which value these precision variables converge (and hence the posterior variance σ_n^2) and it is left as a future work.

Further we look at the convergence behaviour of the mean value of the posteriors passed across the graph $M_{k,n}$. Substituting the value of $M_{m,k}$ in the expression of $M_{k,n}$ in (9.38), we obtain

$$M_{k,n} = A_{k,n}^{-1}(y_k - \sum_{m\neq n}\sum_{i\neq k} A_{k,m}A_{i,m}^2 S_m^{*-1}S_i^* M_{i,m}), \tag{9.51}$$

where S_i^*, S_m^* are the converged values of the precision variables S_i, S_m, respectively. Defining $\mathbf{m}^{(t)}$ as a vector of length MN, representing the values $M_{k,n}$ at iteration t. So $\mathbf{m}^{(t)} = [M_{1,1}, M_{1,2}, ...1,M, ..., M_{N,M}]^T$. Also, we define $\mathbf{N}$ to be a diagonal matrix of length $MN\times MN$ with entries $A_{k,n}^{-1}$ and $\mathbf{M}$ to be a $MN\times MN$ matrix with $((i-1)M+m)^{th}$ entry of the k^{th} row of M being defined as $A_{k,m}A_{i,m}^2 S_m^{-1}S_i$, but equal to zero when either $i=k$ or $m=n$ or $i=k$ and $m=n$.

$$\mathbf{m}^{(t+1)} = -\mathbf{M}\mathbf{m}^{(t)} + \mathbf{N}(\mathbf{y}\otimes\mathbf{1}_M). \tag{9.52}$$

The above iterations (9.52) converges if $\rho(\mathbf{M}) < 1$.

9.5.3 Scalar Iterations

Further defining the following terms

$$\begin{aligned} Z_{k,n} &= (y_k - \sum_{m \neq n} A_{k,m} M_{m,k}), \\ \text{So } M_{k,n} &= A_{k,n}^{-1} Z_{k,n}. \end{aligned} \tag{9.53}$$

Also, assume that in the large system limit, $M_{n,k}$ can be written as, $M_{n,k} = M_n + \delta_{n->k}$, where $\delta_{n->k}$ is of the $O(\frac{1}{\sqrt{N}})$. This approximation follows from writing

$$\begin{aligned} M_{n,k} &= S_{n,k}^{-1} \sum_{i \neq k} S_{i,n} M_{i,n} \\ &= S_{n,k}^{-1} \sum_{i} S_{i,n} M_{i,n} - M_{k,n}. \end{aligned} \tag{9.54}$$

Substituting $M_{n,k}$ in $Z_{k,n}$

$$Z_{k,n} = (y_k - \sum_m A_{k,m} M_m - \sum_m A_{k,m} \delta_{m->k} + A_{k,n} M_n + O(\frac{1}{N})) \quad = Z_k + \delta_{k->n}, \tag{9.55}$$

all the terms containing $A_{i,j}^2 or A_{i,j} \delta_{j->i}$ becomes $O(\frac{1}{N})$ and $\delta_{k->n} = A_{k,n} M_n$, also here

$$Z_k = (y_k - \sum_m A_{k,m} M_m - \sum_m A_{k,m} \delta_{m->k}). \tag{9.56}$$

$$\begin{aligned} M_{n,k} &= S_{n,k}^{-1} (\frac{1}{\gamma} + \tau'_{BP})^{-1} \sum_{i \neq k} A_{i,n} Z_{i,n} \\ &= S_n^{-1} (\frac{1}{\gamma} + \tau'_{BP})^{-1} \sum_{i \neq k} A_{i,n} Z_{i,n}. \end{aligned} \tag{9.57}$$

As in the papers by Montanari et. al. [196], for general priors, it is possible to write $M_{n,k} = f_n(\sum_{i \neq k} A_{i,n} Z_{i,n})$. Here f_n is a linear function for the Gaussian case (i.e. $f_n(x) = S_n^{-1}(\frac{1}{\gamma} + \tau)^{-1} x_n$). So if we consider the case of Gamma priors for ξ etc, then this parameterization in terms of an f becomes easy to write the recursions. Now doing a first order Taylor series approximation of f around $\sum_i A_{i,n} Z_{i,n}$, $M_{n,k} = f_n(\sum_i A_{i,n} Z_{i,n}) - A_{k,n} Z_{k,n} f'_n(\sum_i A_{i,n} Z_{i,n})$, f'_n being derivative evaluated at
$\sum_i A_{i,n} Z_{i,n}$. Further substituting for $Z_{i,n}$ from (9.53)

$$\begin{aligned} M_{n,k} &= M_n + \delta_{n->k}, \\ M_n &= f_n(\sum_i A_{i,n} Z_i + \sum_i A_{i,n} \delta_{i->n}) \\ \text{and } \delta_{n->k} &= -A_{k,n} Z_k f'_n(\sum_i A_{i,n} Z_i). \end{aligned} \tag{9.58}$$

Note that term $A_{k,n} \delta_{k,n}$ becomes $O(\frac{1}{N})$. Substituting for $\delta_{i->n}$ and with the large system approximation $\sum_i A_{i,n}^2 -> 1$

$$\begin{aligned} M_n &= f_n(\sum_i A_{i,n} Z_i + \sum_i A_{i,n}^2 M_n) \\ &= f_n(\sum_i A_{i,n} Z_i + M_n). \end{aligned} \tag{9.59}$$

Now further writing as a vector M (with each element $M_n, \forall n$). $M = f(A^T Z + M)$, which is the AMP recursion for the mean and $f_n(\cdot)$ represents each of the scalar components in $f(\cdot)$. Also in (9.53), substituting for $\delta_{n->k}$ from (9.58)

$$\begin{aligned} Z_k &= (y_k - \sum_m A_{k,m} M_m) + (\frac{1}{\delta}) Z_k (\frac{1}{n}) \sum_m f'_m(\sum_i A_{i,m} Z_i) \\ &= (y_k - \sum_m A_{k,m} M_m) + \frac{1}{M} Z_k \sum_m f'_m(\sum_i A_{i,m} Z_i), \end{aligned} \tag{9.60}$$

where $(\frac{1}{M}) Z_k \sum_m f'_m(\sum_i A_{i,m} Z_i)$ is the Onsager term.

9.5.4 Original AMP Iterations and SBL-AMP

The difference in AMP vs SBL-AMP is that in AMP the denoising function (or called as the shrinkage function in AMP literature) $f_m(x) = f(x)$, i.e the same function for every component. The original AMP iterations (for any Lipschitz-continuous component-wise shrinkage function $\mathbf{f}$ and i.i.d $\mathbf{x}$) can be written as

$$\begin{aligned} \mathbf{z}_t &= \mathbf{y} - \mathbf{A}\widehat{\mathbf{x}}_t + \frac{1}{\delta}\mathbf{z}_{t-1} < \mathbf{f}'(\widehat{\mathbf{x}}_{t-1} + \mathbf{A}^T \mathbf{z}_{t-1}) >, \\ \widehat{\mathbf{x}}_{t+1} &= \mathbf{f}(\widehat{\mathbf{x}}_t + \mathbf{A}^T \mathbf{z}_t). \end{aligned} \tag{9.61}$$

Onsager correction serves to decouple the input to AMP [196], $\mathbf{r}_t = \widehat{\mathbf{x}}_t + \mathbf{A}^T \mathbf{z}_t = \mathbf{x} + \mathcal{N}(\mathbf{0}, \tau_t \mathbf{I}_M)$.

$$\begin{aligned} &\text{in case of } \mathcal{N}(\mathbf{0}, \frac{1}{\xi}\mathbf{I}), \text{we get LMMSE } \widehat{\mathbf{x}}_{t+1} = \mathbf{f}(\mathbf{r}_t) = b_t \mathbf{r}_t, \\ &\qquad b_t = \frac{\frac{1}{\xi}}{\frac{1}{\xi} + \tau_t}, \\ &\text{and State Evolution (SE) } \tau_{t+1} = \frac{1}{\gamma} + \frac{1}{\delta}(1 - b_t)^2 \frac{1}{\xi} + \frac{1}{\delta} b_t^2 \tau_t \\ &\qquad\qquad = \frac{1}{\gamma} + \frac{1}{\delta}(\xi + \tau_t^{-1})^{-1}. \end{aligned} \tag{9.62}$$

However, in SBL-AMP (for SBL $\mathbf{x} \sim \mathcal{N}(0, \Xi^{-1})$), iterations decouple for $\mathbf{r}_t$ as follows, $\mathbf{r}_t = \mathbf{x} + \mathcal{N}(\mathbf{0}, \tau_t \mathbf{I})$ leading to $\widehat{\mathbf{x}}_{t+1} = \mathbf{f}(\mathbf{r}_t) = \mathbf{F}_t \mathbf{r}_t$, with diagonal $\mathbf{F}_t = (\mathbf{I}_M + \tau_t \Xi)^{-1}$. Define $\mathbf{A}_m$ as the m^{th} column of $\mathbf{A}$ and $\mathbf{A}_{\bar{m}}$ as the matrix excluding column m, vector $\boldsymbol{\delta}_{\bar{m}->k}$ contains as entries $\delta_{n->k}, n \neq m$:

$$\begin{aligned} &\text{Consider } m^{th} \text{ noise element } n_{m,t} = \mathbf{A}_m^T \mathbf{A}_{\bar{m}} \widetilde{\mathbf{x}}_{\bar{m},t} - \mathbf{A}_m^T \Delta_m + \mathbf{A}_m^T \mathbf{v}, \\ &\qquad \Delta_{m,k} = \mathbf{A}_{k,\bar{m}} \boldsymbol{\delta}_{\bar{m}->k}, \\ &\text{leading to } \tau_{t+1} = \frac{1}{\gamma} + \frac{1}{\delta}\frac{1}{M}\sum_{n=1}^{M}(\xi_n + \tau_t^{-1})^{-1}. \end{aligned} \tag{9.63}$$

9.6 Bayesian SAGE (BSAGE)

In this section, we consider a Bayesian version of the space alternating generalized EM (SAGE) algorithm proposed in [186, 197]. In BSAGE, we consider the estimation of x_i by fixing the other

variables and splitting $x_k = \widehat{x}_k + \widetilde{x}_k, \forall k \neq i$. We define $\mathbf{\Sigma}_{\bar{i}}$ is the diagonal matrix with entries as the posterior variances $\sigma_k^2, k \neq i$. So we write the observation model as

$$\mathbf{y} - \mathbf{A}_{\bar{i}}\widehat{\mathbf{x}}_{\bar{i}} \stackrel{\Delta}{=} \mathbf{y}_i = \mathbf{A}_i x_i + \mathbf{A}_{\bar{i}}\widetilde{\mathbf{x}}_{\bar{i}} + \mathbf{v}, \tag{9.64}$$

Further we obtain the LMMSE estimate of x_i as

$$\begin{aligned} \sigma_i^2 &= \xi_i + \mathbf{A}_i^T(\mathbf{A}_{\bar{i}}\mathbf{\Sigma}_{\bar{i}}\mathbf{A}_{\bar{i}}^T + \frac{1}{\gamma}\mathbf{I}_N)^{-1}\mathbf{A}_i, \\ \widehat{x}_i &= \sigma_i^2\mathbf{A}_i^T(\mathbf{A}_{\bar{i}}\mathbf{\Sigma}_{\bar{i}}\mathbf{A}_{\bar{i}}^T + \frac{1}{\gamma}\mathbf{I}_N)^{-1}\mathbf{y}_i \end{aligned} \tag{9.65}$$

We further define, $\mathbf{E}_i$ as the diagonal matrix with i^{th} entry $\frac{1}{\xi_i}$ and rest of the elements same as $\mathbf{\Sigma}$. Also, define $\mathbf{V}_i = \mathbf{A}(\mathbf{E}_i^{-1}\gamma^{-1} + \mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T$. Further applying matrix inversion lemma [14] and substituting for $\mathbf{y}_i$, we obtain

$$\widehat{x}_i = \tfrac{\gamma}{\xi_i}\mathbf{A}_i^T\mathbf{y} - \mathbf{A}_i^T\mathbf{V}_i\tfrac{\gamma}{\xi_i}\mathbf{y} - \tfrac{\gamma}{\xi_i}\mathbf{A}_i^T\mathbf{A}_{\bar{i}}\widehat{\mathbf{x}}_{\bar{i}} + \mathbf{A}_i^T\mathbf{V}_i\tfrac{\gamma}{\xi_i}\mathbf{A}_{\bar{i}}\widehat{\mathbf{x}}_{\bar{i}}. \tag{9.66}$$

Further, in order to write it in the vector form, we define the matrix $\mathbf{B}^T$ (of size $M \times N$) with the rows as $\mathbf{A}_i^T\mathbf{V}_i$. We obtain the expressions in the vector form as

$$\begin{aligned} \widehat{\mathbf{x}}^{(k+1)} &= -\mathbf{M}\widehat{\mathbf{x}}^{(k)} + \mathbf{N}\mathbf{y}, \\ \text{where, } \mathbf{M} &= \gamma\mathbf{\Xi}^{-1}(\mathbf{H} - \mathbf{L}), \\ \mathbf{L} &= (\mathbf{B}^T\mathbf{A} - \operatorname{diag}(\mathbf{B}^T\mathbf{A})), \\ \mathbf{H} &= (\mathbf{A}^T\mathbf{A} - \operatorname{diag}(\mathbf{A}^T\mathbf{A})), \\ \mathbf{N} &= \gamma\mathbf{\Xi}^{-1}(\mathbf{A} - \mathbf{B})^T. \end{aligned} \tag{9.67}$$

The per-iteration complexity of BSAGE is also $\mathcal{O}(M^2N)$, hence same as BP. The convergence condition can be written as $\rho(\mathbf{M}) < 1$. Further comparing the convergence conditions for SAVE and BSAGE, $\rho_{SAVE} = \rho([\gamma\operatorname{diag}(\mathbf{A}^T\mathbf{A}) + \mathbf{\Xi}]^{-1}\text{offdiag}(\gamma\mathbf{A}^T\mathbf{A}))$ and $\rho_{BSAGE} = \rho(\mathbf{\Xi}^{-1}\text{offdiag}(\gamma(\mathbf{A}-\mathbf{B})^T\mathbf{A}))$. It can be observed that if $\mathbf{A}^T\mathbf{A}$ is diagonally dominant (which is also one of the conditions for the convergence of SAVE to the true means), then the effect of the offdiagonal terms of $(\mathbf{A}-\mathbf{B})^T\mathbf{A}$ or $\mathbf{A}^T\mathbf{A}$ is negligible and the dominating factor is the first term in the expression of ρ. Since $[\gamma\operatorname{diag}(\mathbf{A}^T\mathbf{A}) + \mathbf{\Xi}]^{-1} < \mathbf{\Xi}^{-1}$, we can conclude that $\rho_{SAVE} < \rho_{BSAGE}$ explaining the faster convergence of SAVE as noted in [168] and [198].

9.7 Concluding Remarks on Combined BP-MF-EP DAR-SBL

Motivated by the need for low complexity solutions for sparse signal recovery, we looked at various approximate inference techniques for SBL whose complexity is of the order of the length of the sparse signal. In this chapter, we attempt to provide convergence analysis for SBL under approximate inference techniques such as VB, BP or EP. However, much remains to be done. The convergence values of the posterior variances for BP still needs to be understood. One possible future direction is to analyze the convergence behaviour with estimated hyperparameters. Another extension of the present work is when the dictionary matrix is unknown, for example structured dictionary matrices as in [199, 200].

9.8 Towards a Convergent AMP-SBL Solution

It is of great importance to analyze the convergence conditions of approximate message passing based algorithms. State evolution (SE) analysis done on the class of i.i.d matrices ([196, 201]) show that the mean square error converges to the Bayes optimal value in the large system limit. Unfortunately, while AMP performs well for zero-mean i.i.d. projections, performance tends to drastically decline if the measurement matrix deviates even slightly from this case. The authors in [202] have shown even for i.i.d non-zero mean measurement matrix, the AMP algorithm tends to diverge. Hence to overcome these issues several techniques have been proposed in the literature including adaptive damping, mean removal [203] and sequential AMP (called Swept AMP) [204]. However, issues with damping is that it may further slow down the convergence rate, thus making the algorithm highly complex. Also, it is not yet sure how to determine an optimal damping factor.

First, we look at a small variation of our AMP-SBL algorithm detailed in the previous section, where we avoid approximating $\sum_i A_{ij}^2 \approx 1$. Here is the outline of the derivation, which starts from BP expressions and follows the similar lines as AMP-SBL. We start from the BP-SBL message passing expressions in Section 9.5. In the large system limit, we can approximate (neglecting terms of $\mathcal{O}(A_{i,j}^2)$) $\sigma_{n,k}^{-2} = \xi_n + \sum_i \sigma_{i,n}^{-2} = \sigma_n^{-2}$, independent of k. Further we define $\mathbf{\Sigma} = \text{diag}(\sigma_n^2)$. Considering the term $\sigma_{k,n}^{-2} = A_{k,n}^2(\frac{1}{\gamma} + \sum_{m\neq n} A_{k,m}^2 \sigma_{m,k}^2)^{-1}$, in the LSL it can be approximated by $\sigma_{k,n}^{-2} = A_{k,n}^2(\frac{1}{\gamma} + \mathbf{A}_{k,:}\mathbf{\Sigma}\mathbf{A}_{k,:}^T)^{-1}$. $\mathbf{A}_{k,:}\mathbf{\Sigma}\mathbf{A}_{k,:}^T = \tau_k$. $\mathbf{A}_{k,:}$ represents the k^{th} row of $\mathbf{A}$. From posterior belief variances, it follows that $MSE = \text{tr}\{\mathbf{\Sigma}\}$. Further we obtain, $\sigma_n^{-2} = \xi_n + \sum_i (\frac{1}{\gamma} + \tau_i)^{-1} A_{i,n}^2$. Further, we write the variance recursions in matrix form as

$$\begin{aligned} \Sigma_t^{-1} &= \Xi + \text{diag}(\mathbf{A}^T[\text{diag}(\frac{1}{\gamma}\mathbf{I}_N + \mathbf{A}\Sigma_{t-1}\mathbf{A}^T)]^{-1}\mathbf{A}), \\ MSE &= \text{tr}\{\mathbf{\Sigma}_t\} \end{aligned} \tag{9.68}$$

Further, we define $Z_{k,n}$ and arrive at the approximate expression in terms of Z_k, which will be the same as in Section 9.5.3. Substituting for $\sigma_{n,k}^2 = \sigma_n^2$ and $\sigma_{k,n}^2 = A_{k,n}^2(\frac{1}{\gamma} + \tau_k)^{-1}$, the expression of $\widehat{x}_{n,k} = \sigma_{n,k}^2 \sum_{i\neq k} \sigma_{i,n}^{-2} \widehat{x}_{i,n}$ becomes

$$\widehat{x}_{n,k} \approx \sigma_n^2 \sum_{i\neq k} A_{i,n}(\tfrac{1}{\gamma} + \tau_i)^{-1} z_{i,n}. \tag{9.69}$$

We can write $\widehat{x}_{n,k} = f_n(\sum_{i\neq k} A_{i,n}(\frac{1}{\gamma} + \tau_i)^{-1} z_{i,n})$. Here f_n is a linear function for the Gaussian case (i.e. $f_n(x) = \sigma_n^2 x$ and $f_n(x)' = \sigma_n^2$, also $s_i = (\frac{1}{\gamma} + \tau_i)^{-1}$).

Performing a first order Taylor series approximation of f around $\sum_i A_{i,n} s_i z_{i,n}$

$$\widehat{x}_{n,k} = f_n(\sum_i A_{i,n} s_i z_{i,n}) - A_{k,n} s_i z_{k,n} f_n'(\sum_i A_{i,n} s_i z_{i,n}), \tag{9.70}$$

f_n' being derivative evaluated at $\sum_i A_{i,n} s_i z_{i,n}$. Further substituting for $z_{i,n}$ from (9.53)

$$\begin{aligned} \widehat{x}_{n,k} &= \widehat{x}_n + \delta_{n\to k}, \\ \widehat{x}_n &= f_n(\sum_i A_{i,n} s_i z_i + \sum_i A_{i,n} s_i \delta_{i\to n}) \\ \text{and } \delta_{n\to k} &= -A_{k,n} s_k z_k f_n'(\sum_i A_{i,n} s_i z_i + \sum_i A_{i,n} s_i \delta_{i\to n}). \end{aligned} \tag{9.71}$$

Define $\mathbf{S} = \text{diag}(s_i)$. Substituting for $\delta_{i\to n} = A_{i,n}\widehat{x}_n$, $\widehat{x}_n = f_n(\sum_i A_{i,n}s_i z_i + \sum_i s_i A_{i,n}^2 \widehat{x}_n)$.
In vector form: $\widehat{\mathbf{x}} = \mathbf{f}(\mathbf{A}^T\mathbf{S}\mathbf{z} + \text{diag}(\mathbf{A}^T\mathbf{S}\mathbf{A})\widehat{\mathbf{x}})$, which is the AMP recursion for the mean, where $(\mathbf{f}(\mathbf{x}))_n = f_n(x_n)$. Also from (9.56), substituting $\delta_{n\to k}$ from (9.71) and defining $\mathbf{z}_t = [z_1, \cdots, z_N]^T$ at iteration t :

$$\mathbf{z}_t = (\mathbf{y} - \mathbf{A}\widehat{\mathbf{x}}_t) + \Big(\mathbf{S}(\mathbf{A}\circ\mathbf{A})\mathbf{f}'\Big(\mathbf{A}^T\mathbf{S}\mathbf{z}_{t-1} + \text{diag}(\mathbf{A}^T\mathbf{S}\mathbf{A})\widehat{\mathbf{x}}_t\Big)\Big)\circ\mathbf{z}_{t-1}, \tag{9.72}$$

where $\Big(\mathbf{S}(\mathbf{A}\circ\mathbf{A})\mathbf{f}'(\mathbf{A}^T\mathbf{S}\mathbf{z}_{t-1} + \text{diag}(\mathbf{A}^T\mathbf{S}\mathbf{A})\widehat{\mathbf{x}}_t)\Big)\circ\mathbf{z}_{t-1}$ is the Onsager term. Finally, we arrive at the modified AMP-SBL iterations (which is same as the GAMP-SBL in [166])

$$\begin{aligned}
\mathbf{z}_t &= (\mathbf{y} - \mathbf{A}\widehat{\mathbf{x}}_{t-1}) + \Big(\mathbf{S}_{t-1}(\mathbf{A}\circ\mathbf{A})\mathbf{f}'\Big(\text{diag}(\mathbf{A}^T\mathbf{S}_{t-1}\mathbf{A})^{-1}\mathbf{A}^T\mathbf{S}\mathbf{z}_{t-1} + \widehat{\mathbf{x}}_{t-1}\Big)\Big)\circ\mathbf{z}_{t-1},\\
\mathbf{S}_t &= [\,\text{diag}(\frac{1}{\gamma}\mathbf{I}_N + \mathbf{A}\Sigma_{t-1}\mathbf{A}^T)]^{-1},\\
\Sigma_t^{-1} &= \Xi + \text{diag}(\mathbf{A}^T[\,\text{diag}(\frac{1}{\gamma}\mathbf{I}_N + \mathbf{A}\Sigma_{t-1}\mathbf{A}^T)]^{-1}\mathbf{A}),\\
\widehat{\mathbf{x}}_{t+1} &= \mathbf{f}(\underbrace{[\,\text{diag}(\mathbf{A}^T\mathbf{S}_t\mathbf{A})]^{-1}\mathbf{A}^T\mathbf{S}_t\mathbf{z}_t + \widehat{\mathbf{x}}_t}_{\mathbf{r}_t}) = \mathbf{F}_t\mathbf{r}_t,\\
\mathbf{F}_t &= \text{diag}(\mathbf{A}^T\mathbf{S}_t\mathbf{A})\Big(\Xi + \text{diag}(\mathbf{A}^T\mathbf{S}_t\mathbf{A})\Big)^{-1}.
\end{aligned} \tag{9.73}$$

One remark here is that GAMP-SBL does not converge without damping or mean removal procedure (for non-zero mean $\mathbf{A}$) as is noted in [166]. Hence we are inclined to explore other alternatives which are converging.

9.8.1 Fixed Points of Bethe Free Energy and GSwAMP-SBL

When the computation of the posterior distribution becomes intractable, our aim would become to perform probabilistic inference by minimizing the variational free energy (VFE) over an approximate posterior $q(\boldsymbol{\theta})$. The VFE can be written as [205]

$$\mathscr{F}(q) = KLD(q(\boldsymbol{\theta})||P_0(\boldsymbol{\theta})) - <\log p_{\mathbf{y}|\boldsymbol{\theta}}(\mathbf{y}|\boldsymbol{\theta})>_q \tag{9.74}$$

where KLD denotes the Kullback-Leibler divergence and $<>_q$ represents the expectation over the approximate distribution q and the prior $P_0(\boldsymbol{\theta}) = p_{\mathbf{x}}(\mathbf{x}|\Xi)p_{\boldsymbol{\xi}}(\boldsymbol{\xi})p_\gamma(\gamma|c,d)$. We shall further discuss here briefly the mean field VFE and Bethe free energy (BFE). Under the mean field (MF) approximation, where we consider that the q factorizes over the individual scalar parameters, we can obtain the approximate distribution as $q_{\theta_i}(\theta_i) \propto \exp(<\log p_{\mathbf{y}|\boldsymbol{\theta}}(\mathbf{y}|\boldsymbol{\theta})P_0(\boldsymbol{\theta})>_{q_{\theta_{\bar{i}}}(\boldsymbol{\theta}_{\bar{i}})})$. In [205], the authors show that the fixed points of the GAMP MP equations are the stationary points of the cost function termed approximate Bethe Free Energy, which is written below. This simplified form of the Bethe free energy is obtained using the same approximations which lead to GAMP from BP in the large system limit. We denote the MMSE estimate of x_i as $\widehat{x}_i$ and the posterior

Algorithm 16: GSwAMP-SBL

Input: $\mathbf{y}, \mathbf{A}$
Initialize: $\widehat{\gamma}_0 = \frac{c}{d}, \widehat{\xi}_0 = \frac{a}{b}, \sigma_m^2 = 1/(\|\mathbf{A}_m\|^2 \gamma + \widehat{\xi}_m), \forall m, \widehat{\mathbf{x}} = \mathbf{A}^T\mathbf{y}, V_k = 0, \forall k, w_k = 0.$
repeat
  for $k = 1$ to M **do**
    $g_k^{(t)} = \frac{y_k - w_k^{(t;N+1)}}{\frac{1}{\widehat{\gamma}} + V_k^{t;N+1}}$
    $V_k^{(t+1;1)} = \sum_m A_{km}^2 \sigma_m^{2(t)}$
    $w_k^{(t+1;1)} = \sum_m A_{k,m} \widehat{x}_m^{(t)} - V_k^{(t+1;1)} g_k^{(t)}$
  end for
  S = RandomPermute($[1, 2..., N]$)
  for $n = 1$ to N **do**
    $m = S_n$
    $\tau_m^{(t+1)} = \left[\sum_k \frac{A_{km}^2}{\frac{1}{\widehat{\gamma}} + V_k^{(t+1;k)}}\right]^{-1}$
    $r_m^{(t+1)} = \widehat{x}_m^{(t)} + \tau_m^{(t+1)} \sum_k A_{km} \frac{y_k - w_k}{\frac{1}{\widehat{\gamma}} + V_k^{(t+1;k)}}$
    $\widehat{x}_m^{(t+1)} = f_1(r_m^{(t+1)}, \tau_m^{(t+1)})$
    $\sigma_m^{2(t+1)} = f_2(r_m^{(t+1)}, \tau_m^{(t+1)})$
  end for
  Hyperparameter Estimation (using MF [168, Section 3]
  for $m = 1$ to N **do**
    $\widehat{\xi}_m^{(t+1)} = \frac{a+1/2}{|\widehat{\mathbf{x}}_m^{(t+1)}|^2 + \sigma_m^{2(t)}}, \widehat{\gamma}^{(t+1)} = \frac{c+N/2}{\frac{\|\mathbf{y}-\mathbf{A}\mathbf{x}\|^2}{2} + d}.$
  end for
until convergence

variance as σ_i^2.

$$\begin{aligned} \mathscr{F}_{GAMP}^{Bethe}(r_m, \tau_m, w_k, \widehat{x}_m, \sigma_m^2) = &-\sum_k \log \mathscr{Z}_k - \sum_m \frac{\sigma_m^2 + (\widehat{x}_m - r_m)^2}{2\tau_m} \\ &- \sum_k \frac{(w_k - \sum_m A_{km}\widehat{x}_m)^2}{2V_k} - \sum_m \log Z(r_m, \tau_m) \\ \text{with } V_k = &\sum_m A_{k,m}^2 \sigma_m^2, \\ \mathscr{Z}_k = &\int \frac{e^{-\frac{(w_k - z_k)^2}{2V_k}}}{\sqrt{2\pi V_k}} P_{y_k|z_k}(y_k|z_k) dz_k \end{aligned} \tag{9.75}$$

where $\mathbf{z} = \mathbf{A}\mathbf{x}$, with z_k being the k^{th} element. $Z(r_m, \tau_m)$ represents the normalization constant, which gets defined as

$$Z(r_m, \tau_m) = \int P_{x_m}(x_m|\xi_m) e^{-\frac{(x_m - r_m)^2}{2\tau_m}} dx_m. \tag{9.76}$$

By optimizing (9.75) alternatingly w.r.t $r_m, \tau_m, w_k, \widehat{x}_m, \sigma_m^2$, we reach the Algorithm 16, which is termed as sequential GAMP or Swept GAMP based SBL (GSwAMP-SBL). In Algorithm 16, the functions f_1, f_2 are defined as follows (which represent MMSE estimate in the Gaussian case as

in SBL)

$$f_1(r_m,\tau_m) = r_m \frac{\xi_m^{-1}}{\xi_m^{-1}+\tau_m}, \qquad f_2(r_m,\tau_m) = (\xi_m + \tau_m^{-1})^{-1}. \tag{9.77}$$

9.9 GSwAMP-SBL based Dynamic AR-SBL

Time varying sparse signal $\mathbf{x}_t$ is modeled using an AR(1) process with a diagonal correlation coefficient matrix $\mathbf{F}$, which can be written as follows

$$\begin{aligned} \text{State Update: } \mathbf{x}_t &= \mathbf{F}\mathbf{x}_{t-1} + \mathbf{w}_t, \\ \text{Observation: } \mathbf{y}_t &= \mathbf{A}^{(t)}\mathbf{x}_t + \mathbf{v}_t, \end{aligned} \tag{9.78}$$

where $\mathbf{x}_t = \left[x_{1,t},\ ...,\ x_{N,t}\right]^T$. Diagonal matrices $\mathbf{F}$ and Ξ are defined with its elements, $\mathbf{F}_{i,i} = f_i, f_i \in (-1,1)$ and $\Xi = \text{diag}(\boldsymbol{\xi}), \boldsymbol{\xi} = [\xi_1,...\xi_N]$. Further, $\mathbf{w}_t \sim \mathscr{CN}(\mathbf{0}, \boldsymbol{\Lambda}^{-1})$, where $\boldsymbol{\Lambda}^{-1} = \Xi^{-1}(\mathbf{I} - \mathbf{F}\mathbf{F}^H) = \text{diag}(\frac{1}{\lambda_1},..., \frac{1}{\lambda_N})$ and $\mathbf{v}_t \sim \mathscr{CN}(\mathbf{0}, \frac{1}{\gamma}\mathbf{I})$. $\mathbf{w}_t$ are the complex Gaussian mutually uncorrelated state innovation sequences. Hence we sparsify the prediction error variance $\mathbf{w}_t$ also, with the same support as $\mathbf{x}_0$ and henceforth enforces the same support set for $\mathbf{x}_t, \forall\, t$. $\mathbf{v}_t$ is independent of the $\mathbf{w}_t$ process. Although the above signal model seems simple, there are numerous applications such as 1) Bayesian adaptive filtering [185], 2) Wireless channel estimation: multipath parameter estimation as in [186]. In this case, $\mathbf{x}_t$ = FIR filter response, and Ξ represents example the power delay profile. We also denote the unknown parameter vector $\boldsymbol{\theta}_t = \{\mathbf{x}_t, \boldsymbol{\Lambda}, \gamma, \mathbf{F}\}$ and θ_i represents each scalar in $\boldsymbol{\theta}$. Note that we only estimate the reparametrized innovation sequence precision instead of the precision variables ξ_i.

9.10 GSwAMP-SBL for Nonlinear Kalman Filtering

The joint distribution $p(\mathbf{y}_t, \boldsymbol{\theta}_t | \mathbf{y}_{1:t-1})$ can be written as ($\boldsymbol{\Sigma}_{t|t-1}$ represents the diagonal prediction covariance matrix)

$$\begin{aligned} \ln p(\mathbf{y}_t, \boldsymbol{\theta}_t | \mathbf{y}_{1:t-1}) = \frac{N}{2}\ln\gamma - \frac{\gamma}{2}\left||\mathbf{y}_t - \mathbf{A}_t\mathbf{x}_t\right||^2 + -M\det(\widehat{\boldsymbol{\Sigma}}_{t|t-1}) - \frac{1}{2}\left(\mathbf{x}_t - \widehat{\mathbf{x}}_{t|t-1}\right)^T \widehat{\boldsymbol{\Sigma}}_{t|t-1}^{-1}\left(\mathbf{x}_t - \widehat{\mathbf{x}}_{t|t-1}\right) \\ + (c-1)\ln\gamma + c\ln d - d\gamma + \text{constants}. \end{aligned} \tag{9.79}$$

9.10.1 Diagonal AR(1) (DAR(1)) Prediction Stage

In the prediction stage, similar as in KF, we compute the posterior, $p(\mathbf{x}_t|\mathbf{y}_{1:t-1})$, where $\mathbf{y}_{1:t-1}$ refers to the observations till time $t-1$. For more detailed derivation, we refer to our previous work [187] or to Section 9.2.2.1 in this thesis. This part gets computed using MF, however, the interation between $x_{m,t}$ and f_m requires Gaussian projection, using expectation propagation (EP) [187]. The resulting Gaussian distribution is parameterized as $x_{l,t} \sim \mathscr{N}(\widehat{x}_{l,t|t-1}, \sigma^2_{l,t|t-1})$.

9.10.2 Measurement Update (Filtering) Stage

For the measurement update stage, the posterior for $\mathbf{x}_t$ is inferred using GSwAMP-SBL in Algorithm 16. The posterior mean and diagonal covariance matrix of the estimate computed at $\mathbf{x}_t$

are denoted by $\widehat{\mathbf{x}}_{t|t}, \mathbf{\Sigma}_{t|t}$. We denote each entries in $\widehat{\mathbf{x}}_{t|t}$ as $\widehat{\mathbf{x}}_{l,t|t}$ respectively. In the measurement stage, the prior for $\mathbf{x}_t$ gets replaced by the posterior estimate from the prediction stage. We refer to our previous work [198] for detailed discussions on the fitering stage. One remark here is that compared to our previous work using BP in [198], using GSwAMP gives a more computationally feasible implementation and accurate posterior variances, where $\sigma^2_{l,t|t}$ incorporates the effect of all $\sigma^2_{l',t|t}, l' \neq l$. $\sigma^2_{l,t|t}$ represents the diagonal elements of the posterior covariance matrice $\mathbf{\Sigma}_{t|t}$.

9.10.3 Lag-1 Smoothing Stage

We obtain the system model for the smoothing stage (by combining the AR(1) stage in the measurement model) as follows

$$\begin{aligned} \mathbf{y}_t &= \mathbf{A}^{(t)}\mathbf{F}\mathbf{x}_{t-1} + \widetilde{\mathbf{v}}_t, \\ \text{where } \widetilde{\mathbf{v}}_t &= \mathbf{A}^{(t)}\mathbf{w}_{t-1} + \mathbf{v}_t, \end{aligned} \tag{9.80}$$

where $\widetilde{\mathbf{v}}_t \sim \mathscr{CN}(0, \widetilde{\mathbf{R}}_t)$ with $\widetilde{\mathbf{R}}_t = \mathbf{A}^{(t)}\mathbf{\Lambda}^{-1}\mathbf{A}^{(t)\,H} + \frac{1}{\gamma}\mathbf{I}$. We show in [198, Lemma 1] that KF is not enough to adapt the hyperparameters, instead we need at least a lag 1 smoothing (i.e. the computation of $\widehat{\mathbf{x}}_{t-1|t}, \mathbf{\Sigma}_{t-1|t}$ through GSwAMP-SBL). Here, we first do a noise whitening by multiplying $\mathbf{y}_t$ with $\widetilde{\mathbf{R}}_t^{-1/2}$. Hence, we can rewrite the observation model as

$$\begin{aligned} \widehat{\mathbf{y}}_t &= \widehat{\mathbf{A}}^{(t)}\mathbf{x}_{t-1} + \widehat{\mathbf{v}}_t, \\ \text{where } \widehat{\mathbf{v}}_t &= \widetilde{\mathbf{R}}_t^{-1/2}\widetilde{\mathbf{v}}_t, \\ \widehat{\mathbf{A}}^{(t)} &= \widetilde{\mathbf{R}}_t^{-1/2}\mathbf{A}^{(t)}\mathbf{F}. \end{aligned} \tag{9.81}$$

The joint distribution can be factorized as, $p(\mathbf{y}_t, \boldsymbol{\theta}|\mathbf{y}_{1:t-1}) = p(\mathbf{y}_t|\boldsymbol{\theta}_t)\, p(\mathbf{x}_{t-1}|\mathbf{y}_{1:t-1})\, p(\mathbf{F}, \mathbf{\Lambda}, \gamma|\mathbf{y}_{1:t-1})$.

$$\begin{aligned} \ln p(\mathbf{y}_t, \boldsymbol{\theta}_{t-1}|\mathbf{y}_{1:t-1}) = \frac{-1}{2}\ln\det\widetilde{\mathbf{R}}_t - |f_m|^2|x_m|^2\mathbf{A}_m^{(t)\,T}\widetilde{\mathbf{R}}_t^{-1}\mathbf{A}_m^{(t)} \\ + 2\Re(f_m^H x_m^H \mathbf{A}_m^{(t)\,H}\widetilde{\mathbf{R}}_t^{-1}(\mathbf{y}_t - \mathbf{A}_{\tilde{m}}^{(t)}\mathbf{F}_{\tilde{m}}\mathbf{x}_{\tilde{m},t})) + c_f, \end{aligned} \tag{9.82}$$

where c_f being the terms independent of f_m, $\mathbf{A}_{\tilde{m}}^{(t)}, \mathbf{x}_{\tilde{m},t}$ represents the matrix $\mathbf{A}^{(t)}$ or the vector $\mathbf{x}_t$ with m^{th} column or element removed. Note that we propose to compute $\widetilde{\mathbf{R}}_t$ by substituting the point estimates of $\mathbf{\Lambda}, \gamma$. We also define $\widehat{\mathbf{F}}_{\tilde{m}|t} = \text{diag}\,(\widehat{f}_{n|t}, n \neq m)$ with m^{th} element removed. Further applying the MF rule, we write the mean and variance of the resulting Gaussian distribution for f_m as

$$\begin{aligned} \sigma^{-2}_{f_m|t} &= (|\widehat{x}_{m,t-1|t}|^2 + \sigma^2_{m,t-1|t})\mathbf{A}_m^{(t)\,T}\widetilde{\mathbf{R}}_t^{-1}\mathbf{A}_m^{(t)}, \\ \widehat{f}_{m|t} &= \sigma^2_{f_m|t}\widehat{x}^H_{m,t-1|t}\mathbf{A}_m^{(t)\,H}\widetilde{\mathbf{R}}_t^{-1}(\mathbf{y}_t - \mathbf{A}_{\tilde{m}}^{(t)}\widehat{\mathbf{F}}_{\tilde{m}|t}\widehat{\mathbf{x}}_{\tilde{m},t-1|t}). \end{aligned} \tag{9.83}$$

9.11 Simulation Results

To elucidate further the excellent convergence proporties of the GSwAMP-SBL algorithm from other state of the art AMP-SBL versions, we evaluate the normalized MSE (NMSE) performance under different scenarios of $\mathbf{A}$ matrices such as ill-conditioned, non-zero mean matrices for static SBL. We also illustrate the performance of the BP based DAR-SBL compared to our suboptimal methods which are based on MF. Note that simulations are performed with dimensions of $\mathbf{A}$, $M = 150, N = 250$. The power delay profile (variances of x_i) for the SBL model in Section 9.9 is chosen as d^{i-1}, with $d = 0.93$ and starting with index $i = 1$. Further we analysis the following scenarios in the simulations. In Figure 9.6 and Figure 9.7, we also assume that the hyperparameters are unknown and get estimated as proposed in our Algorithm 17.

Algorithm 17: GSwAMP based DAR-SBL

Initialization $\widehat{f}_{l|0}, \widehat{\lambda}_{l|0} = \frac{a}{b}, \widehat{\gamma}_0 = \frac{c}{d}, \widehat{x}_{l,0|0} = 0, \sigma^2_{l,0|0} = 0, \forall l$. Define $\mathbf{\Sigma}_{t-1|t-1} = \text{diag}(\sigma^2_{l,t|t-1})$.
for $t = 1 : T$ do
Prediction Stage:

1. Compute $\widehat{x}_{l,t|t-1}, \sigma^2_{l,t|t-1}$ using EP and MF [198], $\widehat{x}_{l,t|t-1} = \widehat{f}_{l|t-1}\widehat{x}_{l,t-1|t-1}$, $\sigma^2_{l,t|t-1} = |\widehat{f}_{l|t-1}|^2\sigma^2_{l,t-1|t-1} +$
$\sigma^2_{f_l|t-1}(|\widehat{x}_{l,t-1|t-1}|^2 + \sigma^2_{l,t-1|t-1}) + \widehat{\lambda}^{-1}_{l|t-1}$.

Filtering Stage:

1. Compute $\widehat{x}_{l,t|t}, \sigma^{-2}_{l,t|t}$ using GSwAMP (iterated convergence).

Smoothing Stage:
Initialization: $\mathbf{\Sigma}^{(0)}_{t-1|t} = \mathbf{\Sigma}_{t-1|t-1}, \widehat{\mathbf{x}}^{(0)}_{t-1|t} = \widehat{\mathbf{x}}_{t-1|t-1}$. Update $\widetilde{\mathbf{R}}_t, \widehat{\mathbf{A}}^{(t)}$.

1. Compute $\widehat{\mathbf{x}}_{t-1|t}, \mathbf{\Sigma}_{t-1|t}$ using GSwAMP (iterated until convergence)

Estimation of hyperparameters (Define: $x'_{k,t} = x_{k,t} - f_k x_{k,t-1}, \zeta_t = \beta\zeta_{t-1} + (1-\beta) < \left\|\mathbf{y}_t - \mathbf{A}^{(t)}\mathbf{x}_t\right\|^2 >$**) :**

1. Compute $\widehat{f}_{l|t}, \sigma^2_{f_{l|t}}$ from (9.83), $\widehat{\gamma}_t = \frac{c+N}{(\zeta_t+d)}$ and $\lambda_{l|t} = \frac{(a+1)}{(<\left|x'_{k,t}\right|^2>_{|t}+b)}$.

9.11.1 ill-conditioned A case:

We construct the matrix $\mathbf{A}$ with condition number $\kappa > 1$. Let $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$, where $\mathbf{U}, \mathbf{V}^T$ are the left and right singular vectors of an i.i.d-Gaussian matrix. Further, we select the singular values such that $\frac{\Sigma_{i,i}}{\Sigma_{i,i+1}} = \kappa^{1/(M-1)}$, for $i = 1, 2..., M-1$ and $\Sigma_{i,i}$ is the i^{th} diagonal element of $\mathbf{\Sigma}$. The more the condition number, the more $\mathbf{A}$ deviates from the i.i.d-Gaussian case. In Figure 9.6, we plot the NMSE values as a function of the condition number for different algorithms such as original SBL (LMMSE-SBL), SAVE-SBL [168], proposed GSwAMP-SBL and damped GAMP-SBL [166]. In fact, in the simulations we observed that GAMP-SBL does not converge without using damping and there does not exist any closed form solution for the optimal damping value. Hence depending on the particular scenario being considered and also on the dimensions, the damping value may change. However, the proposed GSwAMP-SBL is more robust in the sense that it does not require any damping and convergence to a local optimum is guaranteed.

9.11.2 Non-zero mean A case:

In this case, we generate each entries of $\mathbf{A}$ as i.i.d Gaussian with a non-zero mean, $A_{i,j} \sim \mathcal{N}(\mu, \frac{1}{M})$. We plot the NMSE performance for different algorithms in Figure 9.7 as a function of the mean of $\mathbf{A}$. We observe that GAMP-SBL does not converge, in this case apart from damping we may require mean removal procedure also as in noted in [203]. However, SAVE-SBL and the proposed GSwAMP-SBL converges without any mean removal procedure. Hence, GSwAMP-SBL would be preferred from an implementation complexity perspective. SAVE-SBL has the incorrect posterior variance issue which we have observed in our previous papers.

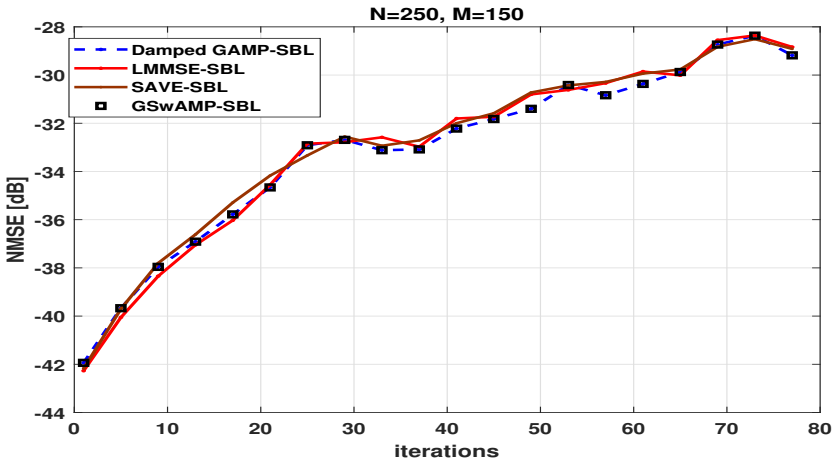


Figure 9.6: NMSE vs Condition number of the measurement matrix **A**.

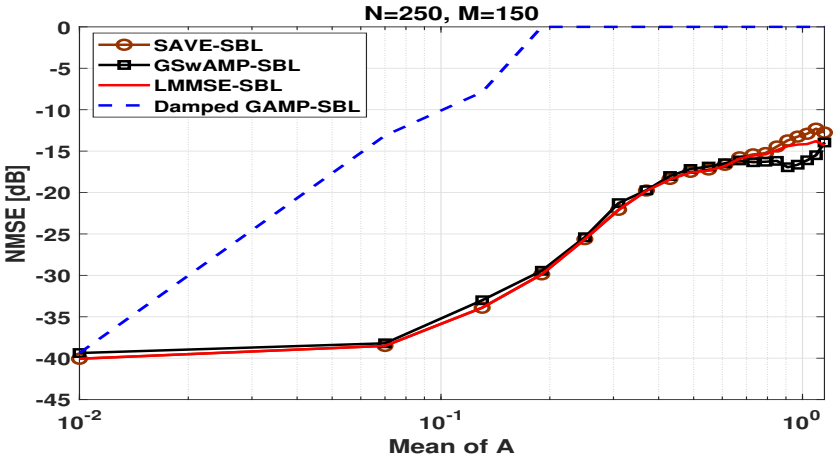


Figure 9.7: NMSE vs the mean of **A**.

9.11.3 Rank Deficient A case (Figure 9.8):

We generate a rank deficient **A** by $\mathbf{A} = \frac{1}{N}\mathbf{HG}$, with $\mathbf{H} \in \mathscr{R}^{M\times R}, \mathbf{G} \in \mathscr{R}^{R\times N}$ and $R < M$. The entries of $\mathbf{H}, \mathbf{G}$ are generated as i.i.d Gaussian with zero mean and unit variance. The rank ratio $\frac{R}{N}$ indicates the deviation of **A** from i.i.d Gaussian case.

9.12 Conclusions

In this chapter, we look at the robustness of the SBL algorithm under deviations from i.i.d Gaussian assumptions of measurement matrix. Towards this direction, we propose a GSwAMP-SBL algorithm which implements the GAMP sequentially rather than parallel as in the original GAMP version by Rangan [143]. Among the many techniques proposed for improving the convergence properties of AMP algorithms, GSwAMP stands out due to the low cost per iteration, compared to the highly complex nature of damping or mean removal based algorithms in the literature. We also integrate hyperparameter estimation (by MF) and an extension of the GSwAMP-SBL for a time varying sparse signal is also proposed.

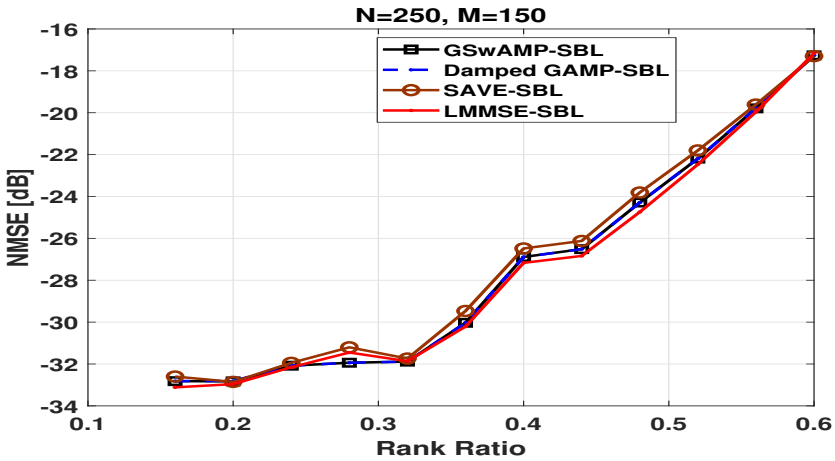


Figure 9.8: NMSE vs the rank ratio.

9.12.1 Conclusions and Perspectives

Conclusions and Perspectives 10

- There are many Bayesian estimation problems, many of which are LMMSE (Wiener, Kalman), which contain hyperparameters to be tuned, using various approaches.
- BP-SBL computational complexity is similar to AMP (except that BP-SBL may have more memory requirements due to large number of messages being passed at any iteration) if we consider Gaussian BP. In SBL, it is Gaussian BP for a fixed estimate of hyperparameters. But BP will have higher computational complexity if we go to more general measurement models or prior distributions. xAMP variants does not start with Gaussian prior or posterior. Hence we remark that for possibly nonlinear measurement model or more general priors or even for the case of joint hyperparameter plus sparse state vector estimation $(\widehat{\mathbf{x}}, \boldsymbol{\xi})$, our MP based SBL algorithms are more relevant.
- In message passing (approximate iterative) based inference techniques, we can conclude that it is easy to get the mean (estimate) correct but more difficult to get correct posterior variances.
- MP based approximate inference algorithms can be unified under free energy optimization framework. We used mCRB formulation for split in various MP simplification levels and it allows performance-complexity trade-off.
- We relied on large system analysis for yielding simplified asymptotic performance analysis, allowing to show Bayes optimality for some special cases and to justify algorithmic simplifications.
- Proposed new versions of AMP for SBL such as GSwAMP-SBL which is seen to converge for measurement matrices which deviate from i.i.d model for $\mathbf{A}$.
- A very good overview of the state of the art on fast sparse Bayesian techniques proposed here can also be found in [206]. Apart from the topics discussed here in, the paper also gives an overview of other competing methods (similar to SBL) such as Stein's

Conclusions and Perspectives 10 (cont.)

unbiased risk estimator (SURE), empirical Bayes and kernel based hyperparameter estimation.

Chapter 10

SPARSE BAYESIAN LEARNING FOR TENSOR SIGNAL PROCESSING

In many applications such as Multiple Input Multiple Output (MIMO) radar [207], massive MIMO channel estimation [132], image, and video processing, etc., the received signals are multidimensional (i.e tensors). Moreover, these signals can be represented as a low rank tensor. To fully exploit the structure of such signals, tensor decomposition methods such as CANDECOMP/PARAFAC (CP) [208, 209] or Canonical Polyadic Decomposition (CPD) [210] have been introduced. Explicitly accounting for this tensorial structure can be more beneficial than the matricized or vectorized representations of the data since the matrix decomposition cannot fully exploit the multi-dimensional subspace structure of the data. One initial work in this direction is based on the concept of multi-dimensional SVD applied to multi-dimensional harmonic retrieval problems [211]. In this paper, we consider a generalized problem in which the dictionary matrix can be factorized as a Kronecker product [212], the received tensor signal $\mathbf{Y}$ can be represented as

$$\mathbf{y} = (\mathbf{A}_1 \otimes \mathbf{A}_2 .... \otimes \mathbf{A}_N)\mathbf{x} + \mathbf{w}, \tag{10.1}$$

where $\mathbf{y} = vec(\mathbf{Y})$, $\otimes$ represents the Kronecker product between two matrices, $vec(\cdot)$ representing the vectorized version of the tensor or matrix $(\cdot)$, $\mathbf{Y} \in \mathscr{C}^{I_1 \times I_2 ... \times I_N}$ is the observations or data, $\mathbf{A}_{j,i} \in \mathscr{C}^{I_j}$, the factor matrix $\mathbf{A}_j = [\mathbf{A}_{j,1}, ..., \mathbf{A}_{j,P_j}]$ which is unknown and the tensor product is represented by $[[\mathbf{A}_1, ..., \mathbf{A}_N; \mathbf{x}]]$, $\mathbf{x}$ is the $M(= \prod_{j=1}^{N} P_j)$-dimensional sparse signal and $\mathbf{w}$ is the additive noise. $\mathbf{x}$ contains only K non-zero entries, with $K << M$ and thus the dictionary matrix to be learned allows a low rank representation. $\mathbf{w}$ is assumed to be a white Gaussian noise, $\mathbf{w} \sim \mathscr{N}(0, \gamma^{-1}\mathbf{I})$. To address this problem when the dictionary matrix is known, a variety of algorithms such as the orthogonal matching pursuit [135], the basis pursuit method [136] and the iterative re-weighted l_1 and l_2 algorithms [137] exist in the literature. The SBL introduced by [2, 139], is developed around a sparsity-promoting prior for $\mathbf{x}$, whose realizations are softly sparse in a sense that most entries are small in magnitude and close to zero.

CPD can be viewed as a general extension of the singular value decomposition (SVD) to the high-order tensors, with the difference that the factor matrices need not be orthogonal. In certain applications such as wireless channel estimation, these factors have specific forms such as Vandermonde or Toeplitz or Hankel. To find the tensor factor matrices, the most popular solution is the alternating least squares (ALS) [213], which iteratively optimizes one factor matrix at a time while keeping the others fixed. Most of the existing algorithms [214–218] focus on either maximum likelihood based schemes, LS or K-SVD algorithms. Knowledge of tensor rank is a prerequisite to implement these algorithms and it takes large number of iterations for them to

converge. Moreover, classical algorithms do not take account the potential statistical knowledge of the factor matrices. While we focus on a Bayesian approach to the estimation of the factor matrices in this chapter, with automatic relevance determination.

Many research papers had looked at various model mismatches and impairments affecting channel estimation over the last couple of decades. These mismatched models and hardware impairments (such as clock offsets, timing advance jitter) are the motivation for the rather non-parametric approach that we follow in this tensor work. For the antenna array responses, there is the calibration, mutual coupling, individual antenna responses, etc that make those academic models such as the Vandermonde vectors for ULAs deviate significantly from reality. A residual carrier offset can be handled by the AR(1) temporal model for the multipath complex gains, though a parsimonious parameterization would dictate that the phase in the AR(1) prediction coefficient would be the same for all multipath components, if they are affected by a common frequency offset. The path delay which would lead to a Vandermonde vector of phase shifts in the subcarrier domain, is affected by the unknown Tx/Rx filter diagram which within its passband can perhaps be ignored. In any case, clock drift would lead to temporal variation of that delay component, which can be handled non-parametrically by introducing some forgetting factor in the Kalman filter to allow for (slower) temporal variations in the model (dictionary) that are not captured by the AR(1) model of the path gains. Of course, a (more) parsimonious parameterization would always lead to better estimation quality, but then one has to be sure of the parsimonious model accuracy. We feel that in a (doubly) dispersive MIMO channel estimation scenario, the many dimensions to the problem already lead to the potential of good quality estimation by just exploiting the Kronecker structure.

Another interesting point here is about the better identifiability results by exploiting the Kronecker or Khatri-Rao structured factor matrices. If we have 3 factors, each of dimension N_i, then exploiting the Kronecker structure reduces the number of parameters per rank 1 term in a CPD model from $(\prod_{i=1}^{3} N_i) - 1$ to $\sum_{i=1}^{3}(N_i - 1)$. If each N_i is for example, 10, this goes down from 1000 to 30. The surprising thing is that starting at 3 factors, the CPD becomes essentially unique, i.e., identifiable, which is pretty amazing by itself. Other possibilities are to use parametric expressions for some factors, example, for the delay dimension. It is also to be noted that even though in our work here, we do not consider any parametric forms for the factor matrices, these factors could have a more parsimonious representation with parametric factors such as Vandermonde. We do start with the Khatri-Rao structured matrices which can correspond to the case of a discretized (grid based) version of the dictionary matrix [200] and further refine this model to consider Kronecker structured dictionary matrices. In this thesis, we restrict our attention to the Kronecker structured case only.

10.1 Summary of this Chapter

- We propose novel Space Alternating Variational Estimation based SBL techniques with Khatri-Rao structured and KS dictionary learning called SAVED [200] and SAVED-KS [125, 219], respectively, advancing the SAVE methods which we introduced in [168, 220, 221] which assumed a known dictionary.

- We also propose a joint VB version for the KS dictionary matrix factors which has a better performance compared to SAVED-KS, but at the cost of an increase in computational complexity.

- We also discuss the local identifiability using the non-singularity of the Fisher information matrix (FIM) for KS DL in an SBL setting.
- Simulation results suggest that the proposed solution has a faster convergence rate (and hence lower complexity) than (even) the classical ALS and furthermore has lower reconstruction MSE in the presence of noise.

10.1.1 Tensor Notations

An N-way tensor is represented using a calligraphic font. For example, $\mathscr{X} \in \mathscr{C}^{I_1 \times \cdots \times I_N}$, where I_j represents the dimension along the $j-th$ mode. A Tucker model for the tensor model here can be represented as [213]

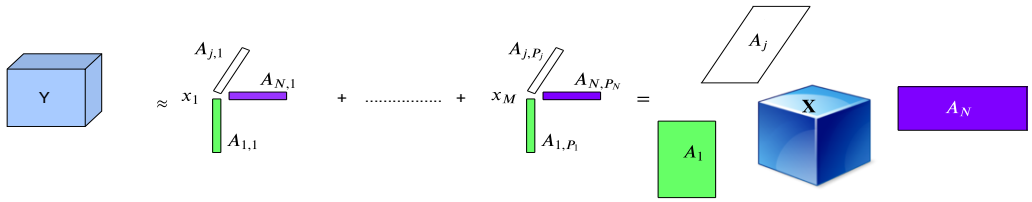
$$\begin{aligned} \mathscr{Y} &= \mathscr{X} \times_1 \mathbf{A}_1 \times_2 \mathbf{A}_2 \cdots \times_N \mathbf{A}_N + \mathscr{W} \\ &= [[\mathscr{X}; \mathbf{A}_1, \mathbf{A}_2 \cdots \mathbf{A}_N]]. \end{aligned} \tag{10.2}$$

Each $\mathbf{A}_j$ is of dimension $I_j \times P_j$. The matricized version of (10.2) (after unfolding along $n-th$ mode) can be represented as

$$\mathbf{Y}^{(n)} = \mathbf{A}_n \mathbf{G}^{(n)} (\mathbf{A}_N \otimes \cdots \otimes \mathbf{A}_{n+1} \otimes \cdots \otimes \mathbf{A}_1)^T + \mathbf{W}^{(n)}. \tag{10.3}$$

CP can be viewed as a special case of Tucker above with $P_1 = P_2 = \cdots = P_N$.

10.2 Hierarchical Probabilistic Model



In the following sections, we represent (10.1) using the tensor decomposition properties from [213]. Let $Y_{i_1,\ldots,i_N}$ represents the $(i_1, i_2, \ldots, i_N)^{th}$ element of the tensor and $\mathbf{y} = [y_{1,1,\ldots,1}, y_{1,1,\ldots,2} \ldots, y_{I_1,I_2,\ldots,I_N}]^T$, then it can be verified that [222]

$$\mathbf{y} = (\mathbf{A}_1 \otimes \mathbf{A}_2 \ldots \otimes \mathbf{A}_N)\mathbf{x} + \mathbf{w}, \tag{10.4}$$

where $\otimes$ represents the Khatri-Rao product between two matrices, $\mathbf{y} \in \mathscr{C}^{(\prod_{i=1}^{N} I_i) \times 1}$ and we denote $\mathbf{A} = \mathbf{A}_1 \otimes \mathbf{A}_2 \ldots \otimes \mathbf{A}_N$. Factor matrix $\mathbf{A}_j$ is of dimension $I_j \times P_j$. In the tensor representation, the notations $\mathscr{Y}, \mathscr{X}$ represent $\mathbf{y}, \mathbf{x}$, respectively. Compared to the case of Khatri-Rao structured dictionary matrices [200] (which indeed would correspond to CPD), the KS case can represent the possible coupling between different factor matrices. To illustrate this, consider the massive MIMO channel estimation problem outlined in Section 10.2.1. It is possible that there exists two paths with same delay response but have different AoA/AoD. This kind of correlation or coupling between different paths can be represented by the KS (Tucker representation) case we consider here. Moreover, not all the possible couplings or path combinations will be present which indeed leads to a sparsification of the tensor $\mathscr{X}$. Since the sparsity measure (number of nonzero components) of $\mathbf{x}$ is unknown, the following VB-SBL algorithm performs automatic rank determination. In Bayesian compressive sensing, a two-layer hierarchical prior is assumed for the $\mathbf{x}$

as in [2]. The hierarchical prior is chosen such that it encourages the sparsity property of $\mathbf{x}$. $\mathbf{x}$ is assumed to have a Gaussian distribution parameterized by $\boldsymbol{\xi} = [\xi_1\ \xi_2\ ...\ \xi_M], \xi_i > 0$ and real, where ξ_i represents the inverse variance or the precision parameter of x_i.

$$\begin{aligned} p(\mathbf{x}|\boldsymbol{\xi}) &= \prod_{i=1}^{M} p(x_i|\xi_i) \\ &= \prod_{i=1}^{M} \mathscr{CN}(0, \xi_i^{-1}). \end{aligned} \tag{10.5}$$

Further a Gamma prior is considered over $\boldsymbol{\xi}$

$$\begin{aligned} p(\boldsymbol{\xi}) &= \prod_{i=1}^{M} p(\xi_i|a,b) \\ &= \prod_{i=1}^{M} \Gamma^{-1}(a) b^a \xi_i^{a-1} e^{-b\xi_i}. \end{aligned} \tag{10.6}$$

The inverse of noise variance, $\gamma > 0$ and real, is also assumed to have a Gamma prior, $p(\gamma) = \Gamma^{-1}(c) d^c \xi_i^{c-1} e^{-d\gamma}$. Now the likelihood distribution can be written as

$$p(\mathbf{y}|\mathbf{x},\gamma) = (2\pi)^{-N} \gamma^N e^{-\gamma||\mathbf{y}-\mathbf{A}\mathbf{x}||^2}. \tag{10.7}$$

We consider factor matrices to be unstructured because the parametric forms are uncertain. For example, in massive MIMO channel estimation [223], the array response at the mobile station (MS) is not exploitable. Even the array response at the base station (BS) will typically require calibration to be exploitable. Doppler shifts are clear Vandermonde vectors. Delays could be more or less clear if one goes to the frequency domain in OFDM, and one only takes into account the range of subcarriers for which the Tx/Rx filters can be considered f-flat. Then over those subcarriers, it is also Vandermonde. Let $\mathbf{A}_{j,i}$ represents the i^{th} column of $\mathbf{A}_j$. For the unstructured factor matrices also, we consider $\mathbf{A}_{j,i} = [1\, \mathbf{a}_{j,i}^H]^H$ and further $\mathbf{a}_{j,i}$ is unconstrained and deterministic (in all the Vandermonde cases, it is perfect, or in all cases of phasors). Assuming first entry to be 1 is even better than $\|\mathbf{A}_{j,i}\| = 1$ because $\|\mathbf{A}_{j,i}\| = 1$ still leaves a phase ambiguity. With first entry= 1, the factors are unique, up to permutation in the sum of terms of course.

We define the unfolding operation on an N^{th} order tensor $\mathbf{Y} = [[\mathbf{A}_1, ..., \mathbf{A}_N; \mathbf{x}]]$ as [213] ($\mathbf{Y}^{(n)}$ is of size $I_n \times \prod_{i=1, i\neq n}^{N} I_i$ below, $\mathbf{X} = \text{diag}(\mathbf{x})$)

$$\mathbf{Y}^{(n)} = \mathbf{A}_n \mathbf{X} (\mathbf{A}_N \otimes \mathbf{A}_{N-1} ... \mathbf{A}_{n+1} \otimes \mathbf{A}_{n-1} ... \otimes \mathbf{A}_1)^T. \tag{10.8}$$

10.2.1 Application-Multipath Wireless Channel Estimation

We get for the matrix impulse response of a time-varying frequency-selective MIMO channel $\mathbf{H}(t,\tau)$ [148], In the case of distributed antenna systems (near field), or very wideband regime, the array responses become a function of the position parameters of the (last) path scatterers. The fast variation of the phase in $e^{j2\pi f_i t}$ and possibly the variation of the A_i (when the nominal path represents, in fact, a superposition of paths with similar parameters) correspond to the fast fading. All the other parameters (including the Doppler frequency) vary on a slower time scale and correspond to slow fading.

The channel impulse response $\mathbf{H}$ has per path a rank one contribution in four dimensions (Tx and Rx spatial multi-antenna dimensions, delay spread, and Doppler spread) [223]. Hence, going to the frequency domain, we get

$$\text{vec}(\mathbf{H}(1:t, f_1:f_2)) = \sum_{i=1}^{L} A_i \, \mathbf{h}_t(\psi_i) \otimes \mathbf{h}_r(\phi_i) \otimes \mathbf{v}_f(\tau_i) \otimes \mathbf{v}_t(f_i). \tag{10.9}$$

where $\mathbf{v}_f(.)$, $\mathbf{v}_t(.)$ are appropriate Vandermonde vectors (possibly subsampled in the case of $\mathbf{v}_f(.)$). Hence we get a sum of rank one $4D$ tensors. $\mathbf{h}_r$, $\mathbf{h}_t$ could themselves have a Kronecker structure in the case of polarization or the case of $2D$ antenna arrays with separable structure [224]. In the model above, each of the four Kronecker factors is assumed to be parametric. For instance, $\mathbf{h}_t(.)$ is also a Vandermonde vector in the case of a basic Uniform Linear Array depending on azimuth only, neglecting antenna coupling. Whereas more generally $\mathbf{h}_t(.)$ may be known or learned at the BS side, it is less reasonable to assume a parametric form for $\mathbf{h}_r$ on the UE side, especially in the case of a hand-held device (orientation, way of holding it). In the following sections, we represent (10.9) using the tensor decomposition properties from [213]. Let $Y_{i_1,\ldots,i_N}$ represents the $(i_1 i_2 \ldots i_N)^{th}$ element of the tensor (after correlating with the pilot symbols) and the vectorized version $\mathbf{y} = [y_{1,1,1,,1}, y_{1,1,1,2}\ldots., y_{I_1,I_2,I_3,I_4}]^T$, then it can be verified that [222]

$$\begin{aligned} \mathbf{y} &= (\mathbf{H}_t \otimes \mathbf{H}_r \otimes \mathbf{V}_f \otimes \mathbf{V}_t)\mathbf{x} + \mathbf{w} \\ &= \mathbf{A}\mathbf{x} + \mathbf{v}, \\ \text{where } \mathbf{A} &= (\mathbf{H}_t \otimes \mathbf{H}_r \otimes \mathbf{V}_f \otimes \mathbf{V}_t), \\ \mathbf{x} &= [A_1, \ldots, A_M]^T, \end{aligned} \tag{10.10}$$

where $\mathbf{H}_t, \mathbf{H}_r, \mathbf{V}_f, \mathbf{V}_t$ represent the matrices with sizes $I_1 \times M, I_2 \times M, I_3 \times M, I_4 \times M$ respectively and the columns represent the vectors $\mathbf{h}_t(\psi_i), \mathbf{h}_r(\phi_i), \mathbf{v}_f(\tau_i), \mathbf{v}_t(f_i)$. Since the actual number of multi-paths is not known, we assume that $M \gg L$ and the following SBL based compressed sensing method performs automatic rank determination. Also, we denote $N = \prod_{i=1}^{4} I_i$.

10.3 Variational Bayesian Inference for Joint Dictionary Learning and Sparse Signal Recovery

The computation of the posterior distribution of the parameters is usually intractable. In order to address this issue, in VB framework, the posterior distribution $p(\mathbf{x}, \boldsymbol{\xi}, \gamma, \mathbf{A}|\mathbf{y})$ is approximated by a variational distribution $q(\mathbf{x}, \boldsymbol{\xi}, \gamma, \mathbf{A})$ that has the factorized form (we define $\boldsymbol{\theta} = (\mathbf{x}, \boldsymbol{\xi}, \gamma, vec(\mathbf{A}))$ as the vector of all parameters to be estimated and $\boldsymbol{\theta}_k$ represents any subset which is statistically independent from others in the approximate posterior below):

$$\begin{aligned} q(\mathbf{x}, \boldsymbol{\xi}, \gamma, \mathbf{A}) &= q_\gamma(\gamma) \prod_{i=1}^{M} q_{x_i}(x_i) \prod_{i=1}^{M} q_{\xi_i}(\xi_i) \prod_{i=1}^{M} \prod_{j=1}^{N} q_{\mathbf{a}_{j,i}}(\mathbf{a}_{j,i}) \\ &= \prod_k q_k(\boldsymbol{\theta}_k) \end{aligned} \tag{10.11}$$

In the equation (10.11) above, which describes the form of the approximate posterior, the parameters $\gamma, x_i, \xi_i, \mathbf{a}_{j,i}, \forall j, i$ are assumed to be independent. If all the path responses are independent, this factorization may become optimal or match the true posterior. On the other hand, once the path responses become correlated, for example, two paths with same delay but with

different AoA/AoD can be correlated. In this case, the above factorization may be quite sub-optimal leading to a performance gap compared to more complex methods such as joint VB based DL which we consider in the later sections here. VB compute the factors q by minimizing the Kullback-Leibler distance between the true posterior distribution $p(\mathbf{x},\boldsymbol{\xi},\gamma,\mathbf{A}|\mathbf{y})$ and the $q(\mathbf{x},\boldsymbol{\xi},\gamma,\mathbf{A})$. From [146]

$$KLD_{VB} = KL\big(p(\mathbf{x},\boldsymbol{\xi},\gamma,\mathbf{A}|\mathbf{y})||q(\mathbf{x},\boldsymbol{\xi},\gamma,\mathbf{A})\big) \tag{10.12}$$

The KL divergence minimization is equivalent to maximizing the evidence lower bound (ELBO) [147]. To elaborate on this, we can write the marginal probability of the observed data as

$$\begin{aligned} \ln p(\mathbf{y}) &= L(q) + KLD_{VB}, \text{ where,} \\ L(q) &= \int q(\boldsymbol{\theta}) \ln \frac{p(\mathbf{y},\boldsymbol{\theta})}{q(\boldsymbol{\theta})} d\boldsymbol{\theta}, \\ KLD_{VB} &= -\int q(\boldsymbol{\theta}) \ln \frac{p(\boldsymbol{\theta}|\mathbf{y})}{q(\boldsymbol{\theta})} d\boldsymbol{\theta}. \end{aligned} \tag{10.13}$$

where $\boldsymbol{\theta} = \{\mathbf{x},\boldsymbol{\xi},\gamma,\mathbf{A}\}$ and θ_i represents each independent factor in $\boldsymbol{\theta}$. Since $KLD_{VB} \geq 0$, it implies that $L(q)$ is a lower bound on $\ln p(\mathbf{y})$. Moreover, $\ln p(\mathbf{y})$ is independent of $q(\boldsymbol{\theta})$ and therefore maximizing $L(q)$ is equivalent to minimizing KLD_{VB}. This is called as ELBO maximization and doing this in an alternating fashion for each variable in $\boldsymbol{\theta}$ leads to

$$\begin{aligned} \ln(q_i(\theta_i)) &= < \ln p(\mathbf{y},\boldsymbol{\theta}) >_{k\neq i} + c_i, \\ p(\mathbf{y},\boldsymbol{\theta}) &= p(\mathbf{y}|\mathbf{x},\boldsymbol{\xi},\gamma)\, p(\mathbf{x}|\boldsymbol{\xi})\, p(\boldsymbol{\xi})\, p(\gamma). \end{aligned} \tag{10.14}$$

Here $<>_{k\neq i}$ represents the expectation operator over the distributions $q_k(\theta_k)$ for all $k \neq i$.

10.4 Kronecker Structured Dictionary Learning

In the following sections, we represent (10.1) using the tensor decomposition properties from [213]. Let $Y_{i_1,\ldots,i_N}$ represents the $(i_1, i_2, \ldots, i_N)^{th}$ element of the tensor and $\mathbf{y} = [y_{1,1,\ldots,1}, y_{1,1,\ldots,2} \ldots, y_{i_1,i_2,\ldots,i_N}]^T$, then it can be verified that [222, 225]

$$\mathbf{y} = (\mathbf{A}_1 \otimes \mathbf{A}_2 \ldots \otimes \mathbf{A}_N)\mathbf{x} + \mathbf{w} = (\bigotimes_{j=1}^{N} \mathbf{A}_j)\mathbf{x} + \mathbf{w}, \tag{10.15}$$

where we denote $\mathbf{A} = \bigotimes_{j=1}^{N} \mathbf{A}_j$. Since the sparsity measure (number of nonzero components) of $\mathbf{x}$ is unknown and the following VB-SBL algorithm performs automatic rank determination.

We emphasize that the presented algorithm do not exploit parametric forms, because those parametric forms are uncertain. For example, considering the massive MIMO channel estimation problem [223], the array response at the mobile station (MS) is not exploitable. Even the array response at the base station (BS) will typically require calibration to be exploitable. Doppler shifts lead to Vandermonde vectors. Delays could be more or less clear if one goes to the frequency domain in OFDM, and one only takes into account the range of subcarriers for which the Tx/Rx filters can be considered frequency-flat. Then over those subcarriers, it is also Vandermonde. We consider $\mathbf{A}_{j,i} = [1\, \mathbf{a}_{j,i}^H]^H$ and further $\mathbf{a}_{j,i}$ is unconstrained and deterministic (in all the Vandermonde cases, it is perfect, or in all cases of phasors). Assuming first entry to be 1 is

even better than $\|\mathbf{A}_{j,i}\| = 1$ because $\|\mathbf{A}_{j,i}\| = 1$ still leaves a phase ambiguity. With first entry= 1, the factors are unique, up to permutation in the sum of terms. It is to be noted that one major difference compared to the DL for Khatri-Rao structured matrix factors as looked upon in our paper [200] is that we avoid considering a discretized dictionary and instead of the sparsity for $\mathbf{x}$ comes from considering the cases of multi-paths with same delay having different AoA or AoDs.

10.4.1 SAVED-KS Sparse Bayesian Learning

In this section, we propose a Space Alternating Variational Estimation (SAVE) based alternating optimization between each elements of $\boldsymbol{\theta}$. For SAVE, not any particular structure of $\mathbf{A}$ is assumed, in contrast to AMP which performs poorly when $\mathbf{A}$ is not i.i.d. or sub-Gaussian. Based on a quadratic loss function, the Bayesian estimator of a parameter is the posterior mean; we therefore define the VB estimators of parameters $\boldsymbol{\theta}$ as the means of the variational approximation to the posterior distribution. The joint distribution can be written as

$$\begin{aligned}\ln p(\mathbf{y},\boldsymbol{\theta}) = N\ln\gamma - \gamma\left||\mathbf{y}-\mathbf{A}\mathbf{x}|\right|^2 + \sum_{i=1}^{M}\left(\ln\xi_i - \xi_i|x_i|^2\right)\\ + \sum_{i=1}^{M}((a-1)\ln\xi_i + a\ln b - b\xi_i) + (c-1)\ln\gamma + c\ln d - d\gamma + \text{constants}.\end{aligned} \tag{10.16}$$

In the following, $c_{x_i}, c'_{x_i}, c_{\xi_i}$, $c_\gamma, c_{a_{ji}}$ etc. represents normalization constants for the respective pdfs.

Update of $q_{x_i}(x_i)$: Using (10.14), $\ln q_{x_i}(x_i)$ turns out to be quadratic in x_i and thus can be represented as a Gaussian distribution as follows. Note that we split $\mathbf{A}\mathbf{x}$ as, $\mathbf{A}\mathbf{x} = \mathbf{C}_i x_i + \mathbf{C}_{\bar{i}}\mathbf{x}_{\bar{i}}$, where $\mathbf{C}_i$ represents the i^{th} column of $\mathbf{A}$, $\mathbf{C}_{\bar{i}}$ represents the matrix with i^{th} column of $\mathbf{A}$ removed, x_i is the i^{th} element of $\mathbf{x}$, and $\mathbf{x}_{\bar{i}}$ is the vector without x_i. In fact, we can represent $\mathbf{C}_i = (\bigotimes_{j=1}^{N}\mathbf{A}_{j,p_{ji}})$. To show the relation to the columns of the KS factor matrices $(p_1, p_2, ..., p_N)$ which generates $\mathbf{C}_i$, $i = 1 + \sum_{k=1}^{N}(p_k-1)J_k, J_k = \prod_{m=N,m\neq i}^{k+1} P_m, P_{N+1} = 1$. So we denote $\mathbf{A}_{j,p_{ji}}$ as the column vector from $\mathbf{A}_j$ which generates $\mathbf{C}_i$. From the property of the Kronecker products [222] that $(\mathbf{A}\otimes\mathbf{B})(\mathbf{C}\otimes\mathbf{D}) = \mathbf{AC}\otimes\mathbf{BD}$, we can verify that $\|\mathbf{C}_i\|^2 = (\bigotimes_{j=1}^{N}\mathbf{A}_{j,p_{ji}})^H(\bigotimes_{j=1}^{N}\mathbf{A}_{j,p_{ji}}) = \prod_{j=1}^{N}\left\|\mathbf{A}_{j,p_{ji}}\right\|^2$. Clearly, the mean and the variance of the resulting Gaussian distribution becomes

$$\begin{aligned}\sigma_i^2 &= \frac{1}{<\gamma>\prod_{j=1}^{N}<\left\|\mathbf{A}_{j,p_{ji}}\right\|^2> + <\xi_i>},\\ <x_i> &= \widehat{x}_i\\ &= \sigma_i^2((<\mathbf{C}_i^H>\mathbf{y} - <(\mathbf{C}_i^H\mathbf{C}_{\bar{i}})><\mathbf{x}_{\bar{i}}>)<\gamma>,\end{aligned} \tag{10.17}$$

where $\widehat{x}_i$ represents the point estimate of x_i and $\widehat{\mathbf{A}}_{j,i} = [1 < \mathbf{a}_{j,i}^H >]^H$, $< \mathbf{a}_{j,i} >$ being the mean of $\mathbf{a}_{j,i}$ which follows from the below derivation for $\mathbf{a}_{j,i}$. Also, note that in $< (\mathbf{C}_i^H\mathbf{C}_{\bar{i}}) >$, there are cross terms of the form $< \prod_{j=1}^{N}\mathbf{A}_{j,p_i}^H\mathbf{A}_{j,p_k} >, i \neq k$ which can be written as $= \prod_{j=1}^{N} < \mathbf{A}_{j,p_i}^H >< \mathbf{A}_{j,p_k} >$ because of the independence of the approximate distribution q of each columns of the factor matrices.

Update of $q_{\mathbf{a}_{j,i}}(\mathbf{a}_{j,i})$: Here we go back to the tensor representation. For simplicity, we define

$\mathbf{V}_j = <\mathbf{X}^{(j)}><(\overset{1}{\underset{k=N,k\neq j}{\otimes}}\mathbf{A}_k)^T>, \mathbf{W}_j = <\mathbf{X}^{(j)}(\overset{1}{\underset{k=N,k\neq j}{\otimes}}\mathbf{A}_k)^T(\overset{1}{\underset{k=N,k\neq j}{\otimes}}\mathbf{A}_k)^*\mathbf{X}^{(j)\,H}>$. The variational approximation for the vector $\mathbf{a}_{j,i}$ results in

$$\begin{aligned}\ln q_{\mathbf{a}_{j,i}}(\mathbf{a}_{j,i}) &= -<\gamma><\left\{\|\mathscr{Y}-[[\mathscr{X};\mathbf{A}_1,...,\mathbf{A}_N]]\|^2\right\}> \\ &\overset{(a)}{=} -<\gamma>\mathrm{tr}\left\{-\mathbf{Y}^{(j)}\mathbf{V}_j^H\mathbf{A}_j^H+\mathbf{A}_j\mathbf{V}_j\mathbf{Y}^{(j)}+\mathbf{A}_j\mathbf{W}_j\mathbf{A}_j^H\right\}+c_{\mathbf{a}_{ji}}.\end{aligned} \tag{10.18}$$

In (a), we used the fact that [213] $\|\mathbf{A}\|^2 = \mathrm{tr}\{\mathbf{A}^{(k)}(\mathbf{A}^{(k)})^H\}$ for a tensor $\mathbf{A}$ and further we denote $\mathbf{A}_N\otimes...\mathbf{A}_{j+1}\otimes\mathbf{A}_{j-1}...\otimes\mathbf{A}_1 = \overset{1}{\underset{k=N,k\neq j}{\otimes}}\mathbf{A}_k$. In (10.18), $\mathrm{tr}\{\mathbf{A}_j\mathbf{W}_j\mathbf{A}_j^H\}$ can be written as, $\mathrm{tr}\{\mathbf{A}_{j,i}\mathbf{W}_j\mathbf{A}_{j,i}^H\}$+ "terms independent of $\mathbf{a}_{j,i}$", which gets simplified as $\mathrm{tr}\{\mathbf{W}_j\}\left\|\mathbf{a}_{j,i}\right\|^2$ + "others". Finally, the mean ($<\mathbf{a}_{ji}>=\widehat{\mathbf{a}}_{ji}$) and covariance ($\Upsilon_{j,i}$) of the resulting Gaussian distribution can be written as (after expanding $\mathbf{V}_j,\mathbf{W}_j$)

$$\begin{aligned}\widehat{\mathbf{a}}_{ji} &= (\mathbf{b}_j)_{\overline{1}}, \\ \mathbf{b}_j &= (\mathbf{Y}^{(j)}<\mathbf{X}^{(j)}><(\overset{1}{\underset{k=N,k\neq j}{\otimes}}\mathbf{A}_k)^T>)_i, \\ \Upsilon_{j,i} &= \beta_{j,i}\mathbf{I}, \\ \beta_{j,i} &= \mathrm{tr}\{(\overset{1}{\underset{k=N,k\neq j}{\otimes}}<\mathbf{A}_k^T\mathbf{A}_k^*>)<\mathbf{X}^{(j)\,H}\mathbf{X}^{(j)}>\},\end{aligned} \tag{10.19}$$

where $(\cdot)_i$ represents the i^{th} column of the matrix $(\cdot)$ and $(\mathbf{b}_j)_{\overline{1}}$ represents the vector formed by all the elements except the first one of the vector $\mathbf{b}_j$. For the computation of the elements of the matrix $<\mathbf{X}^{(j)\,H}\mathbf{X}^{(j)}>$, the diagonal elements contain terms of the form $<|x_l|^2>$ the expressions for which are provided below in (10.20). The non-diagonal terms contain terms of the form $<x_lx_k>, l\neq k$ which gets simplified due to the independence of the corresponding q distributions, $<x_lx_k>=\widehat{x}_l\widehat{x}_k$. Also, we can write $<\left\|\mathbf{A}_{j,i}\right\|^2>=1+\left\|\widehat{\mathbf{a}}_{j,i}\right\|^2+\beta_{j,i}I_j$, which gets used in (10.17).

Update of $q_{\xi_i}(\xi_i), q_\gamma(\gamma)$: The variational approximation leads to the Gamma distribution for the $q_{\xi_i}(\xi_i)$ and $q_\gamma(\gamma)$, which are parameterized by its mean. The detailed derivation for this is omitted here, since it is provided in our paper [168]. The mean of the Gamma distribution for $q_{\xi_i}(\xi_i), q_\gamma(\gamma)$ is given by

$$\begin{aligned}<\xi_i> &= \frac{a+\frac{1}{2}}{\left(<|x_i|^2>+b\right)}, \\ <\gamma> &= \frac{c+\frac{N}{2}}{(<\left\|\mathbf{y}-(\overset{N}{\underset{j=1}{\otimes}}\mathbf{A}_j)\mathbf{x}\right\|^2>+d)},\end{aligned} \tag{10.20}$$

where

$$\begin{aligned}<\|\mathbf{y}-(\overset{N}{\underset{j=1}{\otimes}}\mathbf{A}_j)\mathbf{x}\|^2> &= \left\|\mathbf{y}\right\|^2-2\mathbf{y}^H(\overset{N}{\underset{j=1}{\otimes}}<\widehat{\mathbf{A}}_j>)\widehat{\mathbf{x}}+\mathrm{tr}((\overset{N}{\underset{j=1}{\otimes}}<\mathbf{A}_j^H\mathbf{A}_j>)(\widehat{\mathbf{x}}\widehat{\mathbf{x}}^H+\Sigma)), \\ \Sigma &= \mathrm{diag}(\sigma_1^2,...,\sigma_M^2), \\ \widehat{\mathbf{x}} &= [\widehat{x}_1,\widehat{x}_2,...,\widehat{x}_M]^H.\end{aligned} \tag{10.21}$$

and

$$<|x_i|^2>=|\widehat{x}_i|^2+\sigma_i^2,. \tag{10.22}$$

From (10.17), it can be seen that the estimate $\widehat{\mathbf{x}}$ converges to the L-MMSE equalizer $\widehat{\mathbf{x}} = (\mathbf{A}^H\mathbf{A} + \frac{1}{<\gamma>}\boldsymbol{\Sigma}^{-1})^{-1}\mathbf{A}^H\mathbf{y}$. This version of the SAVE where each columns of the factor matrices are updated independently is called as SAVED-KS (SAVE with KS Dictionary learning).

10.4.2 Joint VB for KS Dictionary Learning

In this section, we treat the columns of the factor matrix $\mathbf{A}_j$ jointly in the approximate posterior using VB. We also define for the convenience of the analysis, $\mathbf{A}_j = [\mathbf{1}\,\mathbf{A}_{\bar{1},j}^H]^H$, where $\mathbf{A}_{\bar{1},j}$ represents all other rows except the first and $\mathbf{1}$ represents a column vector (of size P_j) with all ones. $\ln q_{\mathbf{A}_j}(\mathbf{A}_j) = \text{tr}\{-\mathbf{Y}^{(j)}\mathbf{V}_j^H\mathbf{A}_j^H - \mathbf{A}_j\mathbf{V}_j\mathbf{Y}^{(j)\,H} + \mathbf{A}_j\mathbf{W}_j\mathbf{A}_j^H\} + c_{\mathbf{A}_j}$, Defining $\mathbf{B}_j$ as with the first row of $(\mathbf{Y}^{(j)}\mathbf{V}_j^H)$ removed. So $\text{tr}\{-\mathbf{Y}^{(j)}\mathbf{V}_j^H\mathbf{A}_j^H\} = \sum_{i=1}^{M}(\mathbf{Y}^{(j)}\mathbf{V}_j^H)_{1,i} + \text{tr}\{\mathbf{B}_j\mathbf{A}_{\bar{1},j}^H\}$, $(\mathbf{Y}^{(j)}\mathbf{V}_j^H)_{1,i}$ represents the $(1,i)^{th}$ element of the matrix. Now expanding the term $\mathbf{A}_j\mathbf{W}_j\mathbf{A}_j^H = [\mathbf{1}\,\mathbf{A}_{\bar{1},j}^H]^H\mathbf{W}_j[\mathbf{1}\,\mathbf{A}_{\bar{1},j}^H]$ which simplifies $\ln q_{\mathbf{A}_j(\mathbf{A}_j)}$ as

$$\ln q_{\mathbf{A}_j}(\mathbf{A}_j) = <\gamma>\text{tr}\{\mathbf{B}_j\mathbf{A}_{\bar{1},j}^H\} + <\gamma>\text{tr}\{\mathbf{A}_{\bar{1},j}\mathbf{B}_j^H\} - <\gamma>\text{tr}\{\mathbf{A}_{\bar{1},j}\mathbf{W}_j\mathbf{A}_{\bar{1},j}^H\}. \tag{10.23}$$

This corresponds to the functional form of a circularly-symmetric complex matrix normal distribution [226]. This can be represented for a random matrix $\mathbf{X} \in \mathbf{C}^{n\times p}$ as $p(\mathbf{X}) \propto \exp(-\text{tr}\{\boldsymbol{\Psi}^{-1}(\mathbf{X}-\mathbf{M})^H\boldsymbol{\Phi}^{-1}(\mathbf{X}-\mathbf{M})\})$, which is denoted as $\mathscr{CMN}(\mathbf{X}|\mathbf{M},\boldsymbol{\Phi},\boldsymbol{\Psi})$. Thus the variational approximation for $\mathbf{A}_{\bar{1},j}$ gets represented as $\mathscr{CMN}(\mathbf{A}_{\bar{1},j}|\mathbf{M}_j,\mathbf{I}_M,\boldsymbol{\Psi}_j)$.

$$\begin{aligned}\mathbf{M}_j &= \widehat{\mathbf{A}}_{\bar{1},j} = <\gamma>\mathbf{B}_j\boldsymbol{\Psi}_j,\\ \boldsymbol{\Psi}_j &= (<\gamma><\mathbf{X}^{(j)}(\bigotimes_{k=N,k\neq j}^{1}<\mathbf{A}_k^T\mathbf{A}_k^*>)\mathbf{X}^{(j)\,H}>)^{-1}.\end{aligned} \tag{10.24}$$

Note that $vec(\mathbf{A}_{\bar{1},j}) \sim \mathscr{N}(vec(\mathbf{M}_j), \boldsymbol{\Psi}_j \otimes \mathbf{I}_M)$, so the terms of the form $< \|\mathbf{A}_{j,i}\|^2 >$ in (10.17) becomes, $< \|\mathbf{A}_{j,i}\|^2 >= 1 + \|\mathbf{M}_{j,i}\|^2 + (\boldsymbol{\Psi}_j)_{i,i}$. $(\boldsymbol{\Psi}_j)_{i,i}$ is the i^{th} diagonal element of $\boldsymbol{\Psi}_j$ and $\mathbf{M}_{j,i}$ represents the i^{th} column of $\mathbf{M}_j$. Also, we can represent $\mathbf{A}_j^H\mathbf{A}_j = \mathbf{1}\mathbf{1}^H + \mathbf{M}_j^H\mathbf{M}_j + (I_j-1)\boldsymbol{\Psi}_j$.

For our proposed SAVED-KS, it is evident that we do not need any matrix inversions compared to [148, 166]. Update of all the variable $\mathbf{x}, \boldsymbol{\xi}, \gamma$ involves simple addition and multiplication operations. We also introduce the following notations, $\mathbf{x}_{i-} = [x_1 ... x_{i-1}]^T$, $\mathbf{x}_{i+} = [x_{i+1} ... x_M]^T$.

10.5 Identifiability of KS Dictionary Learning

The local identifiability (upto permutation ambiguity) of the KS DL is ensured if the FIM is non-singular [228]. We can write $\mathbf{A}_{j,i} = \mathbf{F}_j^{(i)}\boldsymbol{\theta}_j$, $\boldsymbol{\theta}_j = vec(\mathbf{A}_j)$ and $\mathbf{F}_j^{(i)} = [\mathbf{0}_{I_j\times I_j(i-1)}\,\mathbf{I}_{I_j}\,\mathbf{0}_{I_j\times I_j(P_j-i)}]$ and we define $\mathbf{F}_r = \bigotimes_{p_{ji},\forall j}\mathbf{F}_j^{(p_{ji})}$, $r = \sum_{j=1}^{N}(p_{ji}-1)J_j + p_{Ni}$, $J_j = \prod_{r=j+1}^{N}P_r$. We observe that we can separate the contributions of $\boldsymbol{\theta}$ and $\mathbf{x}$ in (10.15) as,

$$\mathbf{y} = \underbrace{(\sum_{r=1}^{M}x_r\mathbf{F}_r)}_{\mathbf{F}(\mathbf{x})}\underbrace{(\bigotimes_{j=1}^{N}\boldsymbol{\theta}_j)}_{\mathbf{f}(\boldsymbol{\theta})} + \mathbf{w}. \tag{10.25}$$

Writing $\ln p(\mathbf{y},\boldsymbol{\theta},\mathbf{x}) = \ln p(\mathbf{y}|\mathbf{x},\boldsymbol{\xi},\gamma,\mathbf{A}) + \ln p(\mathbf{x}|\boldsymbol{\xi})p(\boldsymbol{\xi})p(\gamma)$, it is clear that the FIM can be split as $FIM = FIM_y + FIM_{prior}$. For FIM_y, extending the derivation of the CRB for the KS dictionary

Algorithm 18: SAVED-KS SBL Algorithm

Given: $\mathbf{y}, \mathbf{A}, I_j, P_j \; \forall j$.
Initialization: a, b, c, d are taken to be very low, on the order of 10^{-10}, thus $p(\xi_i) \propto \xi_i^{-1}, p(\gamma) \propto \gamma^{-1}$ which corresponds to a non-informative Jeffrey's prior [227]. $\xi_i^0 = a/b, \forall i, \gamma^0 = c/d$ and $\sigma_i^{2,0} = \frac{1}{||\mathbf{C}_i^0||^2\gamma^0 + \xi_i^0}, \mathbf{x}^0 = \mathbf{0}$. Random initialization for the dictionary matrix $\mathbf{A}_j \sim \mathscr{CN}(\mathbf{0}, \mathbf{I})$.
At iteration $t+1$ (superscript t is used to denote the iteration stage),

1. Update $\sigma_i^{2,t+1}, \widehat{x}_i^{t+1}, \forall i$ from (10.17) using $\mathbf{x}_{i-}^{t+1}$ and $\mathbf{x}_{i+}^{t}$.
2. Update $\widehat{\mathbf{A}}_{j,i}^{t+1}, \Upsilon_{j,i} \; \forall i, j$ from (10.19) or $\widehat{\mathbf{A}}_j^t, \mathbf{\Psi}_j$ from (10.24).
3. Compute $< x_i^{2,t+1} >$ from (10.20) and update ξ_i^t, γ^{t+1}.
4. Continue steps $1-4$ till convergence of the algorithm.

matrices in [228] to the high-order tensor SBL case, we define the Jacobian matrix of $\mathbf{S} = \mathbf{F}(\mathbf{x})\mathbf{f}(\boldsymbol{\theta})$ as

$$
\begin{aligned}
&\mathbf{J}(\boldsymbol{\theta}, \mathbf{x}) = [\mathbf{J}(\boldsymbol{\theta}) \; \mathbf{J}(\mathbf{x})], \; \mathbf{J}(\boldsymbol{\theta}) = [\mathbf{J}(\boldsymbol{\theta}_1) ..... \mathbf{J}(\boldsymbol{\theta}_N)] \\
&\text{where,} \\
&\mathbf{J}(\boldsymbol{\theta}_j) = \mathbf{F}(\mathbf{x})(\boldsymbol{\theta}_1 \otimes ... \mathbf{I}_{I_j P_j} .... \otimes \boldsymbol{\theta}_N), \\
&\mathbf{J}(\mathbf{x}) = [\mathbf{F}_1(\bigotimes_{j=1}^{N} \boldsymbol{\theta}_j), ..., \mathbf{F}_M(\bigotimes_{j=1}^{N} \boldsymbol{\theta}_j))].
\end{aligned} \tag{10.26}
$$

We define $\mathbf{\Xi} = \text{diag}(\boldsymbol{\xi})$. Further, the FIM for the case of SBL can be written as

$$
FIM = \begin{bmatrix}
\mathrm{E}(\gamma)\mathbf{J}(\boldsymbol{\theta})^H\mathbf{J}(\boldsymbol{\theta}) & \mathbf{0} & \mathbf{0} & \mathbf{0} \\
\mathbf{0} & \mathrm{E}(\gamma)\mathbf{J}(\mathbf{x})^H\mathbf{J}(\mathbf{x}) + \mathrm{E}(\mathbf{\Xi}) & \mathbf{0} & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & a\,\mathrm{E}(\mathbf{\Xi}^{-2}) & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \mathbf{0} & (N+c-1)\,\mathrm{E}(\gamma^{-2})
\end{bmatrix} \tag{10.27}
$$

Here, $\gamma\mathbf{J}(\mathbf{x})^T\mathbf{J}(\boldsymbol{\theta}) = \mathbf{0}$, since $\mathbf{x}$ is zero mean. Further using the expression for the inverse of the block FIM above, for non-singularity, $\mathbf{J}(\boldsymbol{\theta})$ should be full rank. For the FIM analysis, we assume that the support (no. of non-zero elements of $\mathbf{x}$) is known, then $\mathrm{E}(\gamma)\mathbf{J}(\mathbf{x})^H\mathbf{J}(\mathbf{x}) + \mathrm{E}(\mathbf{\Xi})$ and $a\,\mathrm{E}(\mathbf{\Xi}^{-2})$ becomes invertible if $\prod_{j=1}^{N} I_j > K$. Assuming $\prod_{j=1}^{N} I_j > \sum_{j=1}^{N} (I_j - 1)P_j$ ($I_j - 1$ since the columns are scaled to make the first entry 1), i.e. no. of degrees of freedom in the dictionary $< \prod_{j=1}^{N} I_j$, then it is clear FIM is non-singular. Another remark is that here we consider only single measurement vector case and it is evident from the FIM expression that it can be non-singular even in this case under certain conditions on the dimensions of the KS factor matrices. We also observe that identifiability results for a mix of structured (Vandermonde matrices) and unstructured KS matrices for 3−way tensors are discussed in [229]. Note that algorithms which deal with KS dictionary matrices are very recent and fundamental limits of the estimation accuracy for such systems in a minimax setting can be seen in [230].

10.5.1 Identifiability for mix of parametric and non-parametric KS factors

We briefly outline the results for the case of mixture of parametric and non-parametric KS factors. We assume that the parameters $\mathbf{A}_j, j = 1, ..., P, P < N$ are Vandermonde matrices parame-

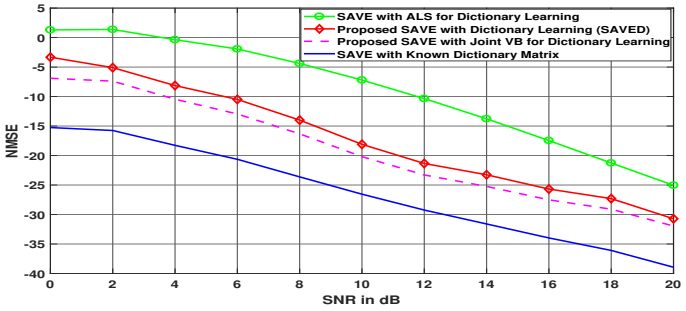


Figure 10.1: NMSE vs SNR in dB.

terized by the spatial response $\phi_{j,l}, l = 1,...,P_j$ and $\mathbf{A}_{j,l} = [1\, e^{ig_j(\phi_{j,l})} .... e^{i(I_j-1)g_j(\phi_{j,l})}]^T, i = \sqrt{-1}$, where for example $g_j(\phi_{j,l}) = \pi \sin(\phi_{j,l})$ and angles are sufficiently separated such that each of the columns $\mathbf{A}_{j,l}$ becomes linearly independent. This corresponds to the case of antenna array response for ULA or frequency response parameterized by a delay. Further vectorizing $\boldsymbol{\theta}_j = vec(\phi_{j,1},...,\phi_{j,P_j})$, so the degrees of freedom reduces to P_j instead of $I_j P_j$ for the unstructured case. So, $\forall\, j = 1,...,P$

$$\mathbf{J}(\boldsymbol{\theta}_j) = \mathbf{F}_{pa}(\mathbf{x})(\boldsymbol{\theta}_1 \otimes ... \mathbf{E}_j \mathbf{A}_j \mathbf{F}_j ... \boldsymbol{\theta}_P \otimes \boldsymbol{\theta}_{P+1} .... \otimes \boldsymbol{\theta}_N), \tag{10.28}$$

where $\mathbf{F}_{pa}(\mathbf{x})$ has the same expression as $\mathbf{F}(\mathbf{x})$ with $\mathbf{F}_j^{(i)}, \forall\, j = 1,..,P$ becomes a matrix with all ones of size $I_j \times I_j$, $\mathbf{E}_j = \text{diag}\,(0,1,...,(I_j-1))$ and $\mathbf{F}_j = i\,\text{diag}\,(g_j'(\phi_{j,1}),...,g_j'(\phi_{j,P_j}))$. Thus for parametric factors, $\mathbf{J}(\boldsymbol{\theta}_j)$ becomes a vector of size $\prod_j I_j \times P_j$. The identifiability conditions can be restated as, assuming $\prod_{j=1}^{N} I_j > \sum_{j=1}^{P} P_j + \sum_{j=P+1}^{N} (I_j - 1)P_j$, i.e. no. of degrees of freedom in the dictionary $< \prod_{j=1}^{N} I_j$, then it is clear FIM is non-singular.

10.5.2 Simulation Results

In this section, we present the simulation results to validate the performance of our SAVED-KS SBL algorithm (Algorithm 1) compared to state of the art solutions. For the simulations, we consider a $3-D$ tensor with dimensions $(4,4,4)$ and the number of non-zero elements of $\mathbf{x}$ or the rank of the tensor (no of non-zero elements of $\mathbf{x}$) is fixed to be 4. All the elements of the dictionary matrix $\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3$ and non-zero elements of $\mathbf{x}$ are generated i.i.d. complex Gaussian, $\mathscr{CN}(0,1)$ and the singular values are modified to convert the matrices such that they have a particular condition number $(= 2)$. This is done to ensure that the system identifiability is not affected by the Krushkal ill-conditioning [213]. Normalized Mean Square Error (NMSE) is defined as $NMSE = \frac{1}{M}||\hat{\mathbf{x}} - \mathbf{x}||^2$, $\hat{\mathbf{x}}$ represents the estimated value, $NMSE_{dB} = 10\log 10(NMSE)$. In Figure 10.1, we depict the normalized MSE (NMSE) performance of our proposed SAVED-KS algorithm with the classical ALS algorithm which does not utilize any statistical information about the dictionary or sparse coefficients. Our SAVED-KS algorithm has much better reconstruction error performance compared to the ALS [213] and our joint VB version performs better than the SAVED-KS version, but comes with a higher computational complexity due to the matrix inversion. It is clear from Figure 10.2 that proposed SAVE approach has a faster convergence rate than the ALS.

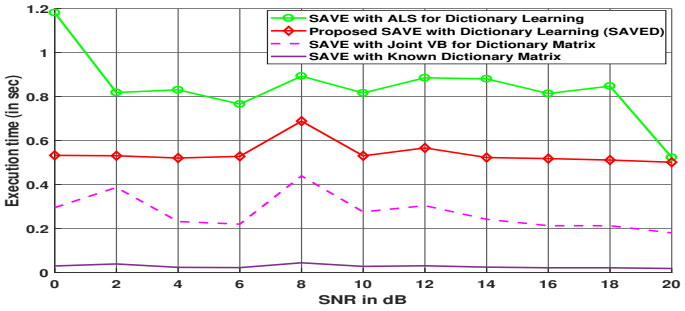


Figure 10.2: Execution time in Matlab for the various algorithms.

10.5.3 Conclusions and Perspectives

Conclusions and Perspectives 11

- We presented a fast SBL algorithm called SAVED-KS, which uses the variational inference techniques to approximate the posteriors of the data, hyperparameters and the factor matrices of the dictionary.
- We showed that the proposed algorithm has a faster convergence rate and better performance in terms of NMSE than even the state of the art ALS solutions for dictionary learning.
- Possible extensions to the current work might include: i) Convex combination of structured and unstructured KS factor matrices, for example, DoA response closeness to the vandermonde. ii) Asymptotic performance analysis and mismatched Cramer-Rao bounds [231] for the SAVED-KS algorithm.
- One of the disadvantage with our algorithm is that in the simulations it is observed that it may not converge when the condition number of the factor matrices is high. However, it is observed that the method proposed in [232] seems to avoid the sensitivity of ALS algorithms to ill-conditioned data. The proposed solution therein is based on semi-algebraic approach that algebraically reformulate the CPD into a set of a simultaneous matrix diagonalization (SMD) problems. However, one drawback of their method is that the rank of the tensor be known in advance. Hence, it would be worth looking at a combination of SBL and such SMD methods to avoid the convergence issue associated with our proposed dictionary learning in this thesis. This is left as a future work.

10.6 Joint Dictionary Learning and Dynamic Sparse State Vector Estimation

The signal model for the recovery of a time varying sparse signal under Kronecker structured (KS) [212, 225] dictionary matrix can be formulated as

$$\begin{aligned} &\text{Observation: } \mathbf{y}_t = (\mathbf{A}_1^{(t)} \otimes \mathbf{A}_2^{(t)} .... \otimes \mathbf{A}_N^{(t)})\mathbf{x}_t + \mathbf{v}_t, \\ &\text{State Update: } \mathbf{x}_t = \mathbf{F}\mathbf{x}_{t-1} + \mathbf{w}_t, \end{aligned} \tag{10.29}$$

10.6.1 Dynamic BP-MF-EP based SBL

The figure 9.2 represents the factor graph (FG) (note that static case is a special case with the state update nodes being not present), where it is divided into two disjoint subsets $\mathscr{A}_{BP} = f_{\delta_{n,t}} \forall n,l,t$ and $\mathscr{A}_{MF}$ represents rest of the factor or variable nodes. To combine BP and MF, we introduce the new variables $h_{n,t} = \mathbf{A}^{(t)}_{n,:}\mathbf{x}_t$, $s_{l,t} = f_l x_{l,t-1}$ and the hard constraint factor nodes, $f_{\delta_{n,t}} = \delta(h_{n,t} - \mathbf{A}^{(t)}_{n,:}\mathbf{x}_t), \forall n \in [1:N], t$, and $f_{\Delta_{l,t}} = \delta(s_{l,t} - f_l x_{l,t-1}), \forall l \in [1:M], t$. We can compute $m_{f_{\delta_{n,t}} \to x_{l,t}}(x_{l,t}) = \int f_{\delta_{n,t}} n_{h_{n,t} \to f_{\delta_{n,t}}}(h_{n,t}) \prod_{l' \neq l} n_{x_{l',t} \to f_{\delta_{n,t}}}(x_{l',t}) \prod_{l' \neq l} dx_{l',t}$. For notational brevity, we denote subscript (l,n) or (n,l) to represent the messages passed from l to n or viceversa. All the messages (beliefs or continuous pdfs) passed between them can be shown to be Gaussian [164] and thus it suffices to represent them by the mean and variance of the beliefs. The joint distribution of all the observations and parameters can be written as, $p(\mathbf{y}_t, \boldsymbol{\theta}_t | \mathbf{y}_{1:t-1}) = p(\mathbf{y}_t | \boldsymbol{\theta}_t) p(\boldsymbol{\theta}_t | \mathbf{y}_{1:t-1})$, where $p(\boldsymbol{\theta}_t | \mathbf{y}_{1:t-1})$ denotes the predictive distribution. Similar as in KF, first we compute the posterior distribution of $\theta_{i,t}$ given the observations till $(t-1)$, which is called as the prediction stage.

10.6.1.1 Diagonal AR(1) (DAR(1)) Prediction Stage

Since there is no coupling between the scalars in the state update (10.29), it is enough to update the prediction stage using MF. However, the interation between $x_{l,t}$ and f_l requires Gaussian projection, using expectation propagation (EP). For more detailed derivation, we refer to our previous work [221] due to space limitations.

10.6.1.2 Measurement Update (Filtering) Stage

For the measurement update stage, the posterior for $\mathbf{x}_t$ is inferred using BP. Note that we represent the mean of the messages by $\widehat{x}^{(t)}_{n,l}, \nu^{(t)}_{n,l}$. The mean and variance of the beliefs computed at $x_{l,t}$ are denoted by $\widehat{x}_{l,t|t}, \sigma^2_{l,t|t}$. In the measurement stage, the prior for $x_{l,t}$ gets replaced by the belief from the prediction stage. We refer to our previous work [198] for detailed derivations and expressions for the messages. We define $d_{l,t} = (\sum_{n=1}^{N} \nu^{(t)-1}_{n,l})^{-1}$, $r_{l,t} = d_{l,t}(\sum_{n=1}^{N} \frac{\widehat{x}^{(t)}_{n,l}}{\nu^{(t)}_{n,l}} + \frac{\widehat{x}_{l,t|t-1}}{\sigma^2_{l,t|t-1}})$. Given the messages, $m_{f_{\delta_{n,t}} \to x_{l,t}}(x_{l,t})$, the belief $q(x_{l,t})$ can be obtained as $(f_{\lambda_i}(\lambda_i) = p(\lambda_k | a,b))$, $q(x_{l,t}) \propto f_{\lambda_i}(\lambda_i) \prod_{n=1}^{N} m_{f_{\delta_{n,t}} \to x_{l,t}} \propto \mathcal{N}(\widehat{x}_{l,t|t}, \sigma^2_{l,t|t})$, where

$$\begin{aligned} \sigma^{-2}_{l,t|t} &= \lambda_{l,t} + d^{-1}_{l,t},\ \widehat{x}_{l,t|t} \\ &= \frac{r_{l,t}}{1 + d_{l,t}\sigma^{-2}_{l,t|t}}. \end{aligned} \tag{10.30}$$

One remark here is that compared to our previous work using VB [168], combining BP and MF gives a more accurate approximation of the error variance as shown in (10.30), where $\sigma^2_{l,t|t}$ incorporates the effect of all $\sigma^2_{l',t|t}, l' \neq l$.

10.6.1.3 Lag-1 Smoothing Stage

We show in [198, Lemma 1] that KF is not enough to adapt the hyperparameters, instead we need at least a lag 1 smoothing (i.e. the computation of $\widehat{x}_{k,t-1|t}, \sigma^2_{k,t-1|t}$ through BP). For the smoothing stage, we use BP with Gaussian Markov Random Fields (GMRF) based factorization. GMRF refers to the representation of BP [183], when the underlying Gaussian distribution is expressed in terms of pairwise connections between scalar variables $x_{i,t}$. We skip the detailed derivation

and instead refer to our paper [198]. Applying the MF rule from (9.8), the resulting Gaussian distribution has mean, $\sigma^{-2}_{f_i|t}$ and variance, $\widehat{f}_{i|t}$, the detailed derivations for which are in [198, Section 3.2.3]. The entire algorithm (a combination of BP, MF and EP, we call it as Combined BP-MF-EP DAR-SBL) is described in Algorithm 15. Also we remark that for the estimation of λ_l, γ, we follow the same approach as in our paper [221] and we refer to it for more details. One remark here is that another version called as Combined Vector BP-MF-EP DAR-SBL follows immediately from the derivations for Algorithm 15, where all the components of $\mathbf{x}_t$ are considered jointly in the FG. Even though the performance will be higher (as observed in the simulations) for the vector case, it comes at the cost of a higher complexity due to the matrix inversion involved. Note that in Algorithm 19, we introduce temporal averaging for certain quantities (represented by $<>_{|t}$) in hyperparameter estimates and β being the temporal weighting coefficient which is less than one, see [221] for more details. For the KS DL, the algorithm remains same as in our previous work [125], which is denoted as space alternating variational estimation with Kronecker structured DL (SAVED-KS DL).

Algorithm 19: Combined BP-MF-EP DAR-SBL with KS DL

Initialization: $\widehat{f}_{l|0}, \widehat{\lambda}_{l|0} = \frac{a}{b}, \widehat{\gamma}_0 = \frac{c}{d}, \widehat{x}_{l,0|0} = 0, \sigma^2_{l,0|0} = 0, \forall l$. Define $\mathbf{\Sigma}_{t-1|t-1} = \text{diag}(\sigma^2_{l,t|t-1})$.
for $t = 1 : T$ do
Prediction Stage: 1. From [221], $\widehat{x}_{l,t|t-1} = \widehat{f}_{l|t-1}\widehat{x}_{l,t-1|t-1}, \sigma^2_{l,t|t-1} =$
$|\widehat{f}_{l|t-1}|^2\sigma^2_{l,t-1|t-1} + \sigma^2_{f_l|t-1}(|\widehat{x}_{l,t-1|t-1}|^2 + \sigma^2_{l,t-1|t-1}) + \widehat{\lambda}^{-1}_{l|t-1}$.
Filtering Stage:

1. Compute $\widehat{x}^{(t)}_{n,l}, v^{(t)}_{n,l}$ from [198, eq. (5)] and update $\widehat{x}_{l,t|t}, \sigma^{-2}_{l,t|t}$ from (10.30).

2. Compute $v^{(t)}_{l,n}, \widehat{x}^{(t)}_{l,n}$ from [198, eq. (7)]. 3. Continue steps 1) to 2) until convergence.

Smoothing Stage:
Initialization: $\mathbf{\Sigma}^{(0)}_{t-1|t} = \mathbf{\Sigma}_{t-1|t-1}, \widehat{\mathbf{x}}^{(0)}_{t-1|t} = \widehat{\mathbf{x}}_{t-1|t-1}$. Define $\mathbf{B}^{(t)} =< \mathbf{F}^T\mathbf{A}^{(t)\,T}\widetilde{\mathbf{R}}^{-1}_t\mathbf{A}^{(t)}\mathbf{F} > +\mathbf{\Sigma}_{t-1|t-1}, \mathbf{h}_t =< \mathbf{F}^T\mathbf{A}^{(t)\,T} > \widetilde{\mathbf{R}}^{-1}_t\mathbf{y}_t$.

1. $P_{i,j} = \frac{-B^{(t)2}_{i,j}}{B^{(t)}_{i,i} + \sum\limits_{k\in\mathcal{N}(i)\setminus j} P_{k,i}}, \mu_{i,j} = (h_{i,t} + \sum\limits_{k\in\mathcal{N}(i)\setminus j} P_{k,i}\mu_{k,i}), \forall i, j.$

2. $\sigma^{-2}_{i,t-1|t} = B^{(t)}_{i,i} + \sum\limits_{k\in\mathcal{N}(i)} P_{k,i}$, $\widehat{x}_{i,t-1|t} = \sigma^2_{i,t-1|t}(h_{i,t} + \sum\limits_{k\in\mathcal{N}(i)} P_{k,i}\mu_{k,i})$

Estimation of hyperparameters (Define: $x'_{k,t} = x_{k,t} - f_k x_{k,t-1}, \zeta_t = \beta\zeta_{t-1} + (1-\beta) < \left\|\mathbf{y}_t - \mathbf{A}^{(t)}\mathbf{x}_t\right\|^2 >$**),** $b'_t = (< \left|x'_{k,t}\right|^2 >_{|t} + b)$.

1. Compute $\widehat{f}_{l|t}, \sigma^2_{f_{l|t}}$ from [198, eq. (11)], $\widehat{\gamma}_t = \frac{c+N}{(\zeta_t+d)}$ and $\lambda_{l|t} = \frac{(a+1)}{b'_t}$.

SAVED-KS DL: $\widehat{\mathbf{a}}_{ji} = (\mathbf{b}_j)_{\overline{i}}$, $\mathbf{b}_j = (\mathbf{Y}^{(j)} < \mathbf{X}^{(j)} >< (\overset{1}{\underset{k=N,k\neq j}{\otimes}} \mathbf{A}_k)^T >)_i$,

$\Upsilon_{j,i} = \beta_{j,i}\mathbf{I}$, $\beta_{j,i} = \text{tr}\{(\overset{1}{\underset{k=N,k\neq j}{\otimes}} < \mathbf{A}^T_k\mathbf{A}^*_k >) < \mathbf{X}^{(j)\,T}\mathbf{X}^{(j)} >\}$.

10.6.2 Suboptimality of SAVED-KS DL and Joint VB

First, we define the unfolding operation on an N^{th} order tensor $\mathbf{Y}_t = [[\mathbf{A}_1^{(t)},...,\mathbf{A}_N^{(t)};\mathbf{x}]]$ as [213] ($\mathbf{Y}_t^{(n)}$ is of size $I_n \times \prod_{i=1,i\neq n}^{N} I_i$)

$$\mathbf{Y}_t^{(n)} = \mathbf{A}_n^{(t)}\mathbf{X}_t^{(n)}(\mathbf{A}_N^{(t)} \otimes \mathbf{A}_{N-1}^{(t)}...\mathbf{A}_{n+1}^{(t)} \otimes \mathbf{A}_{n-1}^{t}... \otimes \mathbf{A}_1^{(t)})^T. \tag{10.31}$$

From the expression for the error covariance in the estimation of the factor $\mathbf{a}_{ji}$ (tr{($\bigotimes_{k=N,k\neq j}^{1}$ $< \mathbf{A}_k^T\mathbf{A}_k^* >$) $< \mathbf{X}^{(j)\,T}\mathbf{X}^{(j)} >$}$\mathbf{I}$), it is clear that it does not take into account the estimation error in the other columns of $\mathbf{A}_j$. The columns of $\mathbf{A}_j$ can be correlated, for example if we consider two paths (say i, j) with same DoA but with different delays, the delay responses $\mathbf{v}_f(\tau_i(t))$ and $\mathbf{v}_f(\tau_i(t))$ may be correlated. However, since it is not clear how to model this dependency, we indeed keep it as a future work. This suboptimality in the error covariance estimate using SAVED-KS resulting from the correlation between the columns, can be avoided by using a joint VB [125]. The joint VB estimates (mean and covariance) can be obtained as

$$\begin{aligned} \mathbf{M}_j^T &= \widehat{\mathbf{A}}_{\overline{1},j}^T = < \gamma > \mathbf{\Psi}_j^{-1}\mathbf{B}_j^T, \\ \mathbf{\Psi}_j &= (< \gamma >< \mathbf{X}^{(j)}(\bigotimes_{k=N,k\neq j}^{1} < \mathbf{A}_k^T\mathbf{A}_k^* >)\mathbf{X}^{(j)\,T} >), \end{aligned} \tag{10.32}$$

where $\mathbf{V}_j = < \mathbf{X}^{(j)} >< (\bigotimes_{k=N,k\neq j}^{1} \mathbf{A}_k)^T >$ and $\mathbf{B}_j$ is defined as with the first row of $(\mathbf{Y}^{(j)}\mathbf{V}_j^T)$ removed. However, the joint VB involves a matrix inversion and is not recommended for large system dimensions. Nevertheless, it is possible to estimate each columns of $\mathbf{A}_j$ by BP, since each column estimate can be expressed as the solution of a linear system of equation from (10.32), $\widehat{\mathbf{a}}_{j,i}^T = \mathbf{\Psi}_j^{-1}\mathbf{b}_{j,i}$. $\mathbf{b}_{j,i}$ represents the i^{th} column of $\mathbf{B}_j^T$. The message passing expressions under BP (using a GMRF based FG) for the factor matrices can be written as

$$\begin{aligned} \zeta_{m,n} &= -(\mathbf{\Psi}_j)_{m,n}^2/(\zeta_{m,m} + \sum_{k\in\mathcal{N}(m)\setminus j} \zeta_{k,m}), \\ \kappa_{m,n} &= \frac{(\zeta_{m,m}\kappa_{m,m}) + \sum_{k\in\mathcal{N}(m)\setminus j} \zeta_{k,m}\kappa_{k,m}}{(\mathbf{\Psi}_j)_{m,n}}, \end{aligned} \tag{10.33}$$

where we initialize $\zeta_{m,m} = (\mathbf{\Psi}_j)_{m,m}$, $\kappa_{m,m} = \frac{(\mathbf{b}_{j,i})_m}{(\mathbf{\Psi}_j)_{m,m},\zeta_{m,n}} = 0, \kappa_{m,n} = 0$. Finally the mean ($\kappa_m$) and variance ($\zeta_m$) of the posterior belief can be computed as

$$\begin{aligned} \kappa_m &= \frac{\zeta_{m,m}\kappa_{m,m} + \sum_{k\in\mathcal{N}(m)} \zeta_{k,m}\kappa_{k,m}}{\zeta_{m,m} + \sum_{k\in\mathcal{N}(m)} \zeta_{k,m}}, \\ \zeta_m &= \zeta_{m,m} + \sum_{k\in\mathcal{N}(m)} \zeta_{k,m}. \end{aligned} \tag{10.34}$$

We remark that, the above BP based low complexity scheme for KS DL represents a major innovation compared to our previous work [125], apart from the extension to the dynamic SBL case.

10.7 Optimal Partitioning of the Measurement Stage and KS DL

In [125], we derived the Fisher Information Matrix (FIM) for the KS DL where the sparse vector is static. Here, we reuse the FIM expressions to derive the optimal partitioning of the variables in the measurement stage. We refer to our paper [233, Lemma 1], where the main message was that if the parameter partitioning in VB is such that the different parameter blocks are decoupled at the level of FIM, then VB is not suboptimal in terms of (mismatched) Cramer-Rao Bound (mCRB). More detailed overview on mCRB can be found in [234]. If a finer partitioning granularity is used (such as up to scalar level as in MF), then VB becomes quite suboptimal, which can be alleviated by using BP instead.

Lemma 7. *For the measurement stage, an optimal partitioning is to apply BP for the sparse vector* $\mathbf{x}_t$ *and VB (SAVED-KS) for the columns of the factor matrices* $\mathbf{A}_{j,i}^{(t)}$ *assuming the vectors* $\mathbf{A}_{j,i}^{(t)}$ *are independent and have zero mean. However, if the columns of* $\mathbf{A}_{j}^{(t)}$ *are correlated, then a joint VB, with the posteriors of the factor matrices assumed independent, should be done for an optimal performance.*

Proof: Let us define

$$\begin{aligned} \mathbf{F}_j^{(i)} &= [\mathbf{0}_{I_j \times I_j(i-1)}\, \mathbf{I}_{I_j}\, \mathbf{0}_{I_j \times I_j(P_j-i)}], \\ \Phi_{j,t} &= vec(\mathbf{A}_j^{(t)}). \end{aligned} \tag{10.35}$$

We observe that we can separate the contributions of $\mathbf{A}^{(t)}$ and $\mathbf{x}_t$ in (10.29) as

$$\mathbf{y}_t = \underbrace{(\sum_{r=1}^{M} x_{r,t}\mathbf{F}_r)}_{\mathbf{F}(\mathbf{x}_t)} \underbrace{(\bigotimes_{j=1}^{N} \Phi_{j,t})}_{\mathbf{f}(\Phi_t)} + \mathbf{w}_t. \tag{10.36}$$

We define

$$\begin{aligned} \mathbf{F}_r &= \bigotimes_{p_{ji},\forall j} \mathbf{F}_j^{(p_{ji})}, \\ r &= \sum_{j=1}^{N} (p_{ji}-1)J_j + p_{Ni}, \\ J_j &= \prod_{r=j+1}^{N} P_r. \end{aligned} \tag{10.37}$$

Further, we can write the FIM as

$$\begin{aligned} \mathbf{J}(\Phi_t, \mathbf{x}_t) &= [\mathbf{J}(\Phi_t)\, \mathbf{J}(\mathbf{x}_t)], \\ \mathbf{J}(\Phi_t) &= [\mathbf{J}(\Phi_{1,t}) ..... \mathbf{J}(\Phi_{N,t})] \\ \text{where,} \mathbf{J}(\Phi_{j,t}) &= \mathbf{F}(\mathbf{x}_t)(\Phi_{1,t} \otimes ... \mathbf{I}_{I_j P_j} .... \otimes \Phi_{N,t}), \\ \mathbf{J}(\mathbf{x}_t) &= [\mathbf{F}_1(\bigotimes_{j=1}^{N} \Phi_{j,t}), ..., \mathbf{F}_M(\bigotimes_{j=1}^{N} \Phi_{j,t}))]. \end{aligned} \tag{10.38}$$

Further, the FIM for the case of SBL can be derived as [125]

$$FIM = \begin{bmatrix} \mathrm{E}(\gamma)\mathbf{J}(\Phi_t)^T\mathbf{J}(\Phi_t) & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathrm{E}(\gamma)\mathbf{J}(\mathbf{x}_t)^T\mathbf{J}(\mathbf{x}_t) + \mathrm{E}(\Xi^{-1}) & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & a\,\mathrm{E}(\Xi^{-2}) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & (N+c-1)\,\mathrm{E}(\gamma^{-2}) \end{bmatrix} \tag{10.39}$$

Here, $\gamma \mathbf{J}(\mathbf{x}_t)^T \mathbf{J}(\Phi_t) = \mathbf{0}$, since $\mathbf{x}_t$ is zero mean. If the all columns of $\mathbf{A}_j^{(t)}$ are independent and zero mean, then $\mathrm{E}(\mathbf{J}(\Phi_t)^T \mathbf{J}(\Phi_t))$ becomes a diagonal matrix with no coupling between the free variables of any two different columns of the factor matrices. However, if any factor matrix is $\mathbf{A}_j^{(t)}$ is correlated, it is suboptimal to factorize the columns of $\mathbf{A}_j^{(t)}$ independently in the approximate posterior. Hence, in this case, a joint VB method (which has higher complexity) would be optimal to estimate the posterior distributions and this indeed justify the superior performance of joint VB approach described in Section 10.6.2.

10.8 Simulation Results

For the observation model, the parameters chosen are $N = 256, M = 200$. For the simulations, we consider a $3-D$ tensor with dimensions $(4,8,8)$ and the number of non-zero elements of $\mathbf{x}_t$ or the rank of the tensor (no of non-zero elements of $\mathbf{x}_t$) is fixed to be $K = 16$. All signals are considered to be real in the simulation. All the elements of the factor matrix $\mathbf{A}_j^{(t)}$ (time varying) are generated i.i.d. from a Gaussian distribution with mean 0 and variance 1. The rows of $\mathbf{A}^{(t)}$ are scaled by $\sqrt{16}$ so that the signal part of any scalar observation has unit variance. Taking the SNR to be 20dB, the variance of each element of $\mathbf{v}_t$ (Gaussian with mean 0) is computed as 0.01.

Consider the state update, $\mathbf{x}_t = \mathbf{F}\mathbf{x}_{t-1} + \mathbf{w}_t$. To generate $\mathbf{x}_0$, the first 16 elements are chosen as Gaussian (mean 0 and variance 1) and then the remaining elements of the vector $\mathbf{x}_0$ are put to zero. Then the elements of $\mathbf{x}_0$ are randomly permuted to distribute the 30 non-zero elements across the whole vector. The diagonal elements of $\mathbf{F}$ are chosen uniformly in $[0.9, 1)$. Then the covariance of $\mathbf{w}_t$ can be computed as $\Xi^{-1}(\mathbf{I} - \mathbf{F}\mathbf{F}^T)$. Note that Ξ contains the variances of the elements of $\mathbf{x}_t$ (including $t = 0$), where for the non-zero elements of $\mathbf{x}_0$ the variance is 1. Following observations can be made from the simulations. In Figure 10.3, which is for static SBL case with DL, there is substantial improvement in NMSE compared to our previous work [125]. Our proposed low complexity algorithm using BP has similar performance as that of joint VB which has higher complexity.

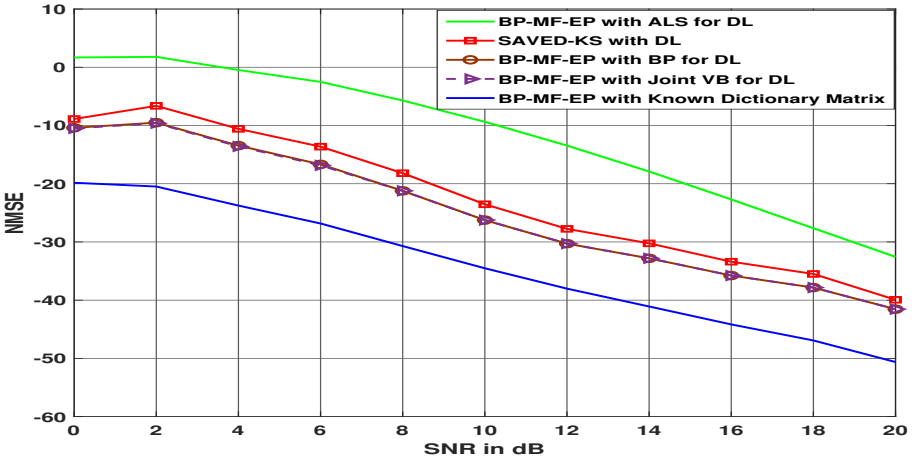


Figure 10.3: Static SBL: NMSE as a function of N.

10.9 Conclusions and Perspectives

Conclusions and Perspectives 12

- We have presented here a low complexity algorithm for KS DL using a combination of BP, VB and EP.
- The motivation behind the proposed algorithm is to circumvent the suboptimality associated with incorrect posterior covariance computation for the columns of the factor matrices in our initial work based on mean field variational Bayes.
- However, we are still unclear whether the proposed algorithm is robust enough to perform comparably with the variations in the model of KS $\mathbf{A}^{(t)}$, which is left as a future work.
- We remark here that even though we do not consider any parametric forms for the kronecker factor matrices, these can have further parsimonious parameterization such as Vandermonde, which may improve the identifiability.
- Moreover, it is an interesting observation that the couplings between different path components can be efficiently handled by BP compared to mean field VB.

Chapter 11

SPARSE BAYESIAN LEARNING FOR A BILINEAR CALIBRATION MODEL AND MISMATCHED CRB

11.1 Introduction

VB estimation allows for approximate Bayesian inference. It determines the closest approximation in the factored form of the posterior distribution by minimizing the Kullback-Leibler distance to the posterior distribution even if this last one is difficult to determine. Despite this well motivated derivation, the performance of VB techniques is not very clear, especially compared to more classical performance bounds. In this chapter, we explore recently introduced mismatched Cramer-Rao bounds (mCRB) for Bayesian estimation in the context of VB estimation. We focus on the case of bilinear signal models. One particular application of these models arises in the context of internal relative reciprocity calibration of Massive antenna arrays, in which the received signals are linear in terms of an intra array channel and the relative calibration factors. A VB approach allows for particularly improved estimation performance that goes beyond the classical CRB, which is now confirmed by the mCRB.

Massive MIMO (Multiple Input Multiple Output) requires CSIT (Channel state information at Tx) acquired using channel reciprocity for a TDD (Time Division Duplexing) system. However, Radio Frequency (RF) components are not reciprocal and we need to calibrate to compensate for this. This calibration is typically achieved by a simple complex scalar multiplication at each transmit antenna. Initial approaches to calibration relied on explicit channel feedback from a user equipment (UE) during the calibration phase to estimate the calibration parameters. This is typically referred to as UE aided calibration. However, what is popular today [235] is to perform the calibration across the antennas of the base station (BS) only and is referred to as internal calibration. In [236], the authors propose a generalized approach towards reciprocity calibration of which the existing estimation techniques are special cases.

Both the classical deterministic estimation theory and Bayesian framework are based on the assumption that the assumed data model and the true data model (pdf) are the same. However, in practice, either we may only have imperfect knowledge of the true data model or due to computational complexities associated with the computation of the true posterior distributions, we prefer approximate Bayesian inference (VB). In such a misspecified estimation framework, it is important to quantify the performance of the estimator using a mismatched Cramer-Rao bounds (mCRB) [231].

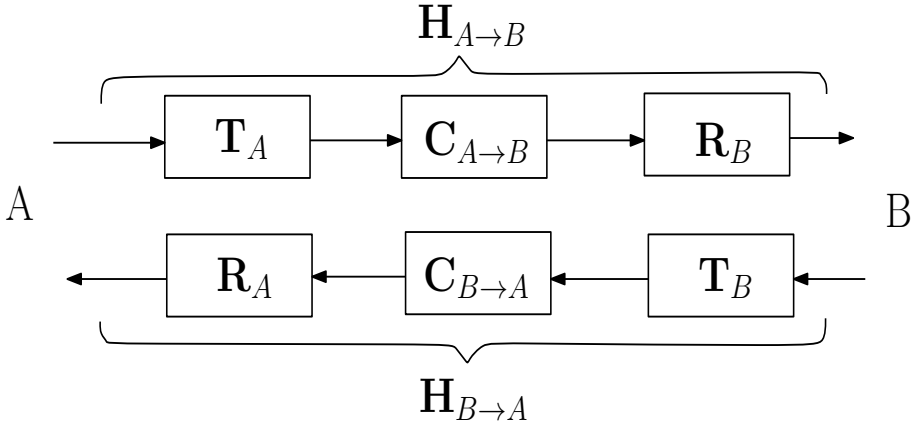


Figure 11.1: Reciprocity Model

11.1.1 Summary of this Chapter:

- We first review the constrained CRB for the case of a bilinear system model (linear in terms of the relative calibration factors and reciprocal channel coefficients).
- We propose a VB (and other variants like AMAP, EC-VB) based estimation algorithm for the joint estimation of the calibration parameters, reciprocal channel coefficients, and hyper-parameters (precisions of the bilinear factors).
- Simulations demonstrate that the mean square error (MSE) of the VB can be lower than that of the deterministic CRB. Motivated by this result, we derive simple and elegant expressions for the mCRB using Laplace approximation for the relative calibration factors.

11.2 Reciprocity Calibration System Model

Consider a system as in Fig. 11.1, where A represents a BS and B represents a UE, each containing M_A and M_B antennas, respectively. The channel as observed in the digital domain, $\mathbf{H}_{A\to B}$ and $\mathbf{H}_{B\to A}$ can be represented by,

$$\begin{aligned}\mathbf{H}_{A\to B} &= \mathbf{R}_B\mathbf{C}_{A\to B}\mathbf{T}_A,\\ \mathbf{H}_{B\to A} &= \mathbf{R}_A\mathbf{C}_{B\to A}\mathbf{T}_B,\end{aligned} \tag{11.1}$$

where (diagonal) matrices $\mathbf{T}_A$, $\mathbf{R}_A$, $\mathbf{T}_B$, $\mathbf{R}_B$ model the response of the transmit and receive RF front-ends, while $\mathbf{C}_{A\to B}$ and $\mathbf{C}_{B\to A}$ model the propagation channels, respectively from A to B and from B to A. Let us consider an antenna array of M elements partitioned into G groups denoted by $A_1, A_2, \ldots, A_G$. Group A_i contains M_i antennas such that $\sum_{i=1}^{G} M_i = M$. Each group A_i transmits a sequence of L_i pilot symbols, defined by matrix $\mathbf{P}_i \in \mathbb{C}^{M_i\times L_i}$ where the rows correspond to antennas and the columns to successive channel uses. After all G groups have transmitted, the received signal for each resource block of bidirectional transmission between antenna groups i and j is given by

$$\begin{cases}\mathbf{Y}_{i\to j} = \mathbf{R}_j\mathbf{C}_{i\to j}H_i\mathbf{P}_i + \mathbf{N}_{i\to j},\\ \mathbf{Y}_{j\to i} = \mathbf{R}_i\mathbf{C}_{j\to i}H_j\mathbf{P}_j + \mathbf{N}_{j\to i}\,.\end{cases} \tag{11.2}$$

We define $\mathbf{F}_i = \mathbf{R}_i^{-T}H_i$ and $\mathbf{F}_j = \mathbf{R}_j^{-T}H_j$ to be the calibration matrices for groups i and j. Also, $\mathbf{f}_i = vec(\mathbf{F}_i)$ represents the vectorized version. This needs to be augmented with a constraint

$\mathscr{C}(\widehat{f},\mathbf{f})=0$. Typical choices for the constraint are

$$\begin{aligned}&\text{1) Norm plus phase constraint (NPC):}\\&\quad\text{norm: } \mathrm{Re}\{\mathscr{C}(\widehat{f},\mathbf{f})\}=||\hat{\mathbf{f}}||^2-c,\ c=||\mathbf{f}||^2,\\&\quad\text{phase: } \mathrm{Im}\{\mathscr{C}(\widehat{f},\mathbf{f})\}=\mathrm{Im}\{\widehat{f}^H\mathbf{f}\}=0.\end{aligned}$$

$$\text{2) Linear constraint: } \mathscr{C}(\widehat{f},\mathbf{f})=\hat{\mathbf{f}}^H\mathbf{g}-c=0\,. \tag{11.3}$$

If we choose the vector $\mathbf{g}=\mathbf{f}$ and $c=||\mathbf{f}||^2$, then the Im{.} part of (11.3) corresponds to (11.3). The most popular linear constraint is the First Coefficient Constraint (FCC), which is (11.3) with $\mathbf{g}=\mathbf{e}_1$, $c=1$. From (11.2), we have

$$\mathbf{Y}_{i\to j}=\underbrace{\mathbf{R}_j\mathbf{C}_{i\to j}\mathbf{R}_i^T}_{\mathscr{H}_{i\to j}}\mathbf{F}_i\mathbf{P}_i+\mathbf{N}_{i\to j}. \tag{11.4}$$

We define $\mathscr{H}_{i\to j}=\mathbf{R}_j\mathbf{C}_{i\to j}\mathbf{R}_i^T$ to be an auxiliary internal channel (not corresponding to any physically measurable quantity) that appears as a nuisance parameter in the estimation of the calibration parameters. Note that the auxiliary channel $\mathscr{H}_{i\to j}$ inherits the reciprocity from the channel $\mathbf{C}_{i\to j}$: $\mathscr{H}_{i\to j}=\mathscr{H}_{j\to i}^T$. Upon applying the vectorization operator for each bidirectional transmission between groups i and j, we have

$$\mathrm{vec}(\mathbf{Y}_{i\to j})=(\mathbf{P}_i^T*\mathscr{H}_{i\to j})\,\mathbf{f}_i+\mathrm{vec}(\mathbf{N}_{i\to j}). \tag{11.5}$$

In the reverse direction, using $\mathscr{H}_{i\to j}=\mathscr{H}_{j\to i}^T$, we have

$$\mathrm{vec}(\mathbf{Y}_{j\to i}^T)=(\mathscr{H}_{i\to j}^T*\mathbf{P}_j^T)\mathbf{f}_j+\mathrm{vec}(\mathbf{N}_{j\to i})^T. \tag{11.6}$$

Alternatively, (11.5) and (11.6) may also be written as

$$\begin{aligned}\mathrm{vec}(\mathbf{Y}_{i\to j})&=\left[(\mathbf{F}_i\mathbf{P}_i)^T\otimes\mathbf{I}\right]\mathrm{vec}(\mathscr{H}_{i\to j})+\mathrm{vec}(\mathbf{N}_{i\to j})\\\mathrm{vec}(\mathbf{Y}_{j\to i}^T)&=\left[\mathbf{I}\otimes(\mathbf{P}_j^T\mathbf{F}_j)\right]\mathrm{vec}(\mathscr{H}_{i\to j})+\mathrm{vec}(\mathbf{N}_{j\to i}).\end{aligned} \tag{11.7}$$

Stacking these observations into a vector
$\mathbf{y}=\left[\mathrm{vec}(\mathbf{Y}_{1\to 2})^T\,\mathrm{vec}(\mathbf{Y}_{2\to 1}^T)^T\,\mathrm{vec}(\mathbf{Y}_{1\to 3})^T\ldots\right]^T$, the above two alternative formulations can be summarized into

$$\begin{aligned}\mathbf{y}&=\mathscr{H}(\mathbf{h},\mathbf{P})\mathbf{f}+\mathbf{n}\\&=\mathscr{F}(\mathbf{f},\mathbf{P})\mathbf{h}+\mathbf{n},\end{aligned} \tag{11.8}$$

where $\mathbf{h}=\left[\mathrm{vec}(\mathscr{H}_{1\to 2})^T\,\mathrm{vec}(\mathscr{H}_{1\to 3})^T\,\mathrm{vec}(\mathscr{H}_{2\to 3})^T\ldots\right]^T$, and $\mathbf{n}$ is the corresponding noise vector. The expressions for the composite matrices $\mathscr{H}$ and $\mathscr{F}$ are the same as given in [237, equation (18)]. The scenario is now identical to that encountered in some blind channel estimation scenarios and hence we can take advantage of some existing tools [238], [239], which we exploit next.

11.2.1 Cramér-Rao bound

Treating $\mathbf{h}$ and $\mathbf{f}$ as deterministic unknown parameters, and assuming that the receiver noise $\mathbf{n}$ is distributed as $\mathcal{CN}(0,\sigma^2\mathbf{I})$, the Fisher Information Matrix (FIM) $\mathbf{J}$ for jointly estimating $\mathbf{f}$ and $\mathbf{h}$ can immediately be obtained from (11.8) as

$$\mathbf{J} = \frac{1}{\sigma^2}\begin{bmatrix}\mathscr{H}^H \\ \mathscr{F}^H\end{bmatrix}\begin{bmatrix}\mathscr{H} & \mathscr{F}\end{bmatrix}. \tag{11.9}$$

The computation of the CRB requires $\mathbf{J}$ to be non-singular. However, for the problem at hand, $\mathbf{J}$ is inherently singular. In fact, the calibration factors (and the auxiliary channel) can only be estimated up to a complex scale factor since the received data (11.8) involves the product of the channel and the calibration factors, $\mathscr{H}\mathbf{f} = \mathscr{F}\mathbf{h}$. As a result the FIM has the following null space [240], [241]

$$\mathbf{J}\begin{bmatrix}\mathbf{f}^T & -\mathbf{h}^T\end{bmatrix}^T = \frac{1}{\sigma^2}\begin{bmatrix}\mathscr{H} & \mathscr{F}\end{bmatrix}^H(\mathscr{H}\mathbf{f}-\mathscr{F}\mathbf{h}) = \mathbf{0}. \tag{11.10}$$

To determine the CRB when the FIM is singular, constraints have to be added to regularize the estimation problem. As the calibration parameters are complex, one complex constraint corresponds to two real constraints. Another issue is that we are mainly interested in the CRB for $\mathbf{f}$, the parameters of interest, in the presence of the nuisance parameters $\mathbf{h}$. Hence we are only interested in the $(1,1)$ block of the inverse of the 2×2 block matrix $\mathbf{J}$ in (11.9). Incorporating the effect of the constraint (11.3) on $\mathbf{f}$, we can derive from [241] the following constrained CRB for $\mathbf{f}$

$$\mathrm{CRB}_{\mathbf{f}} = \sigma^2\mathcal{V}_{\mathbf{f}}\left(\mathcal{V}_{\mathbf{f}}^H\mathscr{H}^H\mathscr{P}_{\mathscr{F}}^{\perp}\mathscr{H}\mathcal{V}_{\mathbf{f}}\right)^{-1}\mathcal{V}_{\mathbf{f}}^H, \tag{11.11}$$

where $\mathscr{P}_{\mathscr{X}} = \mathscr{X}(\mathscr{X}^H\mathscr{X})^H\mathscr{X}^H$ and $\mathscr{P}_{\mathscr{X}}^{\perp} = \mathbf{I}-\mathscr{P}_{\mathscr{X}}$ are the projection operators on resp. the column space of matrix $\mathscr{X}$ and its orthogonal complement, and H corresponds to the Moore-Penrose pseudo inverse. Note that in some group calibration scenarios, $\mathscr{F}^H\mathscr{F}$ can be singular (i.e, $\mathbf{h}$ could be not identifiable even if $\mathbf{f}$ is identifiable or even known). The $M\times(M-1)$ matrix $\mathcal{V}_{\mathbf{f}}$ is such that its column space spans the orthogonal complement of that of $\frac{\partial\mathscr{C}(f)}{\partial\mathbf{f}^*}$, i.e., $\mathscr{P}_{\mathcal{V}_{\mathbf{f}}} = \mathscr{P}_{\frac{\partial\mathscr{C}}{\partial\mathbf{f}^*}}^{\perp}$.

It is shown in [240], [241], [242] that a choice of constraints such that their linearized version $\frac{\partial\mathscr{C}}{\partial\mathbf{f}^*}$ fills up the null space of the FIM results in the lowest CRB, while not adding information in subspaces where the data provides information. One such choice is the set (11.3) (NPC). Another choice is (11.3) with $\mathbf{g} = \mathbf{f}$. With such constraints, $\frac{\partial\mathscr{C}}{\partial\mathbf{f}^*} \sim \mathbf{f}$ which spans the null space of $\mathscr{H}^H\mathscr{P}_{\mathscr{F}}^{\perp}\mathscr{H}$. The CRB then corresponds to the pseudo inverse of the FIM and (11.11) becomes $\mathrm{CRB}_{\mathbf{f}} = \sigma^2\left(\mathscr{H}^H\mathscr{P}_{\mathscr{F}}^{\perp}\mathscr{H}\right)^H$. If the FCC constraint is used instead (i.e., (11.3) with $\mathbf{g} = \mathbf{e}_1$, $c = 1$), where $\mathbf{e}_1$ is an all zero vector with only the first coefficient one, the corresponding CRB is (11.11) where $\mathcal{V}_{\mathbf{f}}$ corresponds now to an identity matrix without the first column (and hence its column space is the orthogonal complement of $\mathbf{e}_1$).

11.2.2 Variational Bayes (VB) Estimation

In VB, a Bayesian estimate is obtained by computing an approximation to the posterior distribution of the parameters $\mathbf{h},\mathbf{f}$ with priors $\mathbf{f}\sim\mathcal{CN}(0,\alpha^{-1}\mathbf{I_M})$, $\mathbf{h}\sim\mathcal{CN}(0,\beta^{-1}\mathbf{I_{N_h}})$ and α,β are assumed to have themselves a uniform prior. N_h is the number of elements in $\mathbf{h}$. This approximation, called the variational distribution, is chosen to minimize the Kullback-Leibler distance between the true posterior distribution $p(\mathbf{h},\mathbf{f},\alpha,\beta|\mathbf{y})$ and a factored variational distribution $q(\mathbf{h},\mathbf{f},\alpha,\beta|\mathbf{y}) = q_{\mathbf{h}}(\mathbf{h})\,q_{\mathbf{f}}(\mathbf{f})\,q_{\alpha}(\alpha)\,q_{\beta}(\beta)$. The factors can be obtained in an alternating fashion

as [243],

$$\ln(q_{\boldsymbol{\theta}_i}(\boldsymbol{\theta}_i)) = <\ln p(\mathbf{y},\mathbf{h},\mathbf{f},\alpha,\beta)>_{k\neq i} + c_i, \tag{11.12}$$

where $\boldsymbol{\theta}_i$ refers to the i^{th} block of $\boldsymbol{\theta} = [\mathbf{h},\mathbf{f},\alpha,\beta]$ and $<>_{k\neq i}$ represents the expectation operator over the distributions $q_{\boldsymbol{\theta}_k}$ for all $k \neq i$. c_i is a normalizing constant. Further considering the constraints on $\mathbf{f}$ ($\mathbf{f}_\perp$ represents the component of $\mathbf{f}$ in the null space of the constraint) and applying VB (11.12)

$$\begin{aligned}
\mathbf{f} &= \mathbf{f}' + \mathcal{V}_f\mathbf{f}_\perp,\\
\mathbf{f}' &= \mathbf{g}\frac{c}{\|\mathbf{g}\|^2},\\
\mathbf{f}^H\mathbf{g} &= c > 0,\\
\mathcal{V}_f^H\mathcal{V}_f &= \mathbf{I},\\
\ln q_{\mathbf{f}}(\mathbf{f}) &= \frac{1}{\sigma^2}(\mathbf{f}'^H + \mathbf{f}_\perp^H\mathcal{V}_f^H) <\mathscr{H}^H> \mathbf{y} + \frac{1}{\sigma^2}\mathbf{y}^H <\mathscr{H}> (\mathbf{f}' + \mathcal{V}_f\mathbf{f}_\perp)\\
&\quad - \frac{1}{\sigma^2}(\mathbf{f}'^H + \mathbf{f}_\perp^H\mathcal{V}_f^H) <\mathscr{H}^H\mathscr{H}> (\mathbf{f}' + \mathcal{V}_f\mathbf{f}_\perp) - <\alpha> \|\mathbf{f}_\perp\|^2 + c_f,\\
\ln q_{\mathbf{h}}(\mathbf{h}) &= \frac{\mathbf{h}^H <\mathscr{F}^H> \mathbf{y} + \mathbf{y}^H <\mathscr{F}> \mathbf{h} - \mathbf{h}^H <\mathscr{F}^H\mathscr{F}> \mathbf{h}}{\sigma^2} - <\beta> \mathbf{h}^H\mathbf{h}.
\end{aligned} \tag{11.13}$$

Here, N_y refers to the number of elements in $\mathbf{y}$ and c is a constant. Here c_p, c_f represents the normalization constants for the respective pdfs. We shall assume here that the noise variance σ^2 is known (or estimated in a separate training procedure). It is now straightforward to see that proceeding as in (11.12), α,β would have a Gamma distribution and a complex normal distribution for $\mathbf{f} \sim \mathscr{CN}(\widehat{f}, \mathbf{C}_{\tilde{f}\tilde{f}})$ and $\mathbf{h} \sim \mathscr{CN}(\widehat{\mathbf{h}}, \mathbf{C}_{\tilde{h}\tilde{h}})$. The detailed expressions are summarized in Algorithm 20. When $G = M$, $\mathbf{C}_{\tilde{f}\tilde{f}}$ and $\mathbf{C}_{\tilde{h}\tilde{h}}$ are diagonal and $<\mathscr{F}^H(\tilde{\mathbf{f}})\mathscr{F}(\tilde{\mathbf{f}})>$, $<\mathscr{H}^H(\tilde{\mathbf{h}})\mathscr{H}(\tilde{\mathbf{h}})>$ can be computed easily (diagonal). However, when $G < M$, these matrices are block diagonal. An approximate version of Algorithm 20, EC-VB (Expectation Consistent [244] VB) [237] where

Algorithm 20: VB Estimation of calibration parameters

1: **Initialization:** Initialize $\widehat{f}$ using existing calibration methods. Use $\widehat{f}$ to determine $\widehat{\mathbf{h}}, <\alpha>, <\beta>$, with $\mathbf{g} = \mathbf{e}_1$.
2: **repeat**
3: $<\mathscr{H}^H\mathscr{H}> = \mathscr{H}^H(\widehat{\mathbf{h}})\mathscr{H}(\widehat{\mathbf{h}}) + <\mathscr{H}^H(\tilde{\mathbf{h}})\mathscr{H}(\tilde{\mathbf{h}})>$.
4: $\widehat{f}_\perp = (\mathcal{V}_f^H(<\mathscr{H}^H\mathscr{H}> + \sigma^2 <\alpha> \mathbf{I})\mathcal{V}_f)^{-1}\mathcal{V}_f^H(<\mathscr{H}^H> \mathbf{y} - <\mathscr{H}^H\mathscr{H}> \mathbf{f}')$
5: $\mathbf{C}_{\tilde{f}\tilde{f}} = \mathcal{V}_f(\mathcal{V}_f^H(\frac{1}{\sigma^2} <\mathscr{H}^H\mathscr{H}> + <\alpha> \mathbf{I})\mathcal{V}_f)^{-1}\mathcal{V}_f^H$
6: $<\mathscr{F}^H\mathscr{F}> = \mathscr{F}^H(\widehat{f})\mathscr{F}(\widehat{f}) + <\mathscr{F}^H(\tilde{\mathbf{f}})\mathscr{F}(\tilde{\mathbf{f}})>$
7: $\widehat{\mathbf{h}} = (<\mathscr{F}^H\mathscr{F}> + \sigma^2 <\beta> \mathbf{I})^{-1}\mathscr{F}^H\mathbf{y}$, $\mathbf{C}_{\tilde{h}\tilde{h}} = (\frac{1}{\sigma^2} <\mathscr{F}^H\mathscr{F}> + <\beta> \mathbf{I})^{-1}$, $<\alpha> = \frac{M}{<\|\mathbf{f}_\perp\|^2>}$, $<\|\mathbf{f}_\perp\|^2> = \widehat{f}_\perp^H\widehat{f}_\perp + tr\{\mathbf{C}_{\tilde{f}_\perp\tilde{f}_\perp}\}$.
8: $<\beta> = \frac{N_h+1}{<\|\mathbf{h}\|^2>}$, $<\|\mathbf{h}\|^2> = \widehat{\mathbf{h}}^H\widehat{\mathbf{h}} + tr\{\mathbf{C}_{\tilde{h}\tilde{h}}\}$.
9: **until** convergence.

the error covariance matrix are approximated to be multiple of identity is also considered in the simulations. Note here that by forcing the matrices $\mathbf{C}_{\tilde{f}\tilde{f}}$, $\mathbf{C}_{\tilde{h}\tilde{h}}$ to zero and α,β to zero, this algorithm reduces to the Alternating Maximum Likelihood (AML) algorithm [238, 239] which iteratively maximizes the likelihood by alternating between the desired parameters $\mathbf{f}$ and the nuisance

parameters **h** for the formulation (11.8). The penalized ML method used in [245] uses quadratic regularization terms for both **f** and **h** which can be interpreted as Gaussian priors and which may improve estimation in ill-conditioned cases. In our case, we arrive at a similar solution from the VB perspective and more importantly, the regularization terms are optimally tuned.

11.3 Mismatched CRB's

As can be seen in Fig. 11.2, VB allows us to attain lower MSE than the CRB (for deterministic parameters). One possibility to evaluate the performance is to consider the Bayesian CRB. However, VB is an approximate Bayesian estimation technique. Also, a Bayesian CRB is valid only if the (Gaussian) priors for **f** and **h** are the correct priors. However, the interest of the VB technique is that it will converge to the most appropriate priors even if the parameters **f** and **h** are deterministic! This requires Mismatched CRBs. In this chapter, we explore the Bayesian mCRB exposed in [234, 246].

Under a mismatched distribution model, it is important to define the convergence point $\overline{\boldsymbol{\theta}}$ (also called as a pseudo true parameter) which is used to evaluate the effectiveness of the estimator, since no true parameter vector may exist under the assumed distribution q. The VB convergence point (of complete $\boldsymbol{\theta}$) is the MAP of $E_p(\sum_i \ln(q_{\boldsymbol{\theta}_i}(\boldsymbol{\theta}_i)))$ (assuming large amount of data), so ln of a product of q's = sum of ln of q's and converges to its expected value according to actual pdf p (law of large numbers). Similar to [234] (misspecified CRBs) which considers deterministic case, we do it also for random $\boldsymbol{\theta}$, but not neglecting priors in the asymptotic regime (considering some fictitious asymptotic regime in which prior information scales similarly as information in data, so that both continue to count, but get a Gaussian concentration around the convergence point).

11.3.1 mCRB Bilinear Model

CRB corresponds to Laplace approximation of MAP or VB. Laplace approximation [243] refers to the evaluation of marginal likelihood or free energy using Laplace's method. This is equivalent to a Gaussian approximation of the posterior q around a maximum a posteriori (MAP) estimate, motivated by the fact that in the asymptotic limit (a large amount of data or high SNR), the posterior approaches a Gaussian around the MAP point [234]. Gradients of $\ln q$ can be taken from the recursions for $\ln q$ (11.12), so it is the gradients of $\ln p$ as usual, except with averaging over $q_{\bar{i}}$ for gradient and Hessian as we will show here. But the final error covariance matrix of Laplace approximation (2^{nd} order Taylor) is the expectation with p. Let $\widehat{\boldsymbol{\theta}}$ be an estimator of $\boldsymbol{\theta}$ based on the approximate posterior q and the assumed prior. Let $\boldsymbol{\zeta} = \widehat{\boldsymbol{\theta}} - \overline{\boldsymbol{\theta}}$, where the estimator mean is evaluated at the point $\overline{\boldsymbol{\theta}}$. First, we need to find $\overline{\boldsymbol{\theta}}$. This corresponds to the peak of the posterior pdf in an asymptotic scenario of a large amount of data or high SNR, computation of which is derived in 11.3.2. Throughout the chapter, the vector $\boldsymbol{\theta}_i$ represents a subset of $\boldsymbol{\theta}$ and θ_i represents a scalar parameter in $\boldsymbol{\theta}$. In this section $\boldsymbol{\theta}$ (a column vector) contains the parameters $\mathbf{h}, \mathbf{f}$ and ψ the precision parameters, α, β and $\boldsymbol{\theta}^0$ denote the true value of $\boldsymbol{\theta}$. $\overline{\boldsymbol{\theta}}$ (or $\overline{\psi}$) can be evaluated as

$$\begin{aligned} \overline{\boldsymbol{\theta}}_i &= \arg\max_{\boldsymbol{\theta}_i} \mathrm{E}_{p(\mathbf{y},\boldsymbol{\theta}^0)} \ln q(\boldsymbol{\theta}_i) \\ &= \arg\max_{\boldsymbol{\theta}_i} \mathrm{E}_{p(\mathbf{y}|\boldsymbol{\theta}^0)} \ln < p(\mathbf{y},\boldsymbol{\theta}) >_{\bar{i}} . \end{aligned} \tag{11.14}$$

Even though the parameters are modeled as random for estimation, but we assume that in reality they are deterministic. So the expectation over $p(\boldsymbol{\theta})$ disappears in (11.14). Also, we define $\widetilde{\overline{\boldsymbol{\theta}}} =$

$\overline{\boldsymbol{\theta}} - \boldsymbol{\theta}^0$, $\widetilde{\boldsymbol{\theta}} = \widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0 = \boldsymbol{\zeta} + \widetilde{\overline{\boldsymbol{\theta}}}$. For any choice of score function $\boldsymbol{\eta}$ using a matrix generalization of the Cauchy Schwartz inequality [231, 246], the error correlation matrix can be written as

$$\mathbf{mCRB} = \mathbf{R}_{\widetilde{\theta}\widetilde{\theta}} = \mathrm{E}_p \widetilde{\boldsymbol{\theta}}\widetilde{\boldsymbol{\theta}}^H \geq \mathbf{R}_{\zeta\eta}\mathbf{R}_{\eta\eta}^{-1}\mathbf{R}_{\eta\zeta} + \widetilde{\overline{\boldsymbol{\theta}}}\widetilde{\overline{\boldsymbol{\theta}}}^H, \tag{11.15}$$

where $\mathbf{R}_{\zeta\eta} = \mathrm{E}(\boldsymbol{\zeta}\boldsymbol{\eta}^H)$ and $\mathbf{R}_{\zeta\zeta} = \mathrm{E}(\boldsymbol{\zeta}\boldsymbol{\zeta}^H)$.

The score function can be written as

$$\begin{aligned}\boldsymbol{\eta} &= \frac{\partial}{\partial \boldsymbol{\theta}^*} \ln q(\boldsymbol{\theta}) \mid_{\overline{\boldsymbol{\theta}}} - \mathrm{E}_{p(\mathbf{y}|\boldsymbol{\theta}^0)} \frac{\partial}{\partial \boldsymbol{\theta}^*} \ln q(\boldsymbol{\theta}) \mid_{\overline{\boldsymbol{\theta}}} \\ &= \frac{\partial}{\partial \boldsymbol{\theta}^*} \ln q(\boldsymbol{\theta}) \mid_{\overline{\boldsymbol{\theta}}} - \mathrm{E}_{p(\mathbf{y}|\boldsymbol{\theta}^0)} (\frac{\partial}{\partial \boldsymbol{\theta}^*} \ln q(\boldsymbol{\theta}) \mid_{\overline{\boldsymbol{\theta}}}).\end{aligned} \tag{11.16}$$

The choice of the score function is motivated by the requirements for the tightness of the CRB detailed in [231] that it should be zero mean and depends on the sufficient statistic for estimating $\boldsymbol{\theta}$. So the score function here is the score function for the deterministic CRB minus its possibly non zero-mean under the true model $p(\mathbf{y}, \boldsymbol{\theta}^0)$. Also, the particular choice score function (11.16) results in $\mathrm{E}_{p(\mathbf{y}|\boldsymbol{\theta}^o)}(\frac{\partial}{\partial \boldsymbol{\theta}^*} \ln q(\boldsymbol{\theta}) \mid_{\overline{\boldsymbol{\theta}}}) = 0$, due to the Laplace approximation of $\boldsymbol{\theta}$ around the asymptotic estimate $\overline{\boldsymbol{\theta}}$. Further, under concentration conditions (data asymptotics or SNR asymptotics, or perhaps prior asymptotics (becoming very precise)), we can do a 2^{nd} order Taylor series of misspecified posterior. The Taylor series expansion of the data likelihood around $\overline{\boldsymbol{\theta}}$ is given by

$$\begin{aligned}\log q(\mathbf{y}, \overline{\boldsymbol{\theta}} + \Delta\theta) = \log q(\mathbf{y}, \overline{\boldsymbol{\theta}}) + \Delta\boldsymbol{\theta}^H \frac{\partial \log q(\mathbf{y}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}^*} \mid_{\overline{\boldsymbol{\theta}}} + \\ \Delta\boldsymbol{\theta}^H \frac{\partial^2 \log q(\mathbf{y}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}^* \boldsymbol{\theta}^T} \mid_{\overline{\boldsymbol{\theta}}} \Delta\boldsymbol{\theta} + o(\|\Delta\boldsymbol{\theta}\|^2).\end{aligned} \tag{11.17}$$

Further neglecting the higher order terms and equating the derivative w.r.t $\Delta\boldsymbol{\theta}^*$ to be zero yields an approximation of the error term $\boldsymbol{\zeta}$ as

$$\boldsymbol{\zeta} = -(\frac{\partial^2 \log q(\mathbf{y}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}^* \boldsymbol{\theta}^T} \mid_{\overline{\boldsymbol{\theta}}})^{-1} \frac{\partial \log q(\mathbf{y}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}^*} \mid_{\overline{\boldsymbol{\theta}}}. \tag{11.18}$$

Note that we can replace the Hessian and $\frac{\partial \log q(\mathbf{y}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}^*}$ in (11.18) by $\mathrm{E}_{p(\mathbf{y}|\boldsymbol{\theta})}(\frac{\partial^2 \log q(\mathbf{y}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}^* \boldsymbol{\theta}^T})$ and $\mathrm{E}_{p(\mathbf{y}|\boldsymbol{\theta})}(\frac{\partial \log q(\mathbf{y}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}^*})$ respectively in the asymptotic limit. Taking the derivative of the data log-likelihood gives

$$\begin{aligned}\frac{\partial \log q(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^*} &= -\frac{1}{\sigma^2} \begin{bmatrix} 0 \\ \mathcal{V}_f^H < \mathcal{H}^H \mathcal{H} > \mathbf{f} - \mathcal{V}_f^H < \mathcal{H} > \mathbf{y} + < \alpha > \mathbf{f}_\perp \\ < \mathcal{F}^H \mathcal{F} > \mathbf{h} - < \mathcal{F}^H > \mathbf{y} + < \beta > \mathbf{h} \end{bmatrix}, \\ \mathrm{E}_{p(\mathbf{y}|\boldsymbol{\theta})} \frac{\partial^2 \log q(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^* \boldsymbol{\theta}^T} &= -\mathcal{V}^H \mathbf{Q} \mathcal{V}, \\ \mathbf{Q} &= \frac{1}{\sigma^2} \mathrm{blkdiag}(0, \mathcal{V}_f^H < \mathcal{H}^H \mathcal{H} > \mathcal{V}_f + < \alpha > \mathbf{I}, < \mathcal{F}^H \mathcal{F} > + < \beta > \mathbf{I}).\end{aligned} \tag{11.19}$$

where blkdiag$(\cdot)$ represents the block diagonal matrix formed by the respective matrix elements in the block. The evaluation of $\mathbf{Q}$ at the asymptotic limit, $\overline{\boldsymbol{\theta}}$, be denoted as $\overline{\mathbf{Q}}$. Let $\mathrm{E}(\frac{\partial \log q(\mathbf{y}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}^*}) \mid_{\overline{\boldsymbol{\theta}}} = f(\overline{\boldsymbol{\theta}})$. The error term $\boldsymbol{\zeta}$ can then be expressed as, $\boldsymbol{\zeta} = \mathcal{V}(\mathcal{V}^H \overline{\mathbf{Q}} \mathcal{V})^{-1} \mathcal{V}^H f(\mathbf{y}, \boldsymbol{\theta})$. Note that $\mathcal{V} = [\mathbf{0}\ \mathbf{I}]$. The cross correlation matrix between $\boldsymbol{\zeta}$ and $\boldsymbol{\eta}$ becomes

$$\begin{aligned}\mathbf{R}_{\zeta\eta} &= -\mathcal{V}(\mathcal{V}^H \overline{\mathbf{Q}} \mathcal{V})^{-1} \mathcal{V}^H f(\mathbf{y}, \boldsymbol{\theta}) f(\mathbf{y}, \boldsymbol{\theta})^H \\ &= -(\mathcal{V}(\mathcal{V}^H \overline{\mathbf{Q}} \mathcal{V})^{-1} \mathcal{V}^H f(\mathbf{y}, \boldsymbol{\theta}) f(\mathbf{y}, \boldsymbol{\theta})^H).\end{aligned} \tag{11.20}$$

Here $f(\overline{\boldsymbol{\theta}})f(\overline{\boldsymbol{\theta}})^H = \mathbf{J}_q$. Finally substituting (11.20) in (11.15), we obtain (define **MFIM** to be the corresponding mismatched FIM)

$$\mathbf{mCRB} = \mathcal{V}(\mathcal{V}^H\overline{\mathbf{Q}}\mathcal{V})^{-1}\mathcal{V}^H\mathbf{J}_q\mathcal{V}(\mathcal{V}^H\overline{\mathbf{Q}}\mathcal{V})^{-1}\mathcal{V}^H + \widetilde{\overline{\boldsymbol{\theta}}}\widetilde{\overline{\boldsymbol{\theta}}}^H. \tag{11.21}$$

Further we derive the mCRB for VB ($\mathbf{mCRB}_{VB}$) with the posteriors of $\mathbf{h},\mathbf{f}$ being factorized.

Lemma 8. *If the parameter partitioning in VB is such that the different parameter blocks are decoupled at the level of the Fisher Information Matrix, then VB is not suboptimal in terms of (mismatched) Cramer-Rao Bound. If a finer partitioning granularity is used (such as up to scalar level as in mean field), then VB becomes quite suboptimal.*

So in the too fine partitioning case, the VB partitioning is applied to the MFIM, taking a too fine block diagonal part, and since that partitioning is finer than the block diagonal MFIM structure, then the inverse of the too fine block diagonal part of the FIM does not give the correct CRB. So $\mathbf{mCRB}_{VB} = (\text{blockdiag}(\mathbf{MFIM}))^{-1} \neq \mathbf{mCRB}$.

$$\begin{aligned}\mathbf{mCRB}_{VB} &= \mathcal{V}_f(\mathcal{V}_f^H(\mathbf{A}_{\mathbf{f},\mathbf{f}})^{-1}\mathcal{V}_f)^{-1}\mathcal{V}_f^H + \widetilde{\overline{\boldsymbol{\theta}}}\widetilde{\overline{\boldsymbol{\theta}}}^H \\ \mathbf{A} &= \mathcal{V}(\mathcal{V}^H\overline{\mathbf{Q}}\mathcal{V})^{-1}\mathcal{V}^H\mathbf{J}_q\mathcal{V}(\mathcal{V}^H\overline{\mathbf{Q}}\mathcal{V})^{-1}\mathcal{V}^H,\end{aligned} \tag{11.22}$$

$\mathbf{A}$ evaluated at $\overline{\boldsymbol{\theta}}$, $\mathbf{A}_{\mathbf{ff}} = (\mathbf{f},\mathbf{f})$ block of $\mathbf{A}$ (here it is the product of block diagonal of 3 factors), mCRB above for given $\boldsymbol{\theta}^o$. Some remarks which follow from our mCRB analysis are stated below.

- mCRB in this chapter are along the lines of [231] and it is applicable to all estimators with same bias and cross-correlation matrix.
- This mCRB, is mismatched because we introduce an artificial prior. Asymptotically (i.e. at high SNR), the MSE of either alternating MAP (AMAP) or VB or EC-VB should match this mCRB.
- Asymptotically, the suboptimality of VB is not in its mean, it is only in the approximation of the error covariance, which should underestimate the actual error covariance: $[(\mathbf{J}_q)^{-1}]_{1,1} > ((\mathbf{J}_q)_{1,1})^{-1}$.
- Our view point of first working for given $\boldsymbol{\theta}$ is compatible with the view that actually the $\boldsymbol{\theta}$ may be be deterministic (prior for $\boldsymbol{\theta} = \delta(\boldsymbol{\theta} - \boldsymbol{\theta}^0)$, dirac delta function at true value) and the idea of doing Bayesian or VB is just to create a bias so that the biased estimator would reach lower MSE, in particular below the CRB. James-Stein estimator [247] was the first instance of this. In case of James-Stein, they are able to show that the deterministic MSE is lowered by adding the prior (with optimized/estimated variance hyperparameter). Then VB (with estimated = optimized hyperparameters) is a way of making sure that this bias is useful, optimizes MSE in some sense, within the class of estimators determined by the stucture of the prior chosen. In other words, these Bayesian estimators provide a way to introduce a useful bias (shrinkage) that allows to lower MSE (from the point of view of deterministic parameters, with a single true value).

11.3.2 Computation of Convergence Point

Starting from (11.14), the resulting (deterministic) $\overline{\boldsymbol{\theta}}$ is obtained by running alternating MAP (initialized by the true $\boldsymbol{\theta}^o$). Or one can also run the VB, by putting $\mathbf{n} = 0$ in $\mathbf{y}$, and considering

$\widetilde{\mathbf{h}} = 0, \widetilde{\mathbf{f}} = 0$,hence also $C_{\widetilde{\mathbf{h}}\widetilde{\mathbf{h}}} = 0, C_{\widetilde{\mathbf{f}}\widetilde{\mathbf{f}}} = 0$. So, the VB converges to $\overline{\boldsymbol{\theta}}$. For computing $\bar{\mathbf{f}}$, substituting for $\mathbf{y} = \mathscr{H}^0\mathbf{f}^0 + \mathbf{n}$ in (11.14) (similarly for the computation of $\overline{\mathbf{h}}$, need to consider the alternative representation of $\mathbf{y}$ (11.8))

$$\begin{aligned} \mathrm{E}_{p(\mathbf{y}|\boldsymbol{\theta})\ln<p(\mathbf{y},\boldsymbol{\theta},\psi)>_{\tilde{i}}} = &-N_y \ln\sigma^2 - \frac{1}{\sigma^2}(< \left\|\mathscr{H}^0\mathbf{f}^0 - \mathscr{H}\mathbf{f}\right\|^2 > + \sigma^{2,0}N_y) \\ &+ (M-1) < \ln\alpha > - < \alpha > \|\mathbf{f}_\perp\|^2 + N_h < \ln\beta > - < \beta >< \|\mathbf{h}\|^2 > + c. \end{aligned} \tag{11.23}$$

The derivative of (11.23) w.r.t $\mathbf{f}, \alpha, \beta, \mathbf{h}$ leads to Algorithm 21. Note that the Algorithm 21 applies

Algorithm 21: Computation of Asymptotic Estimates, $\overline{\boldsymbol{\theta}}$

1: **Initialization:** Initialize $\bar{\mathbf{f}}$ using existing calibration methods ($\bar{\mathbf{f}} = \mathbf{f}' + \mathcal{V}_f\bar{\mathbf{f}}_\perp$).
2: **repeat**
3: $\bar{\mathbf{f}}_\perp = (\mathcal{V}_f^H\overline{\mathscr{H}}^H\overline{\mathscr{H}}\mathcal{V}_f + \sigma^2\overline{\alpha}\mathbf{I})^{-1}(\mathcal{V}_f^H\overline{\mathscr{H}}^H\mathscr{H}^0\mathbf{f}^0 - \mathcal{V}_f^H\overline{\mathscr{H}}^H\overline{\mathscr{H}}\mathbf{f}')$.
4: $\overline{\mathbf{h}} = (\overline{\mathscr{F}}^H\overline{\mathscr{F}} + \sigma^2\overline{\beta}\mathbf{I})^{-1}(\overline{\mathscr{F}}^H\mathscr{F}^0\mathbf{h}^0)$.
5: $\overline{\alpha} = \frac{M}{<\left\|\bar{\mathbf{f}}_\perp\right\|^2>}$, $\overline{\beta} = \frac{N_h+1}{<\left\|\overline{\mathbf{h}}\right\|^2>}$, $\sigma^2 = \sigma^{2,0} + \frac{1}{N_y}\left\|\mathscr{H}^0\mathbf{f}^0 - \overline{\mathscr{H}}\overline{\mathbf{f}}\right\|^2$.
6: **until** convergence.

to any partitioning of the variables in the approximate posterior q, where for VB (11.12), there will only be one iteration with the initial values for $\overline{\mathscr{H}}^H\overline{\mathscr{H}} =< \mathscr{H}^H\mathscr{H} >$ or $\overline{\mathscr{H}} =< \mathscr{H} >$ (by the converged values of VB).

11.4 Simulations

In this section, we assess numerically the performance of various calibration algorithms and also compare them against their CRBs. The Tx and Rx calibration parameters for the BS antennas are assumed to have random phases uniformly distributed over $[-\pi, \pi]$ and amplitudes uniformly distributed in the range $[1-\delta, 1+\delta]$. SNR is defined as the ratio of the average received signal power across channel realizations at an antenna and the noise power at that antenna. In Fig. 11.2, it is clear that VB MSE can go lower than the determinsitic CRB and close to the mCRB. In Fig. 11.3, we compare the MSE performance of various VB variants with $mCRB_{VB}$ and deterministic CRB. It shows the performance improvement of VB w.r.t deterministic CRB or AML at all SNR and also the accurate behaviour of our derived mCRB expressions. We consider transmit schemes that transmit from one antenna at a time ($G = M$) and compare their MSE performance with the CRB. The MSE with FCC for Argos, Rogalin [248] and the VB method in Algorithm 20 is plotted. The curves are generated over one realization of an i.i.d. Rayleigh channel and known first coefficient constraint is used. These curves are compared with the CRB derived in 11.2.1 for the FCC case and it can be seen that the AML curve overlaps with the CRB at higher SNRs. Also plotted is the CRB as given in [245] assuming the internal propagation channel is fully known (the mean is known and the variance is negligible) and a (small) underestimation of the MSE can be observed as expected.

11.5 Conclusions

In this chapter, we came up with a simple and elegant derivation of the mCRB for a general calibration framework that includes as subsets all existing calibration techniques. For the case of

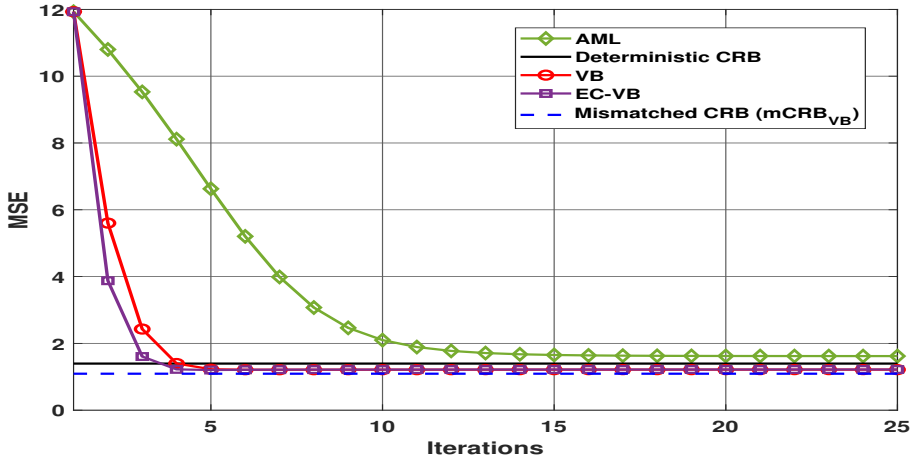


Figure 11.2: Convergence of the various iterative schemes for $M = G = 16$.

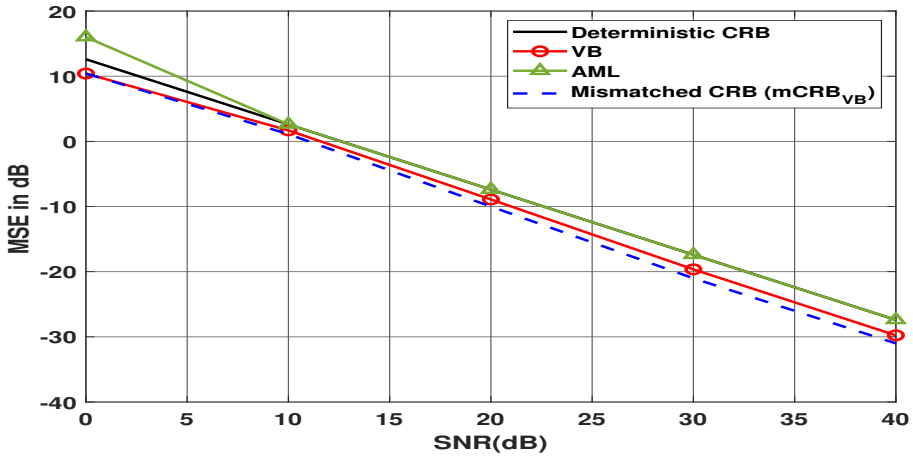


Figure 11.3: Comparison of single antenna transmit schemes with the CRB ($G = M = 16, L_i = 1, \forall i, \delta = 0.5$).

groups involving a single antenna, the conventional CRB derivation assuming the first coefficient known has also been provided. An optimal estimation algorithm based on VB is also introduced along with its variants. We further derived mismatched CRB to validate the performance improvement over deterministic CRB. All these techniques have been compared via simulations in terms of both MSE performance and speed of convergence.

Chapter 12

CONCLUSIONS AND FUTURE WORK

In this thesis, we looked at various aspects of massive MIMO communications. In this final chapter, we provide some concluding remarks for each of three parts of this thesis. Furthermore, we look at various possible straightforward extensions to the current work here and also some future topics which needs to be explored. We try to highlight both the pros and cons of our various proposed solutions and some possible future directions to circumvent the disadvantages associated with them.

12.1 Beamforming Techniques for Massive MIMO

The first part of the thesis focused on beamforming techniques for massive MIMO, focusing on maximizing the sum throughput across all users in the network. One potential solution to circumvent the hardware complexity and power consumption issues in MaMIMO is hybrid beamforming. First part of the thesis started with efficient HBF solutions based on optimizing the WSR. Further, we moved to fully digital solutions under partial CSIT. It has to be mentioned that all the digital BF solutions proposed here can be easily extended to HBF case. We specifically looked at pathwise CSIT, where only the fast fading components are assumed to be unknown. We also looked at pilot contamination issue in MaMIMO networks. We proposed an efficient solution based on the concept of rate splitting to mitigate the pilot contamination issue. Following are the main conclusions on the first part on BF techniques.

- Hybrid beamforming (HBF) provides both beamforming (BF) gain (using analog phase shifters) and spatial multiplexing using digital beamformer.
- For a multi-cell multi-user MIMO system, we proposed a joint beamforming design and power allocation algorithm by maximizing the weighted sum rate (WSR). Weighted sum rate maximization is highly non convex due to the objective function being non convex and non convex constraints (unit modulus constraints on analog BF). Hence existing works mostly focus on suboptimal approaches whereas we try to solve the original WSR problem.
- For the analog phasors, we used a technique called deterministic annealing which can track the global optimum solution starting from the unconstrained analog BF. Our solution surpasses (in terms of spectral efficiency) the existing state of the art designs which are suboptimal.
- In the follow up works, we also extended our design to the case of 1) wideband OFDM system, 2) HBF under per-antenna power constraints which are more realistic and 3) full

duplex backhaul link under a limited dynamic range noise model (which approximates the non linearities in the RF chain).

- For the full duplex system, we considered for the first time in the literature an analog BF which mitigates the self interference such that it avoids the ADC saturation in the receive chain.
- One disadvantage of our above mentioned HBF designs is high computational complexity which may not be practically feasible to implement. However, our algorithms are still useful as a benchmark solutions for the existing suboptimal solutions in the literature. So one possible future direction is to look at a low complexity design, possible approach could be based on deep learning techniques or an efficient minorizer for the WSR which leads to a low complexity BF solution.
- Our HBF solutions can be easily extended for the case of partial CSIT and one possible direction is to use the upper bound of EWSR (ESIP-WSR) to optimize the precoders.
- Another future direction is to look at efficient channel estimation schemes in the case of hybrid systems under mmWave OFDM systems. This can be challenging since at the baseband side, we only observe a low dimensional channel. One possible strategy will be to look at sparse Bayesian learning techniques discussed in this thesis.
- In the next chapters, we move to fully digital BF solutions in MaMIMO. As the digital processing capabilities continue to increase tremendously, it is possible that in the future fully digital solutions may still become feasible for MaMIMO. Moreover, for the spectral efficiency analysis, it becomes more simplistic to consider fully digital solutions. Firstly, we consider a robust BF design under pwCSIT, wherein the BFs are optimized using an upper bound of the EWSR (called ESIP-WSR). We show through analysis and simulation results the extreme SNR behaviour of the pwCSIT based BF design. It can be shown that at high SNR, the ZF across the paths occur (assuming antenna dimensions are enough) and the task gets split between Tx and Rx precoders.
- In the uplink of MaMIMO TDD system, UEs that transmit the same pilot signal contaminate each others channel estimates. This pilot interference not only reduces the CSI quality but also creates the so-called coherent interference, which has been believed to fundamentally limit the spectral efficiency (SE) of MaMIMO, even when $M \to \infty$.
- The aim of our work is to deal with the pilot contamination effect for a finite M in a single cell multi-user MIMO system.
- A possible solution: A rate splitting (RS) approach that splits the UEs messages into common and private parts, encode the commmon parts into a common stream, and private parts into private streams and superpose in a non-orthogonal manner the common stream on top of all private streams. Common stream carriers a part of the message of all or a subset of UEs in the cell. At each UE, the common stream is first decoded, by treating the interference from the private streams as noise. Further successive interference cancellation (thus partially canceling the interference) is done to decode the private stream of the user.
- A maximum ratio (MR) precoding scheme is used for private streams while a precoder based on a weighted combination of the channel estimates of all UEs is adopted for the common stream.

- A novel algorithm is proposed to allocate the power among the common and private streams. In simulations: we observe that the RS scheme does help to mitigate the pilot contamination effect for a finite number of antennas. As K increases, the gain provided by RS decreases. The larger K, the lower the common rate since the common message has to be decoded by all UEs. First time in the literature to consider RS scheme to mitigate the pilot contamination.

12.2 Asymptotic Analysis for Massive MIMO

In MaMIMO systems, the signal and interference powers converge to deterministic quantities (this is called as channel hardening) due to the law of large numbers. This in turn leads to deterministic expressions for the SINR and rate. In this part, we look at spectral efficiency computation using large system analysis results in random matrix theory. Our aim here is to obtain simplified sum rate expressions which give intuitive understanding of the network behavior. Here are the main results and some future perspectives on this topic which we looked at.

- As a starting point for large system analysis, we looked at a simplistic scenario wherein the user channel covariance matrices are treated as a multiple of identity. This scale factor indeed represents the channel attenuation and can be different for different users. One direct implication of this assumption (which may not be very practical) is that different users may have the same scattering geometry around it and this can occur possibly in the case when all users are colocated. Further, we proposed a simplified BF scheme called RO-ZF BF. We showed that the SE performance of RO-ZF BF is quite close to that of the optimal MMSE BF. We obtained simplified sum rate expressions for RO-ZF, ZF and ZF-DPC BFs for both perfect CSIT and partial CSIT case. In the later chapters, we started looking at more complex channel models, where the user channel covariance matrices are all distinct.

- Why large system analysis? Monte-Carlo simulations involving large numbers of antennas and user equipments (UEs) become cumbersome in MaMIMO. Existing large system results on multi-user massive MIMO systems consider spatially uncorrelated fading, which is quite unrealistic.

- In a multi-user MaMIMO system, transmit correlation diversity (different users may have different covariance matrices spanning mutually orthogonal subspaces or at least linearly independent) can be beneficial.

- We consider a stochastic geometric inspired randomization of the user covariance subspaces, i.e. users are randomly placed across the network and this random placing leads to AoA at the BS which are non overlapping for different users.

- We consider a partial CSIT model at the BS side. We consider for our analysis two case of channel estimation error, 1) error being inversely proportional to SNR and 2) constant channel estimation error scenario (finite rate feedback systems (FDD) and pilot contamination).

- We propose a BF design for maximizing an upper bound of the ergodic capacity which is tight in certain massive MIMO limit (M and $K \to \infty$, M being BS antennas and K users).

- We uses an extension of the random matrix theory results in the literature to derive deterministic equivalents of the SINR and rate of different users. We compare both analytically and numerically the spectral efficiency performance of our proposed partial CSIT BF design and other suboptimal designs in the literature under different kind of channel estimates, for e.g. least squares (LS), LMMSE and a subspace projected version of the LS estimate.
- Our simulation results show that the large system approximations we derived are quite accurate even for finite system dimensions and provide useful insights about the system behaviour. The resulting asymptotic expressions are very simple and depends only on few system parameter such as channel covariance rank, M, K, transmit SNR and large scale fading coefficients.
- So many open issues remain here. We are currently looking at an extension of the large system analysis to the case of massive MIMO (hence with multiple antennas at the UE side).
- It would be of interest to extend the current analysis to the case when the number of multipaths exceed that of the number of antennas. In this case, ZF (for CoCSIT or ESIP-WSR based BF designs) will not be possible at high SNR.
- In classical stochastic geometry, state of the art works assign certain random distributions to the user location. Further, they derive the random channel attenuations based on the assumed distribution. In this thesis, we assume that the channel attenuation part (represented by $\lambda_{k,c}$) are known. However, it would be of great interest to understand the system performance when the number of paths exceed that of the number of antennas (overloaded systems).
- Stochastic geometry based large system analysis for the case of potential next-generation technologies like cell-free massive MMO and rate splitting.

12.3 Approximate Bayesian Inference for Sparse Bayesian Learning

As remarked before, even though the main motivation behind the SBL algorithms proposed herein are for sparse wireless channel estimation, it has multiple application even beyond the realm of communication technologies. Another remark here is that unlike the state of the art solutions which restrict the measurement matrices to be parametrized in terms of channel path responses (which are Vandermonde), we do not consider a parametrized matrices in the first stage. Motivation behind this consideration is the practical hardware issues which make the path responses to be far from Vandermonde. Below are the main conclusions on this part of the thesis.

- Massive MIMO or mm Wave channels are sparse in the angular or delay domain (consider an OFDM system) due to the limited number of scattering components in the environment. Leveraging on the sparsity assumptions in the underlying channel which is based on scattering geometry of the propagation environment, we propose a Bayesian compressive sensing algorithm called as space alternating variational estimation (SAVE).
- Original Bayesian compressive sensing algorithm termed as sparse Bayesian learning (SBL) does not scale with the data dimensions due to the underlying matrix inversion associated

with it. Hence we consider a Variational Bayesian inference (VBI) based method called SAVE. In VBI, we try to compute an approximate posterior such that it minimizes the KL divergence between the true posterior and the approximate posterior which is assumed to be factorized (independent) at the scalar level. Compared to the state of the art methods our solution has better normalized MSE performance and has faster convergence, which makes it a practically viable solution.

- Extensions considered: 1) Time varying sparse multi path complex fading coefficients, temporal correlation modeled by a first order auto-regressive process. 2) Dictionary learning (in which we learn the AoA, AoD and the delays also) Disadvantage of SAVE method: SAVE is quite suboptimal in terms of the posterior variance computation. Hence we proposed belief propagation based posterior approximation which can be efficiently implemented using approximate message passing algorithms (AMP) which became popular in the literature in recent years.

- However, we observe that BP may not always converge. When BP converges, it always converge to the true LMMSE solution. Moreover, under certain conditions on the measurement matrix (under i.i.d. Gaussian entries in $\mathbf{A}$), we show that the per-component MSE converges to the true posterior variance. Even though, it is shown in the literature that AMP converges to the true MMSE value under i.i.d. Gaussian $\mathbf{A}$, it is the first time in the literature that the Bayes optimality in terms of per-component MSE is shown. Still, BP has its issues. It is quite sensitive to the characteristics of the measurement matrix.Even a slight deviation from i.i.d. Gaussian assumptions lead to very poor performance of AMP. Motivated by this, other extensions of AMP appeared in the literature such as GAMP and VAMP. However, the Bayes optimality for these algorithms are shown only for either i.i.d. Gaussian or right rotationally invariant $\mathbf{A}$ matrices case. These algorithms need not converge for every variations of $\mathbf{A}$ matrix.

- Motivated by the above arguments, several studies have appeared in the literature looking at convergent AMP alternatives. Among them, the most prominent one is the GSwAMP algorithm. In this thesis, we propose static and dynamic SBL algorithms using GSwAMP. Indeed, in the simulations, we were not able to arrive at a case when it diverges. GSwAMP variant is seen to be converging for non-zero mean, rank deficient and ill-conditioned $\mathbf{A}$ matrices, which itself is an interesting result.

- However, much remains to be done. What remains elusive here is which cost function (possibly a variant of BFE) leads to the alternating optimization as is done in GSwAMP. This is an interesting avenue for future work. One possibly direction is to check whether the concave-convex procedure (combined with an appropriate majorizer function for the non-convex part of the BFE) initially proposed in [249] can lead to GSwAMP or a better convergent alternative (with provable guarantees).

- Another future direction would be to explore the potential of deep learning for sparse linear inverse problems. This stream of research started with [250], which proposes a learned iterative soft thresholding algorithm (LISTA). This is derived by unfolding the iterations into the deep neural network (DNN) layers. In [251], the authors propose a learned AMP (LAMP) algorithm which has better performance than LISTA. The major benefit by moving to DNN as is noted down by all these papers is of faster convergence (by using far few DNN layers compared to the number of iterations in sparse linear inverse problems). Hence,

motivated by these methods, it will be of greater interest to analyze the potential of learned SBL algorithms.

Chapter 13

APPENDICES

A Derivation of LMMSE Estimation

In this section, we look at the derivation of the LMMSE estimate for a MIMO channel between user and the BS. The below derivation is used to obtain the LMMSE channel estimate throughout the thesis. We start from a deterministic least square estimate, $\widehat{\mathbf{H}}_{LS}$ which gets expressed as follows.

$$\begin{aligned} \widehat{\mathbf{H}}_{LS} &= \mathbf{H} + \widetilde{\mathbf{H}}_{LS}, \\ \text{where, } \mathbf{H} &= \mathbf{C}_r^{1/2} \mathbf{H}_v \mathbf{C}_t^{1/2}. \end{aligned} \tag{1}$$

$\widetilde{\mathbf{H}}_{LS}$ represents the estimation error which is independent of $\mathbf{H}$ and its entries are distributed as i.i.d. $\mathscr{CN}(0, \widetilde{\sigma}^2)$. Further vectorizing all the matrices above

$$\underbrace{vec(\widehat{\mathbf{H}}_{LS})}_{\widehat{\mathbf{h}}_{LS}} = \underbrace{vec(\mathbf{H})}_{\mathbf{h}} + \underbrace{vec(\widetilde{\mathbf{H}}_{LS})}_{\widetilde{\mathbf{h}}_{LS}}. \tag{2}$$

The vectorized channel $\mathbf{h}$ can be written as

$$\begin{aligned} \mathbf{h} &\overset{(a)}{=} (\mathbf{C}_t^{1/2} \otimes \mathbf{C}_r^{1/2})\, vec(\mathbf{H}_v) \\ \mathrm{E}(\mathbf{h}\mathbf{h}^H) &= \mathbf{C}_t \otimes \mathbf{C}_r, \text{ since } vec(\mathbf{H}_v) \sim \mathscr{CN}(\mathbf{0}, \mathbf{I}). \end{aligned} \tag{3}$$

where in (3), we used the property $vec(\mathbf{ABC}) = (\mathbf{C}^T \otimes \mathbf{A})\, vec(\mathbf{C})$. Further, computing the correlation matrix of $\widehat{\mathbf{h}}_{LS}$

$$\begin{aligned} \mathrm{E}(\widehat{\mathbf{h}}_{LS}\widehat{\mathbf{h}}_{LS}^H) &\overset{(b)}{=} \mathbf{R}_{\mathbf{hh}} + \mathrm{E}(\widetilde{\mathbf{h}}_{LS}\widetilde{\mathbf{h}}_{LS}^H) \\ &= \mathbf{R}_{\mathbf{hh}} + \widetilde{\sigma}^2 \mathbf{I} \\ &= \mathbf{C}_t \otimes \mathbf{C}_r + \widetilde{\sigma}^2 \mathbf{I}. \end{aligned} \tag{4}$$

Note that (b) is obtained by using the assumption that $\mathbf{H}$ and $\widetilde{\mathbf{H}}_{LS}$ are both zero mean and statistically independent. From standard principles of LMMSE estimation [20], the LMMSE estimate $\widehat{\mathbf{h}}$ can be obtained as

$$\widehat{\mathbf{h}} = \mathbf{R}_{\mathbf{h}\widehat{\mathbf{h}}_{LS}} \mathbf{R}^{-1}_{\widehat{\mathbf{h}}_{LS}\widehat{\mathbf{h}}_{LS}} \widehat{\mathbf{h}}_{LS}. \tag{5}$$

Now, we compute the cross-correlation term $\mathbf{R}_{\mathbf{h}\widehat{\mathbf{h}}_{LS}}$

$$\begin{aligned} \mathbf{R}_{\mathbf{h}\widehat{\mathbf{h}}_{LS}} &= \mathrm{E}(\mathbf{h}\widehat{\mathbf{h}}_{LS}^H) \\ &\overset{(c)}{=} \mathrm{E}(\mathbf{h}\mathbf{h}^H) \\ &= \mathbf{R}_{\mathbf{hh}} = \mathbf{C}_t \otimes \mathbf{C}_r. \end{aligned} \tag{6}$$

In the above equation, (c) is obtained using the statistical independence of $\mathbf{h}$ and $\widetilde{\mathbf{h}}_{LS}$ ($\mathrm{E}(\mathbf{h}\widetilde{\mathbf{h}}_{LS}) = \mathbf{0}$). By combining (4), (5), (6), we arrive at

$$\begin{aligned} \widehat{\mathbf{h}} &= (\mathbf{C}_t \otimes \mathbf{C}_r)\left(\mathbf{C}_t \otimes \mathbf{C}_r + \widetilde{\sigma}^2\mathbf{I}\right)^{-1}\widehat{\mathbf{h}}_{LS} \\ \text{or, } vec(\widehat{\mathbf{H}}) &= (\mathbf{C}_t \otimes \mathbf{C}_r)\left(\mathbf{C}_t \otimes \mathbf{C}_r + \widetilde{\sigma}^2\mathbf{I}\right)^{-1} vec(\widehat{\mathbf{H}}_{LS}), \end{aligned} \tag{7}$$

where $\widehat{\mathbf{H}}$ represents the LMMSE estimate in the matrix format.

Next, we look at the computation of the estimation error covariance matrix. Again, this follows from standard LMMSE principles [20].We denote $\widetilde{\mathbf{H}}$ as the LMMSE estimation error.

$$\begin{aligned} \mathbf{H} = \widehat{\mathbf{H}} + \widetilde{\mathbf{H}} &\implies \\ \mathbf{h} = \widehat{\mathbf{h}} + \widetilde{\mathbf{h}}&, \end{aligned} \tag{8}$$

where $\widehat{\mathbf{h}}$ and $\widetilde{\mathbf{h}}$ are uncorrelated, $\mathrm{E}(\widehat{\mathbf{h}}\widetilde{\mathbf{h}}^H) = \mathbf{0}$. Further, we can write the estimation error covariance matrix as

$$\begin{aligned} \mathrm{E}(\widetilde{\mathbf{h}}\widetilde{\mathbf{h}}^H) &= \mathbf{R}_{\widetilde{\mathbf{h}}\widetilde{\mathbf{h}}} \\ &= \mathbf{R}_{\mathbf{hh}} - \mathbf{R}_{\mathbf{h}\widehat{\mathbf{h}}_{LS}}\mathbf{R}^{-1}_{\widehat{\mathbf{h}}_{LS}\widehat{\mathbf{h}}_{LS}}\mathbf{R}_{\widehat{\mathbf{h}}_{LS}\mathbf{h}}, \end{aligned} \tag{9}$$

where, it can be shown that

$$\mathbf{R}_{\widehat{\mathbf{h}}_{LS}\mathbf{h}} = \mathbf{R}_{\mathbf{hh}} = \mathbf{C}_t \otimes \mathbf{C}_r. \tag{10}$$

Finally, by substituting the expressions in (4), (10) into (9), we obtain

$$\mathbf{R}_{\widetilde{\mathbf{h}}\widetilde{\mathbf{h}}} = (\mathbf{C}_t \otimes \mathbf{C}_r) - (\mathbf{C}_t \otimes \mathbf{C}_r)\left(\mathbf{C}_t \otimes \mathbf{C}_r + \widetilde{\sigma}^2\mathbf{I}\right)^{-1}(\mathbf{C}_t \otimes \mathbf{C}_r) \tag{11}$$

B Derivation of analog phasor design using WSMSE

Consider the terms involving V in the WSMSE.

$$\sum_{k=1}^{K}\left[u_k\, w_k\, e_k(\mathbf{f}_k, \mathbf{V}, \mathbf{g}) + \lambda_{b_k}|\mathbf{V}^{b_k}\mathbf{g}_k|^2\right] \tag{12}$$

where $e_k(\mathbf{f}_k, \mathbf{V}, \mathbf{g})$ is specified in (2.11). Now rewrite each term as a function of $V^{b_k}_{m,n}$. Consider e.g. the generic term

$$\begin{aligned} \mathbf{f}_k^H\mathbf{H}_{k,b_i}\mathbf{V}^{b_i}\mathbf{g}_i &= \sum_{j,l}(\mathbf{f}_k^H\mathbf{H}_{k,b_i})_j\, V^{b_i}_{j,l}\, g_{i,l} \\ &= V^{b_i}_{m,n}\, g_{i,n}(\mathbf{f}_k^H\mathbf{H}_{k,b_i})_m + C^{k,i}_{m,n} \\ \text{where } \; C^{k,i}_{m,n} &= \sum_{(j,l)\neq(m,n)}(\mathbf{f}_k^H\mathbf{H}_{k,b_i})_j\, V^{b_i}_{j,l}\, g_{i,l}. \end{aligned} \tag{13}$$

where $(\mathbf{f}_k^H\mathbf{H}_{k,b_i})_j$ denotes entry j of the row vector $\mathbf{f}_k^H\mathbf{H}_{k,b_i}$. $g_{i,n}$ denotes the n^{th} term of the vector $\mathbf{g}_i$. Substituting $\mathbf{V}^{b_i}_{m,n} = e^{j\theta^{b_i}_{m,n}}$ now yields

$$\left|\mathbf{f}_k^H\mathbf{H}_{k,b_i}\mathbf{V}^{b_i}\mathbf{g}_i\right|^2 = e^{j\theta^{b_i}_{m,n}}\, g_{i,n}(\mathbf{f}_k^H\mathbf{H}_{k,b_i})_m\, C^{k,i\,*}_{m,n} \tag{14}$$

$$+ e^{-j\theta^{b_i}_{m,n}}\, g^*_{i,n}(\mathbf{f}_k^H\mathbf{H}_{k,b_i})^*_m\, C^{k,i}_{m,n} + \text{"terms"} \tag{15}$$

where "terms" denotes terms that do not depend on $\theta^{b_i}_{m,n}$. Expanding the MSE $e_k(\mathbf{f}_k, \mathbf{V}, \mathbf{g})$ gives

$$e_k(\mathbf{f}_k, \mathbf{V}, \mathbf{g}) = -e^{j\theta^{b_k}_{m,n}} g_{k,n}(\mathbf{f}_k^H \mathbf{H}_{k,b_k})_m - e^{-j\theta^{b_k}_{m,n}} g^*_{k,n}(\mathbf{f}_k^H \mathbf{H}_{k,b_k})^*_m$$

$$+ \sum_{i=1}^{K} [e^{j\theta^{b_i}_{m,n}} g_{i,n}(\mathbf{f}_k^H \mathbf{H}_{k,b_i})_m C^{k,i*}_{m,n} \tag{16}$$

$$+ e^{-j\theta^{b_i}_{m,n}} g^*_{i,n}(\mathbf{f}_k^H \mathbf{H}_{k,b_i})^*_m C^{k,i}_{m,n}] + \text{"terms"} \tag{17}$$

where "terms" denote terms that do not depend on $\theta^{b_k}_{m,n}$. Let us define the following quantities

$$\begin{aligned} \alpha^k_{m,n} &= g_{k,n}(\mathbf{f}_k^H \mathbf{H}_{k,b_k})_m \\ \beta^{k,i}_{m,n} &= g_{i,n}(\mathbf{f}_k^H \mathbf{H}_{k,b_i})_m C^{k,i*}_{m,n} \end{aligned} \tag{18}$$

then we can rewrite (17) as

$$\begin{aligned} e_k(\mathbf{f}_k, \mathbf{V}, \mathbf{g}) &= -e^{j\theta^{b_k}_{m,n}} \alpha^k_{m,n} - e^{-j\theta^{b_k}_{m,n}} \alpha^{k*}_{m,n} + \sum_{i=1}^{K} [e^{j\theta^{b_i}_{m,n}} \beta^{k,i}_{m,n} + e^{-j\theta^{b_i}_{m,n}} \beta^{k,i*}_{m,n}] + \text{"terms"} \\ &= -e^{j\theta^{b_k}_{m,n}} \alpha^k_{m,n} - e^{-j\theta^{b_k}_{m,n}} \alpha^{k*}_{m,n} + e^{j\theta^{b_k}_{m,n}} \beta^k_{m,n} + e^{-j\theta^{b_k}_{m,n}} \beta^{k*}_{m,n} + \text{"terms"} \end{aligned} \tag{19}$$

where $\beta^k_{m,n} = \sum_{i:b_i=b_k} \beta^{k,i}_{m,n}$. Now consider the power constraint term

$$\begin{aligned} ||\mathbf{V}^{b_k} \mathbf{g}_k||^2 &= \sum_{l=1}^{N} \left[\sum_{i=1}^{M} V^{b_k}_{l,i} g_{k,i} \right] \left[\sum_{j=1}^{M} V^{b_k *}_{l,j} g^*_{k,j} \right] \\ &= V^{b_k}_{m,n} g_{k,n} \sum_{i=1, i\neq n}^{M_{b_k}} V^{b_k *}_{m,i} g^*_{k,n} + V^{b_k *}_{m,n} g^*_{k,n} \sum_{j=1, j\neq n}^{M_{b_k}} V^{b_k}_{m,j} g_{k,n} + \text{"terms"}. \end{aligned} \tag{20}$$

Defining $\xi^k_{m,n}$ as

$$\xi^k_{m,n} = \lambda_{b_k} g_{k,n} \sum_{i=1, i\neq n}^{M_{b_k}} V^{b_k *}_{m,i} g^*_{k,n}. \tag{21}$$

Then we can write

$$\lambda_{b_k} ||\mathbf{V}^{b_k} \mathbf{g}_k||^2 = \xi^k_{m,n} e^{j\theta^{b_k}_{m,n}} + \xi^{k*}_{m,n} e^{-j\theta^{b_k}_{m,n}} + \text{"terms"}. \tag{22}$$

Now summing the terms (19) and (22) over all k, we get

$$\begin{aligned} \sum_{k=1}^{K} \left[u_k w_k e_k(\mathbf{f}_k, \mathbf{V}, \mathbf{g}) + \lambda_{b_k} \left| |\mathbf{V}^{b_k} \mathbf{g}_k \right|^2 \right] = \\ \sum_{k:b_k=c} [e^{j\theta^c_{m,n}} (u_k w_k (\beta^k_{m,n} - \alpha^k_{m,n}) + \xi^k_{m,n}) \\ + e^{-j\theta^c_{m,n}} (u_k w_k (\beta^{k*}_{m,n} - \alpha^{k*}_{m,n}) + \xi^{k*}_{m,n})] + \text{"terms"} \end{aligned} \tag{23}$$

where "terms" denote terms that do not depend on $\theta^c_{m,n}$. Defining the terms $a^c_{m,n}$ as

$$a^c_{m,n} = \sum_{k:b_k=c} [u_k w_k (\beta^k_{m,n} - \alpha^k_{m,n}) + \xi^k_{m,n}] \tag{24}$$

we can rewrite the term of interest in the cost function (23) as

$$e^{j\theta^c_{m,n}} a^c_{m,n} + e^{-j\theta^c_{m,n}} a^{c*}_{m,n} = 2\Re\{e^{j\theta^c_{m,n}} a^c_{m,n}\} \tag{25}$$

where $a^c_{m,n} = |a^c_{m,n}| e^{j\angle a^c_{m,n}}$. To minimize (25), $e^{j\theta^c_{m,n}} a^c_{m,n}$ has to made real and negative, hence

$$\theta^c_{m,n} = \pi - \angle a^c_{m,n}. \tag{26}$$

This completes the derivation for θ.

C Derivation of analog phasors using WSR

Adding the phase shifter constraint, we identify the dependence of (7.17) on a single element $\mathbf{V}^c_{p,q}$. We simplify each of the quadratic terms in the expression for WSR. First let us consider each element (r,s) of the matrix $\mathbf{G}_k^H \mathbf{V}^{c\,H} \widehat{\mathbf{B}}_k \mathbf{V}^c \mathbf{G}_k$ (for $k : b_k = c$)

$$\begin{aligned}\mathbf{g}_k^{(r)\,H} \mathbf{V}^{c\,H} \widehat{\mathbf{B}}_k \mathbf{V}^c \mathbf{g}_k^{(s)} &= ((\mathbf{V}^c \mathbf{g}_k^{(r)})_p)^H (\widehat{\mathbf{B}}_k)_{p,p} (\mathbf{V}^c \mathbf{g}_k^{(s)})_p \\ &+ ((\mathbf{V}^c \mathbf{g}_k^{(r)})_{\overline{p}})^H (\widehat{\mathbf{B}}_k)_{\overline{p},p} (V^c \mathbf{g}_k^{(s)})_p + ((\mathbf{V}^c \mathbf{g}_k^{(r)})_p)^H (\widehat{\mathbf{B}}_k)_{p,\overline{p}} (\mathbf{V}^c \mathbf{g}_k^{(s)})_{\overline{p}} \\ &+ ((\mathbf{V}^c \mathbf{g}_k^{(r)})_{\overline{p}})^H (\widehat{\mathbf{B}}_k)_{\overline{p},\overline{p}} (\mathbf{V}^c \mathbf{g}_k^{(s)})_{\overline{p}},\end{aligned} \tag{27}$$

where $(\mathbf{x})_p$ represents the p^{th} element of vector $\mathbf{x}$, $(\mathbf{x})_{\overline{p}}$ represents all other elements, $(\mathbf{B})_{p,p}$ represents element (p,p) of matrix $\mathbf{B}$, $(\mathbf{B})_{\overline{p},p}$ represents all elements in column p except for row p, etc. Note that $\left(\mathbf{V}^c \mathbf{g}_k^{(r)}\right)_{\overline{p}}$ does not contain $\mathbf{V}^c_{p,q}$. The p^{th} term of $\mathbf{V}^c \mathbf{g}_k^{(r)}$ can be written in terms of $\mathbf{V}^c_{p,q}$ as :

$$\left(\mathbf{V}^c \mathbf{g}_k^{(r)}\right)_p = \mathbf{V}^c_{p,q} \mathbf{g}^{(r)}_{k,q} + \mathbf{V}^c_{p,\overline{q}} \mathbf{g}^{(r)}_{k,\overline{q}}, \tag{28}$$

where $\mathbf{V}^c_{p,l}$ represents element (p,l) element of $\mathbf{V}^c$ and $\mathbf{g}^r_{k,q}$ represents the q^{th} element of $\mathbf{g}_k^{(r)}$, $\mathbf{g}^{(r)}_{k,\overline{q}}$ represents all other elements, etc. Now substituting $\mathbf{V}^c_{p,q} = e^{j\theta^c_{p,q}}$, (27) can be written as :

$$\begin{aligned}\mathbf{g}_k^{(r)\,H} \mathbf{V}^{c\,H} \widehat{\mathbf{B}}_k \mathbf{V}^c \mathbf{g}_k^{(s)} &= (\mathbf{V}^{c\,H}_{p,\overline{q}} \mathbf{g}^{(r)}_{k,\overline{q}})^H (\widehat{\mathbf{B}}_k)_{p,p} e^{j\theta^c_{p,q}} \mathbf{g}^{(s)}_{k,q} \\ &+ (\mathbf{V}^c_{p,\overline{q}} \mathbf{g}^{(s)}_{k,\overline{q}}) (\widehat{\mathbf{B}}_k)_{p,p} e^{-j\theta^c_{p,q}} \mathbf{g}^{(r)\,H}_{k,q} + ((\mathbf{V}^c \mathbf{g}_k^{(r)})_{\overline{p}})^H (\widehat{\mathbf{B}}_k)_{\overline{p},p} e^{j\theta^c_{p,q}} \mathbf{g}^{(s)}_{k,q} \\ &+ (\widehat{\mathbf{B}}_k)_{p,\overline{p}} (\mathbf{V}^c \mathbf{g}_k^{(s)})_{\overline{p}} e^{-j\theta^c_{p,q}} \mathbf{g}^{(r)\,H}_{k,q} + \text{“terms".}\end{aligned} \tag{29}$$

Here “terms" denote the terms which are independent of $\mathbf{V}^c_{p,q}$. Define the following matrices $\mathbf{C}_k^{p,q}$ and $\mathbf{D}_k^{p,q}$ whose entries are

$$\begin{aligned}(\mathbf{D}_k^{p,q})_{r,s} &= (\mathbf{V}^{c\,H}_{p,\overline{q}} \mathbf{g}^{(r)}_{k,\overline{q}})^H (\widehat{\mathbf{B}}_k)_{p,p} \mathbf{g}^{(s)}_{k,q} + ((\mathbf{V}^c \mathbf{g}_k^{(r)})_{\overline{p}})^H (\widehat{\mathbf{B}}_k)_{\overline{p},p} \mathbf{g}^{(s)}_{k,q}, \\ (\mathbf{C}_k^{p,q})_{r,s} &= (\mathbf{V}^c_{p,\overline{q}} \mathbf{g}^{(s)}_{k,\overline{q}}) (\widehat{\mathbf{B}}_k)_{p,p} \mathbf{g}^{(r)\,H}_{k,q} + (\widehat{\mathbf{B}}_k)_{p,\overline{p}} (\mathbf{V}^c \mathbf{g}_k^{(s)})_{\overline{p}} \mathbf{g}^{(r)\,H}_{k,q}.\end{aligned} \tag{30}$$

Then we can rewrite $\mathbf{G}_k^H \mathbf{V}^{c\,H} \widehat{\mathbf{B}}_k \mathbf{V}^c \mathbf{G}_k$ as

$$\mathbf{G}_k^H \mathbf{V}^{c\,H} \widehat{\mathbf{B}}_k \mathbf{V}^c \mathbf{G}_k = \mathbf{D}_k^{p,q} e^{j\theta^c_{p,q}} + \mathbf{C}_k^{p,q} e^{-j\theta^c_{p,q}} + \mathbf{T}^{k,1}_{\overline{p},\overline{q}}. \tag{31}$$

Similarly we can write

$$\mathbf{G}_k^H \mathbf{V}^{c\,H} (\widehat{\mathbf{A}}_k + \lambda_c \mathbf{I}) \mathbf{V}^c \mathbf{G}_k = \mathbf{E}_k^{p,q} e^{j\theta^c_{p,q}} + \mathbf{F}_k^{p,q} e^{-j\theta^c_{p,q}} + \mathbf{T}^{k,2}_{\overline{p},\overline{q}}. \tag{32}$$

Here $\mathbf{T}^{k,1}_{\overline{p},\overline{q}}, \mathbf{T}^{k,2}_{\overline{p},\overline{q}}$ are matrices with terms independent of $\theta^c_{p,q}$.

D Gradient Derivation - Part I

In this section we derive an expression for the gradient for the terms of the form tr$\{\mathbf{Y}\}$ w.r.t $\mathbf{X}$, where $\mathbf{Y}$ is

$$\begin{aligned}\mathbf{Y} &= \mathbf{A}\,\text{diag}\,(\mathbf{CXD})\mathbf{B}, \\ \mathbf{R} &= \mathbf{CXD},\end{aligned} \tag{33}$$

where $\mathbf{F}(\mathbf{X})$ represents any matrix function in $\mathbf{X}$. Each diagonal element of $\mathbf{Y}$ can be written as

$$\begin{aligned}\mathbf{Y}_{i,i} &= \sum_{m,n} \mathbf{A}_{i,m}\mathbf{R}_{m,n}\mathbf{B}_{n,i}\delta_{m-n},\\ \mathbf{R}_{m,n} &= \sum_{p,q} \mathbf{C}_{m,p}\mathbf{X}_{p,q}\mathbf{D}_{q,n},\\ \mathbf{Y}_{i,i} &= \sum_{m,n} \mathbf{A}_{i,m}(\sum_{p,q} \mathbf{C}_{m,p}\mathbf{X}_{p,q}\mathbf{D}_{q,n})\mathbf{B}_{n,i}\delta_{m-n},\end{aligned} \tag{34}$$

where δ_k represents the Kronecker delta function. The derivative of $\mathrm{tr}\{\mathbf{Y}\}$ w.r.t $\mathbf{X}_{p,q}$ gives

$$\frac{\partial \mathrm{tr}\{\mathbf{Y}\}}{\partial \mathbf{X}_{p,q}} = \sum_{i,i}\sum_{m,n} \mathbf{A}_{i,m}(\mathbf{C}_{m,p}\mathbf{D}_{q,n})\mathbf{B}_{n,i}\delta_{m-n} = \left[\mathbf{D}\,\mathrm{diag}\,(\mathbf{BA})\mathbf{C}\right]^T. \tag{35}$$

E Gradient Derivation - Part II

In this section we derive an expression for the gradient for the terms of the form

$$\begin{aligned}\mathbf{Y} &= \mathbf{A}\,\mathrm{diag}\,(\mathbf{CXD})\mathbf{B} + \mathbf{F}(\mathbf{X}),\\ \mathbf{R} &= \mathbf{CXD},\end{aligned} \tag{36}$$

where $\mathbf{F}(\mathbf{X})$ represents any matrix function in $\mathbf{X}$. Each element of $\mathbf{Y}$ can be written as

$$\begin{aligned}\mathbf{Y}_{i,j} &= \sum_{m,n} \mathbf{A}_{i,m}\mathbf{R}_{m,n}\mathbf{B}_{n,j}\delta_{m-n} + \mathbf{F}(\mathbf{X})_{i,j},\\ \mathbf{R}_{m,n} &= \sum_{p,q} \mathbf{C}_{m,p}\mathbf{X}_{p,q}\mathbf{D}_{q,n},\\ \mathbf{Y}_{i,j} &= \sum_{m,n} \mathbf{A}_{i,m}(\sum_{p,q} \mathbf{C}_{m,p}\mathbf{X}_{p,q}\mathbf{D}_{q,n})\mathbf{B}_{n,j}\delta_{m-n} + \mathbf{F}(\mathbf{X})_{i,j},\end{aligned} \tag{37}$$

where δ_k represents the Kronecker delta function. We define $\mathbf{V}_{r,s}$ as zero-valued matrix except for a unity element at row r and column s and we obtain

$$\begin{aligned}\frac{\partial \det(\mathbf{Y})}{\partial \mathbf{X}} &= \sum_{r,s} \mathbf{V}_{r,s}\frac{\partial \det(\mathbf{Y})}{\partial \mathbf{X}_{r,s}}\\ &= \sum_{r,s} \mathbf{V}_{r,s}\sum_{i,j}\frac{\partial \det(\mathbf{Y})}{\partial \mathbf{Y}_{i,j}}\frac{\det(\mathbf{Y}_{i,j})}{\partial \mathbf{X}_{r,s}}\\ &= \sum_{r,s} \mathbf{V}_{r,s}\sum_{i,j}\frac{\partial \det(\mathbf{Y})}{\partial \mathbf{Y}_{i,j}}[\sum_{m,n}\mathbf{A}_{i,m}\mathbf{C}_{m,r}\mathbf{D}_{s,n}\mathbf{B}_{n,j}\delta_{m-n} + \frac{\det(\mathbf{F}(\mathbf{X})_{i,j})}{\partial \mathbf{X}_{r,s}}]\\ &= \sum_{r,s} \mathbf{V}_{r,s}(\sum_{m,n}\mathbf{C}_{m,r}\mathbf{D}_{s,n}(\sum_{i,j}\frac{\partial \det(\mathbf{Y})}{\partial \mathbf{Y}_{i,j}}\mathbf{A}_{i,m}\mathbf{B}_{n,j})\delta_{m-n} + \sum_{i,j}\frac{\partial \det(\mathbf{Y})}{\partial \mathbf{Y}_{i,j}}\frac{\det(\mathbf{F}(\mathbf{X})_{i,j})}{\partial \mathbf{X}_{r,s}})\\ &= [\mathbf{D}\,\mathrm{diag}\,(\mathbf{B}(\frac{\partial \det(\mathbf{Y})}{\partial \mathbf{Y}})^T\mathbf{A})\mathbf{C}]^T + \mathbf{F}'\end{aligned} \tag{38}$$

For simplicity we call the second term in the summation $\mathbf{F}'$ since that is not of interest here or the required gradients (needed forms of $\mathbf{F}(\mathbf{X})$) are derived in [59]. Further using the result, $\frac{\partial \det(\mathbf{Y})}{\partial \mathbf{X}} = \det(\mathbf{Y})(\mathbf{Y}^{-1})^T$ we can simplify it as

$$\frac{\partial \det(\mathbf{Y})}{\partial \mathbf{X}} = \det(\mathbf{Y})\left[\mathbf{D}\,\mathrm{diag}\,(\mathbf{BY}^{-1}\mathbf{A})\mathbf{C}\right]^T + \mathbf{F}'. \tag{39}$$

F Derivation of Common and Private Stream Powers using Difference of Convex Functions Programming

Let's consider without loss of any generality the optimization of $\rho_k^{(t)}$ for given values of $\{\rho_i^{(t)} : \forall i \neq k\}$ and $\rho_c^{(t)}$. For simplicity, we drop the iteration index t. We begin by rewriting the SE of UE k as (by explicating its dependence from ρ_k)

$$\begin{aligned} \mathsf{SE}_k(\rho_k) &= \frac{\tau_d}{\tau}\log_2\left(\frac{\mathsf{NUM}_k(\rho_k)}{\mathsf{DEN}_k(\rho_k)}\right) \\ &= \frac{\tau_d}{\tau}\left(\log_2\left(\mathsf{NUM}_k(\rho_k)\right) - \log_2\left(\mathsf{DEN}_k(\rho_k)\right)\right) \end{aligned} \tag{40}$$

where $\mathsf{DEN}_k(\rho_k)$ represents the denominator of γ_k in (12) while $\mathsf{NUM}_k(\rho_k) = \mathsf{DEN}_k(\rho_k) + \rho_k|\mathbb{E}\{\mathbf{h}_k^{\mathrm{H}}\mathbf{w}_k\}|^2$. Observe that $-\log_2\left(\mathsf{DEN}_k(\rho_k)\right)$ is a non-concave function of ρ_k. By linearizing it around a tentative value $\widehat{\rho}_k$, the following approximation is obtained:

$$\log_2\left(\mathsf{DEN}_k(\rho_k)\right) \approx \underbrace{\frac{\mathbb{E}\{|\mathbf{h}_k^{\mathrm{H}}\mathbf{w}_k|^2\} - |\mathbb{E}\{\mathbf{h}_k^{\mathrm{H}}\mathbf{w}_k\}|^2}{\mathsf{DEN}_k(\widehat{\rho}_k)}}_{\triangleq \alpha_k}(\rho_k - \widehat{\rho}_k)$$

where the terms independent of ρ_k have been neglected for simplicity. Similarly, we can rewrite SE_i as

$$\mathsf{SE}_i(\rho_k) = \frac{\tau_d}{\tau}\left(\log_2\left(\mathsf{NUM}_i(\rho_k)\right) - \log_2\left(\mathsf{DEN}_i(\rho_k)\right)\right). \tag{41}$$

By linearizing both terms around $\widehat{\rho}_k$

$$\log_2\left(\mathsf{NUM}_i(\rho_k)\right) \approx \frac{\mathbb{E}\{|\mathbf{h}_i^{\mathrm{H}}\mathbf{w}_k|^2\}}{\mathsf{NUM}_i(\widehat{\rho}_k)}(\rho_k - \widehat{\rho}_k) \tag{42}$$

$$\log_2\left(\mathsf{DEN}_i(\rho_k)\right) \approx \frac{\mathbb{E}\{|\mathbf{h}_i^{\mathrm{H}}\mathbf{w}_k|^2\}}{\mathsf{DEN}_i(\widehat{\rho}_k)}(\rho_k - \widehat{\rho}_k) \tag{43}$$

we obtain the following approximation for $\mathsf{SE}_i(\rho_k)$

$$\mathsf{SE}_i(\rho_k) \approx \frac{\tau_d}{\tau}\underbrace{\left(\frac{\mathbb{E}\{|\mathbf{h}_i^{\mathrm{H}}\mathbf{w}_k|^2\}}{\mathsf{NUM}_i(\widehat{\rho}_k)} - \frac{\mathbb{E}\{|\mathbf{h}_i^{\mathrm{H}}\mathbf{w}_k|^2\}}{\mathsf{DEN}_i(\widehat{\rho}_k)}\right)}_{\triangleq \zeta_i}(\rho_k - \widehat{\rho}_k).$$

Following the same approach for the SE of the common message yields

$$\mathsf{SE}_c(\rho_k) \approx \frac{\tau_d}{\tau}\underbrace{\left(\frac{\mathbb{E}\{|\mathbf{h}_{l_{\min}}^{\mathrm{H}}\mathbf{w}_k|^2\}}{\mathsf{NUM}_{c,\min}(\widehat{\rho}_k)} - \frac{\mathbb{E}\{|\mathbf{h}_{l_{\min}}^{\mathrm{H}}\mathbf{w}_k|^2\}}{\mathsf{DEN}_{c,\min}(\widehat{\rho}_k)}\right)}_{\triangleq \zeta_c}(\rho_k - \widehat{\rho}_k)$$

where $\mathsf{NUM}_{c,\min}(\rho_k) = \mathsf{DEN}_{c,\min}(\rho_k) + \rho_c|\mathbb{E}\{\mathbf{h}_{l_{\min}}^{\mathrm{H}}\mathbf{w}_c\}|^2$ and $\mathsf{DEN}_{c,\min}(\rho_k)$ represents the denominator of $\gamma_{l_{\min},c}$ in (5.14). Putting all the above together, an approximation of the sum SE in (5.18) is

$$\underline{\mathsf{SE}}(\rho_k) = \frac{\tau_d}{\tau}\left(\log_2\left(\mathsf{NUM}_k(\rho_k)\right) - \sigma_k^{(2)}(\rho_k - \widehat{\rho}_k)\right) \tag{44}$$

where

$$\sigma_k^{(2)} = \zeta_c + \alpha_k + \sum_{i=1, i\neq k}^{K} \zeta_i. \tag{45}$$

$$\sigma_k^{(1)} = \frac{\mathbb{E}\{|\mathbf{h}_k^{\mathrm{H}}\mathbf{w}_k|^2\}}{\sigma^2 + \widehat{\rho}_c\left(\mathbb{E}\{|\mathbf{h}_k^{\mathrm{H}}\mathbf{w}_c|^2\} - |\mathbb{E}\{\mathbf{h}_k^{\mathrm{H}}\mathbf{w}_c\}|^2\right) + \sum_{i=1, i\neq k}^{K} \widehat{\rho}_i \mathbb{E}\{|\mathbf{h}_k^{\mathrm{H}}\mathbf{w}_i|^2\}} \tag{46}$$

Taking the derivative of its Lagrangian (obtained after adding the power constraint in (5.20)) and equating it to zero yields

$$\frac{\mathbb{E}\{|\mathbf{h}_k^{\mathrm{H}}\mathbf{w}_k|^2\}}{\mathsf{NUM}_k(\rho_k)} - \sigma_k^{(2)} - \mu = 0 \tag{47}$$

from which one obtain (5.21) in the text, with $\sigma_k^{(1)}$ given in (46). A similar approach for ρ_c yields

$$\frac{\mathbb{E}\{|\mathbf{h}_{l_{\min}}^{\mathrm{H}}\mathbf{w}_c|^2\}}{\mathsf{NUM}_{c,\min}(\rho_c)} - \sigma_c^{(2)} - \mu = 0 \tag{46}$$

where $\sigma_c^{(2)}$ can be obtained as done for $\sigma_k^{(2)}$ in (45); details are omitted for space limitation. Solving (46) yields (5.22) in the text, where $\sigma_c^{(1)}$ is

$$\sigma_c^{(1)} = \frac{\mathbb{E}\{|\mathbf{h}_{l_{\min}}^{\mathrm{H}}\mathbf{w}_c|^2\}}{\sigma^2 + \sum_{i=1}^{K} \widehat{\rho}_i \mathbb{E}\{|\mathbf{h}_{l_{\min}}^{\mathrm{H}}\mathbf{w}_i|^2\}}. \tag{47}$$

G Derivation of Common Stream SINR

Rewrite $\widehat{\mathbf{h}}_k = \Phi_k^{1/2}\mathbf{c}$, with $\mathbf{c} \sim \mathscr{CN}(\mathbf{0}, \mathbf{I})$ and define the deterministic matrix $\mathbf{B}_{ik} = (\Phi_k^{1/2})^{\mathrm{H}}\mathbf{R}_i\mathbf{R}_k^{-1}\Phi_k^{1/2}$. By recalling that $\widehat{\mathbf{h}}_i = \mathbf{R}_i\mathbf{R}_k^{-1}\widehat{\mathbf{h}}_k$ yields

$$\mathbb{E}\{\widehat{\mathbf{h}}_k\widehat{\mathbf{h}}_k^{\mathrm{H}}\widehat{\mathbf{h}}_i\widehat{\mathbf{h}}_i^{\mathrm{H}}\} = \Phi_k^{1/2}\mathbb{E}\{\mathbf{c}\mathbf{c}^{\mathrm{H}}\mathbf{B}_{ik}\mathbf{c}\mathbf{c}^{\mathrm{H}}\}(\Phi_k^{1/2})^{\mathrm{H}}. \tag{48}$$

It then follows that

$$\mathbb{E}\left\{[\mathbf{c}\mathbf{c}^{\mathrm{H}}\mathbf{B}_{ik}\mathbf{c}\mathbf{c}^{\mathrm{H}}]_{mn}\right\} = \sum_{j=1}^{M}\sum_{l=1}^{M}[\mathbf{B}_{ik}]_{lj}\mathbb{E}\left\{c_m c_l^* c_j c_n^*\right\}. \tag{49}$$

If $m = n$, then (49) is always zero except for $l = j$:

$$\sum_{l=1}^{M}[\mathbf{B}_{ik}]_{ll}\mathbb{E}\left\{|c_m|^2|c_l|^2\right\} = 3[\mathbf{B}_{ik}]_{mm} + \sum_{l=1, l\neq m}^{M}[\mathbf{B}_{ik}]_{ll} \tag{50}$$

where we have taken into that $\mathbb{E}\left\{|c_m|^4\right\} = 3$. If $m \neq n$, then (49) is always zero except for $l = m$ and $j = n$

$$[\mathbf{B}_{ik}]_{mn}\mathbb{E}\left\{c_m c_m^* c_m c_n^*\right\} = [\mathbf{B}_{ik}]_{mn}. \tag{51}$$

Putting the above results together yields $\mathbb{E}\{\widehat{\mathbf{h}}_k\widehat{\mathbf{h}}_k^{\mathrm{H}}\widehat{\mathbf{h}}_i\widehat{\mathbf{h}}_i^{\mathrm{H}}\} = \mathrm{tr}\{\mathbf{B}_{ik}\}\mathbf{I} + \mathrm{diag}\,(\mathbf{B}_{ik}) + \mathbf{B}_{ik}$.

H Deterministic Equivalent of Auxiliary Quantities

I Proof of Theorem 12

First we compute the deterministic equivalent for $\sigma_k^{(1)}$

$$\begin{aligned}\sigma_k^{(1)} &= \widehat{r}_{\overline{k}}^{-1}\mathbf{g}_k'^H\widehat{\mathbf{S}}_{k,b_k}\mathbf{g}_k' \\ &= \widehat{r}_{\overline{k}}^{-1}\mathbf{g}_k'^H\mathbf{C}_{k,b_k}\mathbf{W}_{k,b_k}\mathbf{C}_{k,b_k}^H\mathbf{g}_k'.\end{aligned} \tag{52}$$

Using the eigen decomposition of $\mathbf{W}_{k,b_k} = \mathbf{V}_{k,b_k}\mathbf{\Lambda}_{k,b_k}\mathbf{V}_{k,b_k}^H$

$$\begin{aligned}\mathbf{g}_k'^H\mathbf{S}_{k,b_k}\mathbf{g}_k' &= \mathbf{g}_k'^H\mathbf{C}_{k,b_k}\mathbf{W}_{k,b_k}\mathbf{C}_{k,b_k}^H\mathbf{g}_k', \\ \mathbf{g}_k' &= \mathbf{g}_k''/\left\|\mathbf{g}_k''\right\|, \\ \mathbf{g}_k'' &= \mathbf{\Gamma}_k^{-1}\mathbf{C}_{k,b_k}\mathbf{v}_{k,b_k}, \\ \mathbf{C}_{k,b_k}^H\mathbf{g}_k' &= \mathbf{C}_{k,b_k}^H\mathbf{\Gamma}_k^{-1}\mathbf{C}_{k,b_k}\mathbf{v}_{k,b_k}/||\mathbf{g}_k''||,\end{aligned} \tag{53}$$

where $\mathbf{\Gamma}_k = \sum_{i\neq k}\beta_i\mathbf{S}_{i,b_k}+\mu_{b_k}\mathbf{I}$. Using large system analysis simplifications shown in (81), $\mathbf{C}_{k,b_k}^H\mathbf{\Gamma}_k^{-1}\mathbf{C}_{k,b_k} = e_{b_k}\mathbf{I}$

$$\begin{aligned}\mathbf{g}_k'^H\mathbf{S}_{k,b_k}\mathbf{g}_k' &= \frac{e_{b_k}^2\mathbf{v}_{k,b_k}^H\mathbf{W}_{k,b_k}\mathbf{v}_{k,b_k}}{\left\|\mathbf{g}_k''\right\|^2} \\ &= \frac{e_{b_k}^2\lambda_{max}(\mathbf{W}_{k,b_k})}{\left\|\mathbf{g}_k''\right\|^2},\end{aligned} \tag{54}$$

We define $\mathbf{\Gamma}_{b_k} = \mathbf{\Gamma}_k + \beta_k\mathbf{S}_{k,b_k}$. Further we consider simplifying $\left\|\mathbf{g}_k''\right\|^2 = \mathbf{v}_{k,b_k}^H\mathbf{C}_{k,b_k}\mathbf{\Gamma}_k^{-2}\mathbf{C}_{k,b_k}\mathbf{v}_{k,b_k}$. By using Lemma 4 from [14] leads to $\left\|\mathbf{g}_k''\right\|^2 = \frac{1}{M_{b_k}}\mathrm{tr}\{\mathbf{\Gamma}_k^{-2}\}\left\|\mathbf{v}_{k,b_k}\right\|^2 = \frac{1}{M_{b_k}}\mathrm{tr}\{\mathbf{\Gamma}_k^{-2}\}$. Further, by using Lemma 6 we approximate $\mathbf{\Gamma}_k^{-1} \approx (\mathbf{\Gamma}_k + \beta_k\mathbf{S}_{k,b_k})^{-1} = \mathbf{\Gamma}_{b_k}^{-1}$. From [14], in the large system limit, for $(1/M_{b_k})\mathrm{tr}\{\mathbf{\Gamma}_{b_k}^{-2}\}$, we have an almost sure convergence value as e_{b_k}', where e_{b_k}' is the derivative of e_{b_k} w.r.t. μ_{b_k}, and thus $\left\|\mathbf{g}_k''\right\|^2 = e_{b_k}'$

$$\begin{aligned}e_{b_k}' &= e_{b_k}^2\Big(\frac{1}{M_{b_k}}\sum_{i=1}^{K}\sum_{r=1}^{L_{i,b_k}}\frac{\beta_i^2\zeta_{i,b_k}^{(r),2}e_{b_k}'}{(1+\beta_i\zeta_{i,b_k}^{(r)}e_{b_k})^2}+1\Big) \\ \implies e_{b_k}' &= \frac{e_{b_k}^2}{1-\frac{e_{b_k}^2}{M_{b_k}}\sum_{i=1}^{K}\sum_{r=1}^{L_{i,b_k}}\frac{\beta_i^2\zeta_{i,b_k}^{(r),2}}{(1+\beta_i\zeta_{i,b_k}^{(r)}e_{b_k})^2}}.\end{aligned} \tag{55}$$

Deterministic limit for $r_{\overline{k}}, r_k$: Each term in $r_{\overline{k}}$ is of the form $p_i\mathbf{g}_i''\mathbf{S}_{k,b_i}\mathbf{g}_i''/\left\|\mathbf{g}_i''\right\|^2$, where $\mathbf{g}_i''^H\mathbf{S}_{k,b_i}\mathbf{g}_i'' = \mathbf{v}_{i,b_i}'^H\mathbf{\Gamma}_i^{-1}\mathbf{C}_{k,b_i}\mathbf{W}_{k,b_i}\mathbf{C}_{k,b_i}^H\mathbf{\Gamma}_i^{-1}\mathbf{v}_{i,b_i}'$ and we defined $\mathbf{v}_{i,b_i}' = \mathbf{C}_{i,b_i}\mathbf{v}_{i,b_i}$. Since $\mathbf{v}_{i,b_i}'$ is independent of all other random quantities in this expression, we apply Lemma 4 and then Lemma 6 to get, $\mathbf{v}_{i,b_i}'^H\mathbf{\Gamma}_i^{-1}\mathbf{C}_{k,b_i}\mathbf{W}_{k,b_i}\mathbf{C}_{k,b_i}^H\mathbf{\Gamma}_i^{-1}\mathbf{v}_{i,b_i}' = \frac{1}{M_{b_i}}\mathrm{tr}\{\mathbf{\Gamma}_{b_i}^{-1}\mathbf{C}_{k,b_i}\mathbf{W}_{k,b_i}\mathbf{C}_{k,b_i}^H\mathbf{\Gamma}_{b_i}^{-1}\}$. Applying Lemma 1 to each of the rows of $\mathbf{V}_{k,b_i}^H\mathbf{C}_{k,b_i}^H\mathbf{\Gamma}_i^{-1}$, then Lemma 4 and 6, we obtain the following simplified expression

$$\begin{aligned}\frac{1}{M_{b_i}}\mathrm{tr}\{\mathbf{\Gamma}_{b_i}^{-1}\mathbf{S}_{k,b_i}\mathbf{\Gamma}_{b_i}^{-1}\} &= \frac{1}{M_{b_i}^2}\mathrm{tr}\{\mathbf{\Gamma}_{b_i}^{-2}\}\mathrm{tr}\{\mathbf{\Lambda}_{k,b_i}\mathbf{B}_{k,b_i}^{-2}\}, \\ \text{where, } \mathbf{B}_{k,b_i} &= \mathrm{diag}\,(1+\beta_k\zeta_{k,b_i}^{(1)}e_{b_i},...,1+\beta_k\zeta_{k,b_i}^{(L)}e_{b_i}).\end{aligned} \tag{56}$$

Finally we obtain,

$$\begin{aligned}\sum_{\substack{i=1,\\ i\neq k,}}^{K} p_i \mathbf{g}_i'^H \mathbf{S}_{k,b_k} \mathbf{g}_i' &= \sum_{\substack{i=1,\\ i\neq k}}^{K} p_i \frac{1}{M_{b_i}} [\sum_{r=1}^{L_{k,b_i}} \frac{\zeta_{k,b_i}^{(r)}}{(1+\beta_k \zeta_{k,b_i}^{(r)} e_{b_i})^2}] \\ &= \Upsilon_{\overline{k}},\end{aligned} \tag{57}$$

Thus we can write $\overline{r}_k$ and $\overline{r}_{\overline{k}}$

$$\begin{aligned}\overline{r}_{\overline{k}} &= 1+\Upsilon_{\overline{k}}, \\ \overline{r}_k &= 1+\Upsilon_{\overline{k}} + p_k \frac{e_{b_k}^2 \lambda_{max}(\mathbf{W}_{k,b_k})}{e'_{b_k}}, \\ \text{Also, } \overline{\beta}_k &= u_k \left(\frac{1}{\overline{r}_{\overline{k}}} - \frac{1}{\overline{r}_k}\right), \\ \overline{\alpha}_k &= \frac{u_k}{\overline{r}_k}.\end{aligned} \tag{58}$$

Finally, combining (54), (55), (58), we can write the deterministic equivalent for $\sigma_k^{(1)}$ as, $\overline{\sigma}_k^{(1)} = \frac{e_{b_k}^2 \lambda_{max}(\mathbf{W}_{k,b_k})}{e'_{b_k}(1+\Upsilon_{\overline{k}})}$. Each term in $\sigma_k^{(2)}$ is of the form $\widehat{\beta}_i \mathbf{g}_k''^H \mathbf{S}_{i,b_k} \mathbf{g}_k''^H / \left\|\mathbf{g}_k''\right\|^2$, which gets simplified as follows:

$$\begin{aligned}\mathbf{g}_k''^H \mathbf{S}_{i,b_k} \mathbf{g}_k''^H &= \mathbf{v}_{k,b_k}'^H \mathbf{\Gamma}_k^{-1} \mathbf{C}_{i,b_k} \mathbf{W}_{i,b_k} \mathbf{C}_{i,b_k}^H \mathbf{\Gamma}_k^{-1} \mathbf{v}_{k,b_k}' \\ &\overset{(a)}{=} \frac{1}{M_{b_k}} \mathrm{tr}\{\mathbf{\Gamma}_k^{-1} \mathbf{C}_{i,b_k} \mathbf{W}_{i,b_k} \mathbf{C}_{i,b_k}^H \mathbf{\Gamma}_k^{-1}\},\end{aligned} \tag{59}$$

where (a) follows from Lemma 4 (since $\mathbf{v}_{k,b_k}'$ is independent of all other matrices involved). By following the same steps as in (56)-(57), this gets simplified and we write $\overline{\sigma}_k^{(2)}$ as

$$\begin{aligned}\overline{\sigma}_k^{(2)} &= \sum_{i=1, i\neq k}^{K} \overline{\beta}_i \mathbf{g}_k''^H \mathbf{S}_{i,b_k} \mathbf{g}_k''^H / \left\|\mathbf{g}_k''\right\|^2 \\ &= \frac{1}{M_{b_k}} \sum_{i=1, i\neq k}^{K} \overline{\beta}_i [\sum_{r=1}^{L_{i,b_k}} \frac{\zeta_{i,b_k}^{(r)}}{(1+\overline{\beta}_i \zeta_{i,b_k}^{(r)} e_{b_k})^2}].\end{aligned} \tag{60}$$

I Sum Rate Evaluation (At any SNR)

In this section, let $\gamma_{k,L}^{(s)}, \gamma_{k,S}^{(s)}$ denotes the SINRs for user k in the case of LMMSE and subspace channel estimators respectively. The superscript s can be 'Opt' or 'N' or 'E' which represents the ESIP-WSR/naive/EWSMSE BFs respectively. By the large system limit, we implies that $L, M, K \to \infty$ at a finite ratio $\alpha_c = \frac{\sum_{i=1}^{K} L_{i,c}}{M}$.

I Analytical Solution for μ_c, β_k, e_c

First we consider the value of the Lagrange multiplier at high SNR. From ILA-WF (7.18), at high SNR, if BF does ZF, then $\sigma_k^{(2)}$, which is the leakage power part converges to zero. Hence, $p_k =$

$\frac{1}{mu} - \frac{1}{\sigma_k^{(1)}}$. $\sigma_k^{(1)} \approx \text{tr}\{\mathbf{D}_k\}$ and it will be a finite constant that becomes negligible at high SNR. Hence, $p_k \rightarrow \frac{u_k}{\mu_c}$. $\sum_{k:b_k=c} p_k = P_c = 1/\mu_c \sum_{k:b_k=c} u_k$. So, $\mu_c = \frac{\sum_{k:b_k=c} u_k}{P_c}$. If $u_k = 1$, then $\mu_c = K_c/P_c$ at high SNR. However, if K_c is finite, since the total power $\rightarrow \infty$, $\frac{1}{\mu} \rightarrow \infty$. Hence, $\lim_{P\rightarrow\infty} \mu = 0$. Next, we consider the deterministic equivalent for β_k, which is defined in (7.39). At high SNR, if ZF happens, then $\overline{r}_{\overline{k}} = 0$, hence $\beta_k = u_k(1 - \frac{1}{\overline{r}_k})$. At high SNR, $\overline{r}_k \rightarrow \infty$, hence $\beta_k = u_k(1-0) = u_k = \overline{\beta}_k$. If $u_k = 1$, $\overline{\beta}_k = 1$.

Further we derive an approximate analytical solution for the implicit equation of e_c in (7.46). Writing $e_c^{-1} = f(e_c)$, it is obvious that $f(e_c)$ is a monotonically decreasing function for any SNR and hence e_c is a monotonically increasing function. So we first evaluate the analytical solution of e_c at the extreme points. At very low SNR, all the interference terms can be neglected, so from (7.31), $\mathbf{\Gamma}_k \approx \mu_c \mathbf{I}, b_k = c$. Hence $e_c \mathbf{I} = \mathbf{C}_k^H \mathbf{\Gamma}_k^{-1} \mathbf{C}_k \xrightarrow[a.s]{M\rightarrow\infty} \mu_c^{-1} \mathbf{C}_k^H \mathbf{C}_k = \frac{1}{\mu_c}\mathbf{I} = e_c^0 \mathbf{I}$. For high SNR, we do a first order perturbation analysis in $\widetilde{\sigma}^2$, thus from (7.45), $\lambda_2 \xrightarrow[a.s]{M\rightarrow\infty} \widetilde{\sigma}^2$. So we can approximate the term $\frac{\beta\lambda_2}{1+\beta\lambda_2 e_c} \approx \beta\widetilde{\sigma}^2(1 - \beta\widetilde{\sigma}^2 e_c) \approx \beta\widetilde{\sigma}^2$. Hence, we obtain

$$
\begin{aligned}
\frac{1}{e_c} &= \frac{\alpha}{L}\frac{\beta\lambda_1}{1+\beta\lambda_1 e_c} + \mu_c', \ \alpha\beta\widetilde{\sigma}^2 + \mu_c \\
&= \mu_c', \ \text{solving this leads to,} \\
e_c^{\infty} &= \frac{-(\mu_c' - \lambda_1') + \sqrt{(\mu_c' - \lambda_1')^2 + 4\beta\lambda_1\mu_c'}}{2\beta\lambda_1\mu_c'}, \ \lambda_1' \\
&= \beta\lambda_1(1 - \frac{\alpha}{L}).
\end{aligned} \tag{61}
$$

Further we can deduce from e_c^0 and (61) that at extreme SNR regimes, $\lim_{P\rightarrow 0} e_c^0 = 0$ and $\lim_{P\rightarrow\infty} e_c^{\infty} = \infty$.

II ESIP-WSR BF with LMMSE Channel Estimator

For the convenience of analysis, we write the BF expression (7.34), $\mathbf{g}_k = \mathbf{\Gamma}_k^{-1}\mathbf{C}_k\mathbf{v}_{k,b_k}$. Here $\mathbf{\Gamma}_k = \sum_{i\neq k} \beta_i \mathbf{S}_{i,b_k} + \mu_{b_k}\mathbf{I}$. For the asymptotic analysis below, we first determine the value of $(1/M_{b_k})\text{tr}\{\mathbf{\Gamma}_{b_k}^{-2}\}$ in the large system limit. From [14], in the large system limit, for $(1/M_c)\text{tr}\{\mathbf{\Gamma}_c^{-2}\}$, we have an almost sure convergence value as e_c', where e_c' is the derivative of e_c w.r.t. μ_c

$$
\begin{aligned}
e_c' &= e_c^2\left(\frac{1}{M_c}\sum_{i=1}^{K}\sum_{r=1}^{L_{i,c}} \frac{\beta_i^2 \lambda_{i,c}^{(r),2} e_c'}{(1+\beta_i\lambda_{i,c}^{(r)}e_c)^2} + 1\right) \implies \\
e_c' &= \frac{e_c^2}{1 - \frac{e_c^2}{M_c}\sum_{i=1}^{K}\sum_{r=1}^{L_{i,c}} \frac{\beta_i^2\lambda_{i,c}^{(r),2}}{(1+\beta_i\lambda_{i,c}^{(r)}e_c)^2}}, \\
x_c &= \frac{e_c^2}{M_c}\sum_{i=1}^{K}\sum_{r=1}^{L_{i,c}} \frac{\beta_i^2\lambda_{i,c}^{(r),2}}{(1+\beta_i\lambda_{i,c}^{(r)}e_c)^2}, \\
e_c' &= \frac{e_c^2}{1-x_c}.
\end{aligned} \tag{62}
$$

For the ESIP-WSR BF with LMMSE channel estimate, the computation of max eigenvector, $\mathbf{v}_{k,b_k}$ in (7.34) is not analytically feasible. Hence we consider the simplification that the first element

in $\mathbf{D}_k$ dominates the rest of the elements and we write $\mathbf{D}_k = \text{tr}\{\mathbf{D}_k\}\mathbf{e}_1\mathbf{e}_1^H$. Substituting this in $\mathbf{W}_k$ leads to $\mathbf{v}_{k,b_k} = \mathbf{e}_1$. First we look at the deterministic equivalent for the signal power P_{S_k}

$$\begin{aligned} |\mathbf{g}_k''^H\mathbf{h}_k|^2 &= \mathbf{e}_1^H\mathbf{C}_k^H\boldsymbol{\Gamma}_k^{-1}\mathbf{C}_k\mathbf{d}_k\mathbf{d}_k^H\mathbf{C}_k^H\boldsymbol{\Gamma}_k^{-1}\mathbf{C}_k\mathbf{e}_1 \\ &\xrightarrow[a.s]{M\to\infty} e_{b_k}^2\mathbf{e}_1^H\mathbf{d}_k\mathbf{d}_k^H\mathbf{e}_1 e_{b_k}^2\,\text{E}(\mathbf{e}_1^H\mathbf{d}_k\mathbf{d}_k^H\mathbf{e}_1) \\ &= e_{b_k}^2\text{tr}\{\mathbf{D}_k\}, \\ \|\mathbf{g}_k''\|^2 &= \mathbf{e}_1^H\mathbf{C}_k^H\boldsymbol{\Gamma}_k^{-2}\mathbf{C}_k\mathbf{e}_1 \xrightarrow[a.s]{M\to\infty} e'_{b_k}. \end{aligned} \tag{63}$$

Further we look at the interference power.

$$\begin{aligned} |\mathbf{g}_i'^H\mathbf{h}_{k,b_i}|^2 &= \mathbf{e}_1^H\mathbf{C}_i^H\boldsymbol{\Gamma}_i^{-1}\mathbf{C}_{k,b_i}\mathbf{d}_{k,b_i}\mathbf{d}_{k,b_i}^H\mathbf{C}_i^H\boldsymbol{\Gamma}_i^{-1}\mathbf{C}_i\mathbf{e}_1 \xrightarrow[a.s]{(a)\ M\to\infty} \\ \frac{1}{M_{b_i}}\text{tr}\{\boldsymbol{\Gamma}_i^{-1}\mathbf{C}_{k,b_i}\mathbf{d}_{k,b_i}\mathbf{d}_{k,b_i}^H\mathbf{C}_{k,b_i}^H\boldsymbol{\Gamma}_i^{-1}\} &\overset{(b)}{=} \frac{1}{M_{b_i}}\text{tr}\{\boldsymbol{\Gamma}_{ik}^{-1}\mathbf{C}_{k,b_i}\mathbf{B}_{k,b_i}^{-1}\mathbf{d}_{k,b_i}\mathbf{d}_{k,b_i}^H\mathbf{B}_{k,b_i}^{-1}\mathbf{C}_{k,b_i}^H\boldsymbol{\Gamma}_{ik}^{-1}\} \\ &\xrightarrow[a.s]{M\to\infty} \frac{1}{M_{b_i}}\text{tr}\{\boldsymbol{\Sigma}_{ik}^{-2}\}\frac{1}{M}\text{tr}\{\mathbf{B}_{k,b_i}^{-2}\mathbf{D}_{k,b_i}\}, \\ \text{where, } \boldsymbol{\Sigma}_{ik} &= \boldsymbol{\Gamma}_{ik} - \beta_k\mathbf{C}_{k,b_i}\widetilde{\mathbf{D}}_{k,b_i}\mathbf{C}_{k,b_i}^H, \\ \boldsymbol{\Gamma}_{ik} &= \boldsymbol{\Gamma}_i - \beta_k\widehat{\mathbf{h}}_{k,b_i}\widehat{\mathbf{h}}_{k,b_i}^H, \end{aligned} \tag{64}$$

where we define $\mathbf{B}_{k,b_i} = \text{diag}(1+\beta_k e_{b_i}\lambda_{k,b_i}^{(1)}),....,1+\beta_k e_{b_i}\lambda_{k,b_i}^{(L_{k,b_i})})$. In order to simplify it, for the transition (a), we apply Lemma 3 since $\mathbf{C}_i$ is independent of $\boldsymbol{\Gamma}_i, \mathbf{C}_{k,b_i}, \mathbf{d}_{k,b_i}$. In (b) above, we approximate $\text{tr}\{\boldsymbol{\Gamma}_i^{-1}\mathbf{C}_{k,b_i}\mathbf{B}_{k,b_i}^{-1}\mathbf{d}_{k,b_i}\mathbf{d}_{k,b_i}^H\mathbf{B}_{k,b_i}^{-1}\mathbf{C}_{k,b_i}^H\boldsymbol{\Gamma}_i^{-1}\} = \text{tr}\{\boldsymbol{\Gamma}_{ik}^{-1}\mathbf{B}_{k,b_i}^{-1}\mathbf{C}_{k,b_i}\mathbf{d}_{k,b_i}\mathbf{d}_{k,b_i}^H\mathbf{B}_{k,b_i}^{-1}\mathbf{C}_{k,b_i}^H\boldsymbol{\Gamma}_{ik}^{-1}\}$ by applying Lemma 4. Further we applied matrix inversion lemma to convert each row of the matrix $\mathbf{C}_{k,b_i}^H\boldsymbol{\Gamma}_{ik}^{-1}$ to $\mathbf{C}_{k,b_i}^{(r)H}\boldsymbol{\Gamma}_{ik}^{-1} = \frac{\mathbf{C}_{k,b_i}^{(r)H}\boldsymbol{\Sigma}_{ik}^{-1}}{1+\beta_k\lambda_{k,b_i}^{(r)}e_i}$, where we denote $\mathbf{C}_{k,b_i}^{(r)}$ as the r^{th} column of $\mathbf{C}_{k,b_i}$ and $\lambda_{k,b_i}^{(r)}$ being the r^{th} diagonal element of $\widetilde{\mathbf{D}}_{k,b_i}$. In the above equation (64), we define $\mathbf{B}_{k,b_i} = \text{diag}(1+\beta_k e_{b_i}\lambda_{k,b_i}^{(1)},..,1+\beta_k e_{b_i}\lambda_{k,b_i}^{(r)})$. Finally, we obtain the SINR as

$$\gamma_{k,L}^{(Opt)} = \frac{p_k(1-x_{b_k}^{(L,Opt)})\text{tr}\{\mathbf{D}_k\}}{\frac{1}{M_{b_i}}\sum\limits_{i\neq k}p_i\text{tr}\{\mathbf{D}_{k,b_i}\mathbf{B}_{k,b_i}^{-2}\}+1}. \tag{65}$$

In the case of subspace channel estimator, where $\mathbf{U}_S = \mathbf{I}$, from (7.37), $\mathbf{V}_{max}(\mathbf{W}_S) = \widehat{\mathbf{d}}_k$ and hence the derivations provided in (II) is valid for any SNR and hence the SINR expression is the same as (102).

III Naive EWSR BF with LMMSE/Subspace Channel Estimators

For the naive BF with LMMSE/Subspace channel estimate, $\mathbf{W}_k = \mathbf{U}_k\widehat{\mathbf{d}}_k\widehat{\mathbf{d}}_k^H\mathbf{U}_k$we can write $\mathbf{V}_{max}(\mathbf{W}_k) \propto \mathbf{U}_k\widehat{\mathbf{d}}_k$. Hence $\mathbf{g}_k'' = \boldsymbol{\Gamma}_k^{-1}\mathbf{C}_k\mathbf{U}_k\widehat{\mathbf{d}}_k$. First considering the signal power part

$$\begin{aligned} \mathbf{g}_k''^H\mathbf{h}_k &= \widehat{\mathbf{d}}_k^H\mathbf{U}_k\mathbf{C}_k^H\boldsymbol{\Gamma}_k^{-1}\mathbf{C}_k\mathbf{d}_k \xrightarrow[a.s]{M\to\infty} e_{b_k}\text{tr}\{\mathbf{U}_k\mathbf{D}_k\}, \\ |\mathbf{g}_k''|^2 &= \widehat{\mathbf{d}}_k^H\mathbf{U}_k\mathbf{C}_k^H\boldsymbol{\Gamma}_k^{-2}\mathbf{C}_k\mathbf{U}_k\widehat{\mathbf{d}}_k \xrightarrow[a.s]{M\to\infty} e'_{b_k}\text{tr}\{\mathbf{U}_k^2(\mathbf{D}_k+\widetilde{\sigma}^2\mathbf{I})\}. \end{aligned} \tag{66}$$

Next we look at the interference power part. Since $\mathbf{C}_i$ is independent of $\mathbf{\Gamma}_i, \mathbf{C}_{k,b_i}, \mathbf{d}_{k,b_i}$, we can apply Lemma 3 to obtain

$$
\begin{aligned}
|\mathbf{g}_i''^H \mathbf{h}_{k,b_i}|^2 &= \widehat{\mathbf{d}}_i^H \mathbf{U}_i \mathbf{C}_i^H \mathbf{\Gamma}_i^{-1} \mathbf{C}_{k,b_i} \mathbf{d}_{k,b_i} \mathbf{d}_{k,b_i}^H \mathbf{C}_{k,b_i}^H \mathbf{\Gamma}_i^{-1} \mathbf{C}_i \mathbf{U}_i \widehat{\mathbf{d}}_i \\
&\xrightarrow[a.s]{M\to\infty} \widehat{\mathbf{d}}_i^H \mathbf{U}_i^H \mathbf{U}_i \widehat{\mathbf{d}}_i \frac{1}{M} \mathrm{tr}\{\mathbf{\Gamma}_i^{-2} \mathbf{C}_{k,b_i} \mathbf{d}_{k,b_i} \mathbf{d}_{k,b_i}^H \mathbf{C}_{k,b_i}^H\}.
\end{aligned} \tag{67}
$$

We define $\mathbf{\Xi}_i = \mathbf{\Gamma}_i - \beta_k \mathbf{C}_{k,b_i} \mathbf{U}_{k,b_i} \widehat{\mathbf{d}}_{k,b_i} \widehat{\mathbf{d}}_{k,b_i}^H \mathbf{U}_{k,b_i} \mathbf{C}_{k,b_i}^H$. Further we apply the Lemma 4 twice and then Lemma 3 to convert $\mathbf{C}_{k,b_i}^H \mathbf{\Xi}_i^{-2} \mathbf{C}_{k,b_i} \xrightarrow[a.s]{M\to\infty} \frac{1}{M_{b_i}} \mathrm{tr}\{\mathbf{\Xi}_i^{-2}\}$ which converges to e'_{b_i}

$$
\begin{aligned}
\frac{1}{M} \mathrm{tr}\{\mathbf{\Gamma}_i^{-2} \mathbf{C}_{k,b_i} \mathbf{d}_{k,b_i} \mathbf{d}_{k,b_i}^H \mathbf{C}_{k,b_i}^H\} &\xrightarrow[a.s]{M\to\infty} \frac{1}{M} \mathrm{tr}\{\mathbf{\Xi}_i^{-2} \mathbf{C}_{k,b_i} \mathbf{d}_{k,b_i} \mathbf{d}_{k,b_i}^H \mathbf{C}_{k,b_i}^H\} \\
&\xrightarrow[a.s]{M\to\infty} e'_{b_i} \frac{1}{M} \mathrm{tr}\{\mathbf{D}_{k,b_i}\}.
\end{aligned} \tag{68}
$$

Thus combining (67), (68), we obtain, $|\mathbf{g}_i''^H \mathbf{h}_{k,b_i}|^2 \xrightarrow[a.s]{M\to\infty} \frac{1}{M} \mathrm{tr}\{\mathbf{D}_{k,b_i}\}$. Finally, the SINR can be written as for LMMSE/Subspace channel estimators

$$
\begin{aligned}
\gamma_{k,L}^{(N)} - \frac{(1 - x_{b_k}^{(L,N)}) \, \mathrm{tr}\{\mathbf{D}_k^2 (\mathbf{D}_k + \widetilde{\sigma}_k^2 \mathbf{I})^{-1}\} p_k}{(\sum_{i=1}^{K} \frac{1}{M} p_i \mathrm{tr}\{\mathbf{D}_{k,b_i}\} + 1)} &\xrightarrow[a.s]{M\to\infty} 0, \\
\gamma_{k,S}^{(N)} - \frac{(1 - x_{b_k}^{(S,N)}) \, (\mathrm{tr}\{\mathbf{D}_k\})^2 p_k}{\mathrm{tr}\{(\mathbf{D}_k + \widetilde{\sigma}^2 \mathbf{I})\} (\sum_{i=1}^{K} \frac{1}{M} p_i \mathrm{tr}\{\mathbf{D}_{k,b_i}\} + 1)} &\xrightarrow[a.s]{M\to\infty} 0.
\end{aligned} \tag{69}
$$

Note that $x_{b_k}^{(S,N)} > x_{b_k}^{(L,N)}$ since the eigenvalues of $\mathbf{W}_k$ for the subspace estimator is greater than that of the LMMSE channel estimator. This leads to a reduction in signal power for the subspace channel estimator case at low to mid SNR range (while the interference power remains approximately the same for both) which explains the sub-optimal performance of subspace channel estimators.

IV EWSMSE BF with LMMSE/Subspace Channel Estimators

We denote superscript (E) to denote the SINRs for the case of EWSMSE BFs (for example, $\gamma_{k,L}^{(E)}$). Also, $x_{b_k}^{(E_L)}, x_{b_K}^{(E_S)}$ represents x_{b_k} for the respective channel estimates. In this section, we consider the deterministic equivalent for the SINR expression of the EWSMSE BF with LMMSE/subspace channel estimators. So from (7.41), we can rewrite the BF expression as, $\mathbf{g}_k'' = \mathbf{F}_k^{-1} \mathbf{C}_k \mathbf{U}_k \widehat{\mathbf{d}}_k$. Further we compute the signal power part

$$
\begin{aligned}
|\mathbf{g}_k''^H \mathbf{h}_k| &\overset{(a)}{=} \widehat{\mathbf{d}}_k^H \mathbf{U}_k \mathbf{C}_k^H \left(\mathbf{\Gamma}_k^{-1} - \mathbf{\Gamma}_k^{-1} \mathbf{C}_k \left(\beta_k^{-1} \widetilde{\mathbf{D}}_k^{-1} + e_{b_k} \mathbf{I} \right)^{-1} \mathbf{C}_k^H \mathbf{\Gamma}_k^{-1} \right) \mathbf{C}_k \mathbf{d}_k, \\
\widehat{\mathbf{d}}_k^H \mathbf{U}_k e_{b_k} \left(\mathbf{I} - \mathbf{E}_k^{-1} e_{b_k} \right) \mathbf{U}_k \widehat{\mathbf{d}}_k &\xrightarrow[a.s]{M\to\infty} e_{b_k} \mathrm{tr}\{\mathbf{U}_k \left(\mathbf{I} - \mathbf{E}_k^{-1} e_{b_k} \right) \mathbf{D}_k\}, \\
&\text{where } \mathbf{E}_k = \beta_k^{-1} \widetilde{\mathbf{D}}_k^{-1} + e_{b_k} \mathbf{I}.
\end{aligned} \tag{70}
$$

In (a) above, we applied matrix inversion lemma to the term $\mathbf{F}_k^{-1}$ and also simplified $\mathbf{C}_k^H \mathbf{\Gamma}_k^{-1} \mathbf{C}_k = e_k \mathbf{I}$ by first applying Lemma 3 and then Theorem 8. Further, we look at the normalization factor

for the BF $\mathbf{g}_k$

(71)

$$\begin{aligned}
\left\|\mathbf{g}_k''\right\|^2 = & \\
& \widehat{\mathbf{d}}_k^H \mathbf{U}_k \mathbf{C}_k^H \left(\mathbf{\Gamma}_k^{-1} - \mathbf{\Gamma}_k^{-1}\mathbf{C}_k\left(\beta_k^{-1}\widetilde{\mathbf{D}}_k^{-1} + e_{b_k}\mathbf{I}\right)^{-1}\mathbf{C}_k^H\mathbf{\Gamma}_k^{-1}\right) \\
& \left(\mathbf{\Gamma}_k^{-1} - \mathbf{\Gamma}_k^{-1}\mathbf{C}_k\left(\beta_k^{-1}\widetilde{\mathbf{D}}_k^{-1} + e_{b_k}\mathbf{I}\right)^{-1}\mathbf{C}_k^H\mathbf{\Gamma}_k^{-1}\right)\mathbf{C}_k\mathbf{U}_k\widehat{\mathbf{d}}_k \\
& = \widehat{\mathbf{d}}_k^H \mathbf{U}_k \mathbf{C}_k^H \Big(\mathbf{\Gamma}_k^{-2} - \mathbf{\Gamma}_k^{-1}\mathbf{C}_k\left(\beta_k^{-1}\widetilde{\mathbf{D}}_k^{-1} + e_{b_k}\mathbf{I}\right)^{-1}\mathbf{C}_k^H\mathbf{\Gamma}_k^{-2} - \mathbf{\Gamma}_k^{-2}\mathbf{C}_k\left(\beta_k^{-1}\widetilde{\mathbf{D}}_k^{-1} + e_{b_k}\mathbf{I}\right)^{-1}\mathbf{C}_k^H\mathbf{\Gamma}_k^{-1} \\
& + \mathbf{\Gamma}_k^{-1}\mathbf{C}_k\left(\beta_k^{-1}\widetilde{\mathbf{D}}_k^{-1} + e_{b_k}\mathbf{I}\right)^{-1}\mathbf{C}_k^H\mathbf{\Gamma}_k^{-2}\mathbf{C}_k\left(\beta_k^{-1}\widetilde{\mathbf{D}}_k^{-1} + e_{b_k}\mathbf{I}\right)^{-1}\mathbf{C}_k^H\mathbf{\Gamma}_k^{-1}\Big)\mathbf{C}_k\mathbf{U}_k\widehat{\mathbf{d}}_k \\
& \overset{(a)}{=} \widehat{\mathbf{d}}_k^H \mathbf{U}_k \left(e'_{b_k}\mathbf{I} - 2e'_{b_k}e_{b_k}\mathbf{E}_i^{-1} + e'_{b_k}e_{b_k}\mathbf{E}_k^{-2}\right)\mathbf{U}_k\widehat{\mathbf{d}}_k \\
& = e'_{b_k}\widehat{\mathbf{d}}_k^H \mathbf{U}_k \left(\mathbf{I} - e_{b_k}\mathbf{E}_k^{-1}\right)^2 \mathbf{U}_k\widehat{\mathbf{d}}_k.
\end{aligned}$$

In (a) above, we applied Lemma 3 to convert $\mathbf{C}_k^H\mathbf{\Gamma}_k^{-2}\mathbf{C}_k \xrightarrow[a.s]{M\to\infty} \frac{1}{M_{b_k}}\mathrm{tr}\{\mathbf{\Gamma}_k^{-2}\}\mathbf{I}_{L_{k,b_k}}$ and then applying Theorem 2 from [102], we arrive at $\frac{1}{M_{b_k}}\mathrm{tr}\{\mathbf{\Gamma}_k^{-2}\} \xrightarrow[a.s]{M\to\infty} e'_{b_k}$. Combining (70), (71), we can write the signal power as

$$P_{S_k} = \frac{\left(1 - x_{b_k}^{(L,E)}\right)\left(\mathrm{tr}\{\mathbf{U}_k\left(\mathbf{I} - e_{b_k}\mathbf{E}_k^{-1}\right)\mathbf{D}_k\}\right)^2 p_k}{\mathrm{tr}\{\mathbf{U}_k^2\left(\mathbf{I} - e_{b_k}\mathbf{E}_k^{-1}\right)^2\left(\mathbf{D}_k + \widetilde{\sigma}_k^2\mathbf{I}\right)\}}. \tag{72}$$

We define $\mathbf{C}_{k,b_i}\mathbf{d}_{k,b_i}\mathbf{d}_{k,b_i}\mathbf{C}_{k,b_i}^H = \mathbf{\Theta}_{k,b_i}$, $\beta_i^{-1}\widetilde{\mathbf{D}}_i^{-1} + e_{b_i}\mathbf{I} = \mathbf{E}_i$. Considering the interference power part

$$\begin{aligned}
\left\|\mathbf{g}_i''^H\mathbf{h}_{k,b_i}\right\|^2 &= \widehat{\mathbf{d}}_i^H\mathbf{U}_i\mathbf{C}_i^H\mathbf{F}_i^{-1}\mathbf{\Theta}_{k,b_i}\mathbf{F}_i^{-1}\mathbf{C}_i\mathbf{U}_i\widehat{\mathbf{d}}_i \\
&\overset{(a)}{=} \widehat{\mathbf{d}}_i^H\mathbf{U}_i\mathbf{C}_i^H\left(\mathbf{\Gamma}_i^{-1} - \mathbf{\Gamma}_i^{-1}\mathbf{C}_i\mathbf{E}_i^{-1}\mathbf{C}_i^H\mathbf{\Gamma}_i^{-1}\right)\mathbf{\Theta}_{k,b_i}\left(\mathbf{\Gamma}_i^{-1} - \mathbf{\Gamma}_i^{-1}\mathbf{C}_i\mathbf{E}_i^{-1}\mathbf{C}_i^H\mathbf{\Gamma}_i^{-1}\right)\mathbf{C}_i\mathbf{U}_i\widehat{\mathbf{d}}_i \\
&= \widehat{\mathbf{d}}_i^H\mathbf{U}_i\mathbf{C}_i^H\mathbf{\Gamma}_i^{-1}\mathbf{\Theta}_{k,b_i}\mathbf{\Gamma}_i^{-1}\mathbf{C}_i\mathbf{U}_i\widehat{\mathbf{d}}_i + e_{b_i}^2\widehat{\mathbf{d}}_i^H\mathbf{U}_i\mathbf{E}_i^{-1}\mathbf{C}_i^H\mathbf{\Gamma}_i^{-1}\mathbf{\Theta}_{k,b_i}\mathbf{\Gamma}_i^{-1}\mathbf{C}_i\mathbf{E}_i^{-1}\mathbf{U}_i\widehat{\mathbf{d}}_i \\
&\quad - e_{b_i}\widehat{\mathbf{d}}_i^H\mathbf{U}_i\mathbf{C}_i^H\mathbf{\Gamma}_i^{-1}\mathbf{\Theta}_{k,b_i}\mathbf{\Gamma}_i^{-1}\mathbf{C}_i\mathbf{E}_i^{-1}\mathbf{U}_i\widehat{\mathbf{d}}_i - e_{b_i}\widehat{\mathbf{d}}_i^H\mathbf{U}_i\mathbf{E}_i^{-1}\mathbf{C}_i^H\mathbf{\Gamma}_i^{-1}\mathbf{\Theta}_{k,b_i}\mathbf{\Gamma}_i^{-1}\mathbf{C}_i\mathbf{U}_i\widehat{\mathbf{d}}_i
\end{aligned} \tag{73}$$

From the analysis done for ESIP-WSR BF, note that we already derived in (64), $\mathbf{C}_i^H\mathbf{\Gamma}_i^{-1}\mathbf{\Theta}_{k,b_i}\mathbf{\Gamma}_i^{-1}\mathbf{C}_i = e'_{b_i}\frac{1}{M_{b_i}}\mathrm{tr}\{\mathbf{D}_{k,b_i}\mathbf{B}_{k,b_i}^{-2}\}$. Hence, first simplifying the first term above for multiple of identity of eigenvalue matrix $\mathbf{D}_{k,b_i}$ case

$$\widehat{\mathbf{d}}_i^H\mathbf{U}_i\mathbf{C}_i^H\mathbf{\Gamma}_i^{-1}\mathbf{\Theta}_{k,b_i}\mathbf{\Gamma}_i^{-1}\mathbf{C}_i\mathbf{U}_i\widehat{\mathbf{d}}_i = e'_{b_i}\frac{1}{M_{b_i}}\mathrm{tr}\{\mathbf{D}_{k,b_i}\mathbf{B}_{k,b_i}^{-2}\}\mathrm{tr}\{\mathbf{U}_i^2\left(\mathbf{D}_i + \widetilde{\sigma}_i^2\mathbf{I}\right)\}. \tag{74}$$

Further dividing by the $\left\|\mathbf{g}_i''\right\|^2$ and then summing across all the users, we can write

$$\Upsilon_{1k} = \sum_{i\neq k}\frac{\frac{1}{M_{b_i}}\mathrm{tr}\{\mathbf{D}_{k,b_i}\mathbf{B}_{k,b_i}^{-2}\}\mathrm{tr}\{\mathbf{U}_i^2(\mathbf{D}_i + \widetilde{\sigma}_i^2\mathbf{I})\}p_i}{\mathrm{tr}\{\mathbf{U}_i^2(\mathbf{I} - e_{b_i}\mathbf{E}_i^{-1})^2\left(\mathbf{D}_i + \widetilde{\sigma}_i^2\mathbf{I}\right)\}}. \tag{75}$$

Next, we try to simplify the second term in (73) and then divide by the normalization factor. Further by summing across all the users, we obtain

(76)

$$e_{b_i}^2\widehat{\mathbf{d}}_i^H\mathbf{U}_i\mathbf{E}_i^{-1}\mathbf{C}_i^H\mathbf{\Gamma}_i^{-1}\mathbf{\Theta}_{k,b_i}\mathbf{\Gamma}_i^{-1}\mathbf{C}_i\mathbf{E}_i^{-1}\mathbf{U}_i\widehat{\mathbf{d}}_i = e_{b_i}^2\mathrm{tr}\{\mathbf{D}_{k,b_i}\mathbf{B}_{k,b_i}^{-2}\}\mathrm{tr}\{\mathbf{U}_i^2\mathbf{E}_i^{-2}\left(\mathbf{D}_i + \widetilde{\sigma}_i^2\mathbf{I}\right)\},$$

$$\Upsilon_{2k} = \sum_{i\neq k}\frac{\left(1 - x_{b_k}^{(L,E)}\right)\frac{1}{M_{b_i}}\mathrm{tr}\{\mathbf{D}_{k,b_i}\mathbf{B}_{k,b_i}^{-2}\}\mathrm{tr}\{\mathbf{U}_i^2\mathbf{E}_i^{-2}\left(\mathbf{D}_i + \widetilde{\sigma}_i^2\mathbf{I}\right)\}p_i}{\mathrm{tr}\{\mathbf{U}_i^2\left(\mathbf{I} - e_{b_i}\mathbf{E}_i^{-1}\right)^2\left(\mathbf{D}_i + \widetilde{\sigma}_i^2\mathbf{I}\right)\}}.$$

Next, we try to simplify the third term in (73) and after normalizing the resulting term after summing across all the interfering users is denoted as Υ_{3k}

(77)
$$e_{b_i}\widehat{\mathbf{d}}_i^H\mathbf{U}_i\mathbf{C}_i^H\boldsymbol{\Gamma}_i^{-1}\boldsymbol{\Theta}_{k,b_i}\boldsymbol{\Gamma}_i^{-1}\mathbf{C}_i\mathbf{E}_i^{-1}\mathbf{U}_i\widehat{\mathbf{d}}_i = e_{b_i}\mathrm{tr}\{\mathbf{D}_{k,b_i}\mathbf{B}_{k,b_i}^{-2}\}\mathrm{tr}\{\mathbf{U}_i^2\mathbf{E}_i^{-1}\left(\mathbf{D}_i+\widetilde{\sigma}_i^2\mathbf{I}\right)\},$$
$$\Upsilon_{3k} = \sum_{i\neq k}\frac{\left(1-x_{b_k}^{(L,E)}\right)\frac{1}{M_{b_i}}\mathrm{tr}\{\mathbf{D}_{k,b_i}\mathbf{B}_{k,b_i}^{-2}\}\mathrm{tr}\{\mathbf{U}_i^2\mathbf{E}_i^{-1}\left(\mathbf{D}_i+\widetilde{\sigma}_i^2\mathbf{I}\right)\}p_i}{e_{b_i}\mathrm{tr}\{\mathbf{U}_i^2\left(\mathbf{I}-e_{b_i}\mathbf{E}_i^{-1}\right)^2\left(\mathbf{D}_i+\widetilde{\sigma}_i^2\mathbf{I}\right)\}}.$$

Note that the third and fourth term is the same and by combining (72), (75),(76) and (77), we further obtain the deterministic equivalent for the SINR as

(78)
$$\gamma_{k,L}^{(E)} - \frac{P_{S_k}}{\Upsilon_{1k}+\Upsilon_{2k}-2\Upsilon_{3k}+1}\xrightarrow[a.s]{M\to\infty}0.$$

V ESIP-WSR BF with LS Channel Estimate

For LS only estimation, optimizing $EWSR(\mathbf{g})$ (4.16) leads to the following generalized eigenvalue problem

(79)
$$\left(\widehat{\mathbf{h}}_{k,LS}\widehat{\mathbf{h}}_{k,LS}^H+\widetilde{\sigma}_k^2\mathbf{I}\right)\mathbf{g}_k = \nu_k\left(\sum_{i\neq k}\mathbf{S}_{i,b_k}+\mu_{b_k}\mathbf{I}\right)\mathbf{g}_k,$$
$$\widehat{\mathbf{h}}_{k,LS}\widehat{\mathbf{h}}_{k,LS}^H\mathbf{g}_k = \nu_k\left(\sum_{i\neq k}\mathbf{S}_{i,b_k}+\left(\mu_{b_k}-\frac{\widetilde{\sigma}_k^2}{\nu_k}\right)\mathbf{I}\right)\mathbf{g}_k.$$

where $\mathbf{S}_{i,b_k} = \widehat{\mathbf{h}}_{i,b_k,LS}\widehat{\mathbf{h}}_{i,b_k,LS}^H+\widetilde{\sigma}_i^2\mathbf{I}$ here. Since $\widehat{\mathbf{h}}_{k,LS}^H\mathbf{g}_k$ is a scalar, this leads to $\mathbf{g}_k\propto(\sum_{i\neq k}\mathbf{S}_{i,b_k}+(\mu_{b_k}-\frac{\widetilde{\sigma}_k^2}{\nu_k})\mathbf{I})^{-1}\widehat{\mathbf{h}}_{k,LS}$. We Define $\widehat{\mathbf{S}}_{i,b_k} = \widehat{\mathbf{h}}_{i,b_k,LS}\widehat{\mathbf{h}}_{i,b_k,LS}^H$. To find ν_k, we multiply by $\mathbf{g}_k$ on both sides of (79) and obtain ν_k

(80)
$$\mathbf{g}_k^H\widehat{\mathbf{h}}_{k,LS}\widehat{\mathbf{h}}_{k,LS}^H\mathbf{g}_k = \nu_k\mathbf{g}_k^H\left(\sum_{i\neq k}\mathbf{S}_{i,b_k}+\left(\mu_{b_k}-\frac{\widetilde{\sigma}_k^2}{\nu_k}\right)\mathbf{I}\right)\mathbf{g}_k,$$
$$\text{So, } \nu_k \overset{(a)}{=} \widehat{\mathbf{h}}_{k,LS}^H\left(\sum_{i\neq k}\widehat{\mathbf{S}}_{i,b_k}+\sum_{i\neq K}\widetilde{\sigma}_i^2+\left(\mu_{b_k}-\frac{\widetilde{\sigma}_k^2}{\nu_k}\right)\mathbf{I}\right)^{-1}\widehat{\mathbf{h}}_{k,LS},$$

(a) follows directly from substituting for $\mathbf{g}_k = c(\sum_{i\neq k}\widehat{\mathbf{S}}_{i,b_k}+\sum_{i\neq K}\widetilde{\sigma}_i^2+(\mu_{b_k}-\frac{\widetilde{\sigma}_k^2}{\nu_k})\mathbf{I})^{-1}\widehat{\mathbf{h}}_{k,LS}$, c being some constant. First we compute the deterministic equivalent for ν_k. We define $\boldsymbol{\Gamma}_k = \sum_{i\neq k}\mathbf{S}_{i,b_k}+(\mu_{b_k}-\frac{\widetilde{\sigma}_k^2}{\nu_k})\mathbf{I}$. In the large system limit, $\nu_k\xrightarrow[a.s]{M\to\infty}\mathrm{E}(\mathrm{tr}\{\widehat{\mathbf{h}}_{k,LS}^H\boldsymbol{\Gamma}_k^{-1}\widehat{\mathbf{h}}_{k,LS}\}) = \mathrm{tr}\{\boldsymbol{\Gamma}_k^{-1}\,\mathrm{E}(\widehat{\mathbf{h}}_{k,LS}\widehat{\mathbf{h}}_{k,LS}^H)\} = \frac{1}{M_{b_k}}\mathrm{tr}\{\boldsymbol{\Gamma}_k^{-1}\}\mathrm{tr}\{\mathbf{D}_k\}+\widetilde{\sigma}_k^2\,\mathrm{E}(\mathrm{tr}\{\boldsymbol{\Gamma}_k^{-1}\}) = \mathrm{tr}\{\mathbf{D}_k\}e_{b_k}+\widetilde{\sigma}_k^2M_{b_k}e_{b_k}$, where e_c (b_k being c) is defined as

(81)
$$e_c = \left(\frac{1}{M_c}\sum_{i=1}^{K}\frac{\beta_i\lambda_{i,c}^{(1)}}{1+\beta_i\lambda_{i,c}^{(1)}e_c}+\sum_{i=1}^{K}\widetilde{\sigma}_i^2+\mu_c-\frac{\widetilde{\sigma}_k^2}{\nu_k}\right)^{-1},$$

where $\lambda_{i,c}^{(1)}$ in (81) is $\mathrm{tr}\{\mathbf{D}_{k,c}\}+\widetilde{\sigma}_k^2M_c$. Now $\mathbf{g}_k'' = \boldsymbol{\Gamma}_k^{-1}\widehat{\mathbf{h}}_{k,LS}$, considering the signal part and substituting for $\widehat{\mathbf{h}}_{k,LS} = \mathbf{h}_k+\widetilde{\mathbf{h}}_k$ and using the fact that $\mathbf{h}_k$ and $\widetilde{\mathbf{h}}_k$ are independent

(82)
$$\mathbf{g}_k''^H\mathbf{h}_k = \widehat{\mathbf{h}}_{k,LS}^H\boldsymbol{\Gamma}_k^{-1}\mathbf{h}_k\xrightarrow[a.s]{M\to\infty}\mathrm{E}\left(\mathrm{tr}\{\mathbf{C}_k^H\boldsymbol{\Gamma}_k^{-1}\mathbf{C}_k\mathbf{d}_k\mathbf{d}_k^H\}\right)$$
$$= e_{b_k}\mathrm{tr}\{\mathbf{D}_k\},$$
$$\mathbf{g}_k''^H\mathbf{h}_k\mathbf{h}_k\mathbf{g}_k'' = e_{b_k}^2\mathrm{tr}\{\mathbf{D}_k\}^2.$$

Further, $\left\|\mathbf{g}_k''\right\|^2 = \widehat{\mathbf{h}}_{k,LS}^H \boldsymbol{\Gamma}_k^{-2} \widehat{\mathbf{h}}_{k,LS} \xrightarrow[a.s]{M\to\infty} \mathrm{tr}\{\boldsymbol{\Gamma}_k^{-2}(\mathbf{C}_k\mathbf{D}_k\mathbf{C}_k^H + \widetilde{\sigma}_k^2\mathbf{I})\} = e'_{b_k}\mathrm{tr}\{\mathbf{D}_k\} + \widetilde{\sigma}_k^2 M_{b_k} e'_{b_k}$. Substituting for $e'_{b_k} = \frac{e_{b_k}^2}{1-x_{b_k}^{(LS,Opt)}}$ from (62), finally we obtain the deterministic equivalent of the signal power as

$$P_{S_k} = p_k\left(1 - x_{b_k}^{(LS,Opt)}\right)\frac{\left(\mathrm{tr}\{\mathbf{D}_k\}\right)^2}{\mathrm{tr}\{\mathbf{D}_k\}+\widetilde{\sigma}_k^2 M_{b_k}}. \tag{83}$$

Note that $x_{b_k}^{(LS,Opt)}$ has the same definition as in (62), but with the eigenvalues, $\lambda_{k,b_i}^{(1)} = \mathrm{tr}\{\mathbf{D}_{k,b_i} + \widetilde{\sigma}_k^2\mathbf{I}\} + \widetilde{\sigma}_k^2$, $\lambda_{k,b_i}^{(r)} = 0, \forall\, r = 2,...,L_{k,b_i}$. Further considering the interfering user channel powers

$$\mathbf{g}_i''^H \mathbf{h}_{k,b_i}\mathbf{h}_{k,b_i}^H \mathbf{g}_i'' = \widehat{\mathbf{h}}_{i,b_i,LS}^H \boldsymbol{\Gamma}_i^{-1}\mathbf{h}_{k,b_i}\mathbf{h}_{k,b_i}^H \boldsymbol{\Gamma}_i^{-1}\widehat{\mathbf{h}}_{i,b_i,LS} \tag{84}$$

Now, we split the true channel as $\mathbf{h}_{k,b_i} = \mathbf{h}_{LS,k,b_i} - \widetilde{\mathbf{h}}_{k,b_i}$. Then we obtain

$$\begin{aligned}
\mathbf{g}_i''^H \mathbf{h}_{k,b_i}\mathbf{h}_{k,b_i}^H \mathbf{g}_i'' &= \widehat{\mathbf{h}}_{i,LS}^H \boldsymbol{\Gamma}_i^{-1}\widehat{\mathbf{h}}_{k,b_i,LS}\widehat{\mathbf{h}}_{k,b_i,LS}^H \boldsymbol{\Gamma}_i^{-1}\widehat{\mathbf{h}}_{i,LS} + \widehat{\mathbf{h}}_{i,LS}^H \boldsymbol{\Gamma}_i^{-1}\widetilde{\mathbf{h}}_{k,b_i,LS}\widetilde{\mathbf{h}}_{k,b_i,LS}^H \boldsymbol{\Gamma}_i^{-1}\widehat{\mathbf{h}}_{i,LS} \\
&\overset{(a)}{=} \frac{1}{M_{b_i}}\left(\mathrm{tr}\{\mathbf{D}_i\} + \widetilde{\sigma}_i^2 M_{b_i}\right)\left[\mathrm{tr}\{\boldsymbol{\Gamma}_i^{-1}\widehat{\mathbf{h}}_{k,b_i,LS}\widehat{\mathbf{h}}_{k,b_i,LS}^H \boldsymbol{\Gamma}_i^{-1}\} + \mathrm{tr}\{\boldsymbol{\Gamma}_i^{-1}\widetilde{\mathbf{h}}_{k,b_i}\widetilde{\mathbf{h}}_{k,b_i}^H \boldsymbol{\Gamma}_i^{-1}\}\right] \\
&\overset{(b)}{=} e'_{b_i}\frac{1}{M_{b_i}}\left(\mathrm{tr}\{\mathbf{D}_i\} + \widetilde{\sigma}_i^2 M_{b_i}\right)\left[\frac{\mathrm{tr}\{\mathbf{D}_{k,b_i}\} + \widetilde{\sigma}_k^2 M_{b_i}}{\left(1+\beta_k\lambda_{k,b_i}^{(1)} e_{b_i}\right)^2} + \widetilde{\sigma}_k^2 M_{b_i}\right],
\end{aligned} \tag{85}$$

where (a) follows from the convergence of $\widehat{\mathbf{h}}_{i,LS}^H \boldsymbol{\Gamma}_i^{-1}\widehat{\mathbf{h}}_{k,b_i}\widehat{\mathbf{h}}_{k,b_i}^H \boldsymbol{\Gamma}_i^{-1}\widehat{\mathbf{h}}_{i,LS}$ $\xrightarrow[a.s]{M\to\infty} \mathrm{tr}\{\boldsymbol{\Gamma}_i^{-1}\widehat{\mathbf{h}}_{k,b_i}\widehat{\mathbf{h}}_{k,b_i}^H \boldsymbol{\Gamma}_i^{-1}\,\mathrm{E}(\widehat{\mathbf{h}}_{i,LS}\widehat{\mathbf{h}}_{i,LS}^H)\}$ and substituting $\mathrm{E}(\widehat{\mathbf{h}}_{i,LS}\widehat{\mathbf{h}}_{i,LS}^H) = \mathrm{tr}\{\mathbf{C}_i\mathbf{D}_i\mathbf{C}_i^H + \widetilde{\sigma}_i^2\mathbf{I}\}$. Further we apply the matrix inversion lemma and Lemma 3 to convert $\widehat{\mathbf{h}}_{k,b_i,LS}^H \boldsymbol{\Gamma}_i^{-1}$ to $\frac{\widehat{\mathbf{h}}_{k,b_i,LS}^H \boldsymbol{\Gamma}_{ik}^{-1}}{1+\beta_k e_{b_i}}$, where $\boldsymbol{\Gamma}_{ik} = \boldsymbol{\Gamma}_i - \beta_k\widehat{\mathbf{h}}_{k,b_i,LS}\widehat{\mathbf{h}}_{k,b_i,LS}^H$. In the next step, we convert $\mathrm{tr}\{\boldsymbol{\Gamma}_{ik}^{-1}\widehat{\mathbf{h}}_{k,b_i,LS}\widehat{\mathbf{h}}_{k,b_i,LS}^H \boldsymbol{\Gamma}_{ik}^{-1}\}$ $\xrightarrow[a.s]{M\to\infty} (\mathrm{tr}\{\mathbf{D}_{k,b_i}\} + \widetilde{\sigma}_k^2 M_{b_i})\frac{1}{M_{b_i}}\mathrm{tr}\{\boldsymbol{\Gamma}_{ik}^{-2}\} \xrightarrow[a.s]{M\to\infty} (\mathrm{tr}\{\mathbf{D}_{k,b_i}\} + \widetilde{\sigma}_k^2 M_{b_i})e'_{b_i}$ using Lemma 3 and then Theorem 8. Finally dividing by the normalization term $\left\|\mathbf{g}_i\right\|^2$, we write the interference power as

$$P_{I_k} = \sum_{i\neq k}\frac{p_i}{M_{b_i}}\left[\frac{\mathrm{tr}\{\mathbf{D}_{k,b_i}\}+\widetilde{\sigma}_k^2 M_{b_i}}{\left(1+\beta_k\lambda_{k,b_i}^{(1)} e_{b_i}\right)^2} + M_{b_i}\widetilde{\sigma}_k^2\}\right]. \tag{86}$$

Finally combining (83), (86), we obtain the SINR expression as

$$\gamma_{k,LS}^{(Opt)} - \frac{p_k\left(1-x_{b_k}^{(LS,Opt)}\right)\frac{\mathrm{tr}\{\mathbf{D}_k\}^2}{\mathrm{tr}\{\mathbf{D}_k\}+\widetilde{\sigma}_k^2 M_{b_k}}}{\sum\limits_{i\neq k}\frac{p_i}{M_{b_i}}\frac{\mathrm{tr}\{\mathbf{D}_{k,b_i}\}+\widetilde{\sigma}_k^2 M_{b_i}}{\left(1+\beta_k\lambda_{k,b_i}^{(1)} e_{b_i}\right)^2} + \widetilde{\sigma}_k^2\sum\limits_{i\neq k} p_i + 1} \xrightarrow[a.s]{M\to\infty} 0. \tag{87}$$

From (87), it can be observed that with LS only channel estimator, signal power gets decreased by a factor $\mathrm{tr}\{\mathbf{D}_k\} + \widetilde{\sigma}_k^2 M_{b_k}$. For the interference power part, there exists two terms. The first term decreases with increasing SNR and it represents the ZF to the LS estimates of the interfering user channels. The second term remains a constant with increasing SNR and it leads to an SNR offset w.r.t the perfect CSIT at high SNR in the case of estimation error being $\propto 1/SNR$.

VI Naive BF with LS Channel Estimate

Now we consider the naive BF for the LS only channel estimator. We observe that the SINR expression has the similar form as the ESIP-WSR BF with LS channel estimate, with only change being in the value of $x_{b_k}^{(LS,N)}$. Note that $x_{b_k}^{(LS,N)}$ has the same definition as in (62), but with the eigenvalues, $\lambda_{k,b_i}^{(1)} = \mathrm{tr}\{\mathbf{D}_{k,b_i}\} + \widetilde{\sigma}_k^2 M_{b_k}, \lambda_{k,b_i}^{(r)} = 0, \forall\, r = 2,...,L_{k,b_i}$.

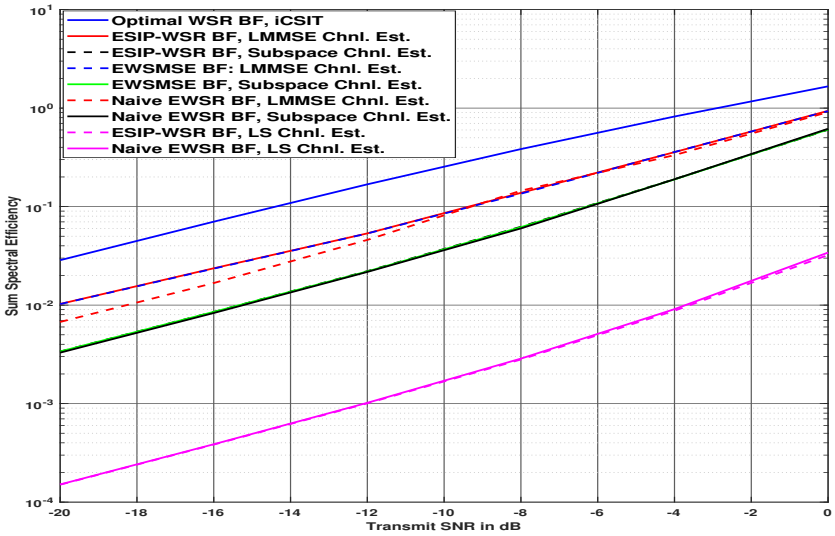


Figure .1: EWSR for $C = 1$ cell, $K_1 = K = 20$ users, $M = 64$, $L = 3$, $\widetilde{\sigma}^2 \propto 1/SNR$, unequal eigenvalues (**D**) .

VII Sum Rate Analysis for Covariance only CSIT case

From the BF expression in (7.36) for CoCSIT case, the deterministic equivalent for signal power part can be written as

$$
\begin{aligned}
|\mathbf{g}_k''^H \mathbf{h}_k|^2 &= \mathbf{e}_{i,max}^H \mathbf{C}_k^H \boldsymbol{\Gamma}_k^{-1} \mathbf{C}_k \mathbf{d}_k \mathbf{d}_k^H \mathbf{C}_k^H \boldsymbol{\Gamma}_k^{-1} \mathbf{C}_k \mathbf{e}_{i,max} \xrightarrow[a.s]{M\to\infty} e_{b_k}^2 \max(\mathbf{D}_k), \\
\left\|\mathbf{g}_k''\right\|^2 &= \mathbf{e}_{i,max}^H \mathbf{C}_k^H \boldsymbol{\Gamma}_k^{-2} \mathbf{C}_k \mathbf{e}_{i,max} \xrightarrow[a.s]{M\to\infty} e'_{b_k},
\end{aligned} \tag{88}
$$

where $\max(\mathbf{D}_k)$ represents the maximum value among the diagonal elements of $\mathbf{D}_k$. Further we obtain the signal power part as, $P_{S_k} = p_k(1 - x_{b_k}^{(C)}) \max(\mathbf{D}_k)$. The analysis for the interference power part remains the same as in Section I for the ESIP-WSR BF with LMMSE channel estimate case.

J Low SNR Analysis ($\widetilde{\sigma}^2 \propto \frac{1}{P}$)

In the Figure .1, we depict the low SNR behaviour of the various BFs with different channel estimates. Further below, we simplify the SINR expressions at low SNR which validate the simulated behaviour. To simplify the analysis, in this section, we consider $M_{b_i} = M, \widetilde{\sigma}_i^2 = \widetilde{\sigma}^2, L_{k,b_i} = L, \forall i,k$.

I ESIP-WSR BF with LMMSE/Subspace Channel Estimate ($\mathbf{D} = \frac{\eta}{L}\mathbf{I}_L$)

In this subsection, we omit the user and BS indices for simplicity. First we consider the simplified case, $\mathbf{D} = \frac{\eta}{L}\mathbf{I}_L$. In this case, the optimal and naive BFs with LMMSE/LS channel estimate converges to the same. Also, we make the approximation that $\widehat{\mathbf{d}} = \mathbf{d} + \widetilde{\mathbf{d}} \approx \widetilde{\mathbf{d}}$ which is accurate at very low SNR.

$$
\begin{aligned}
\mathbf{W}_L &= \widetilde{\sigma}^{-2} \frac{\eta}{L} \left(\widetilde{\sigma}^{-2} \widetilde{\mathbf{d}} \widetilde{\mathbf{d}}^H + \mathbf{I}_L\right) \frac{\eta}{L} + \frac{\eta}{L} \mathbf{I} \\
&= \widetilde{\sigma}^{-4} \frac{\eta}{L} \widetilde{\mathbf{d}} \widetilde{\mathbf{d}}^H + \left(\widetilde{\sigma}^{-2} (\frac{\eta}{L})^2 + \frac{\eta}{L}\right) \mathbf{I}, \mathbf{V}_{max}(\mathbf{W}_L) \propto \widetilde{\mathbf{d}}, \\
\mathbf{W}_S &= \widetilde{\mathbf{d}} \widetilde{\mathbf{d}}^H + \widetilde{\sigma}^2 \mathbf{I}_L, \mathbf{V}_{max}(\mathbf{W}_S) \propto \widetilde{\mathbf{d}}.
\end{aligned} \tag{89}
$$

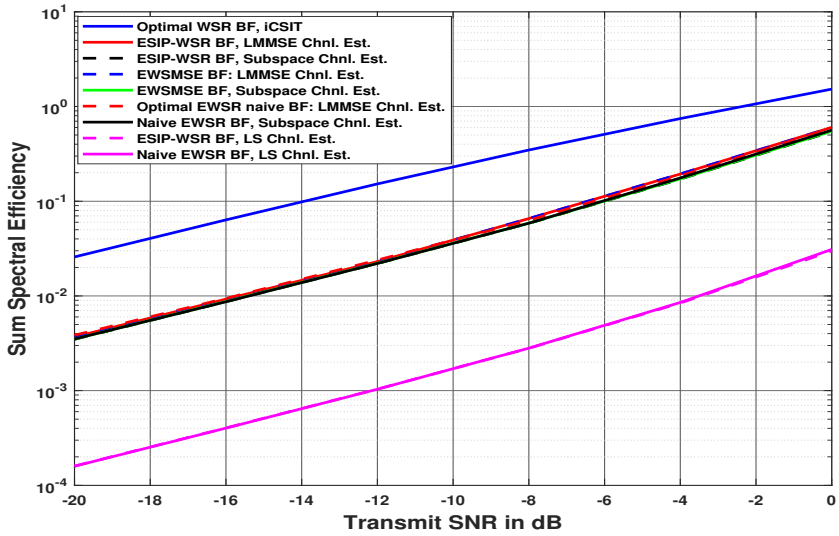


Figure .2: EWSR for $C = 1$ cell, $K_1 = K = 20$ users, $M = 64$, $L = 3$, $\tilde{\sigma}^2 \propto 1/SNR$, $\mathbf{D}_{k,c} = \frac{\eta_{k,c}}{L}\mathbf{I}$.

So $\mathbf{g}''_k = \mathbf{C}\tilde{\mathbf{d}}$ for both LMMSE and subspace in the case of optimal. In the case of naive, $\mathbf{W}_L = \tilde{\sigma}^{-4}\frac{\eta}{L}\tilde{\mathbf{d}}\tilde{\mathbf{d}}^H$ and $\mathbf{W}_S = \tilde{\mathbf{d}}\tilde{\mathbf{d}}^H$ and thus $\mathbf{g}''_k = \mathbf{C}\tilde{\mathbf{d}}$ for both channel estimates. Hence finally computing the signal power part

$$
\begin{aligned}
|\mathbf{g}''^H\mathbf{h}|^2 &= \tilde{\mathbf{d}}^H\mathbf{C}^H\mathbf{C}\mathbf{d}\mathbf{d}^H\mathbf{C}^H\mathbf{C}\tilde{\mathbf{d}} \xrightarrow[a.s]{M\to\infty} \tilde{\sigma}^2\mathrm{tr}\{\mathbf{D}\}.\\
\left\|\mathbf{g}''\right\|^2 &= \tilde{\mathbf{d}}^h\mathbf{C}^H\mathbf{C}\tilde{\mathbf{d}} \xrightarrow[a.s]{M\to\infty} \tilde{\sigma}^2 L.
\end{aligned}
\tag{90}
$$

Hence the SINR becomes $\gamma_{k,L}^{(Opt)} = \gamma_{k,S}^{(Opt)} = \frac{\eta_{k,c}}{L}P$. Also, we observe that for the naive BFs with LMMSE/Subspace channel estimators, $\mathbf{V}_{max}(\mathbf{W}_L) = \mathbf{V}_{max}(\mathbf{W}_S)$ remains the same as the case for ESIP-WSR BF and hence the SINR expressions.

II ESIP-WSR/Naive BFs with LMMSE Channel Estimate (distinct eigenvalues in D)

For simplicity of notation, we drop user and BS indices in this section. So, at low SNR, all the interference are negligible and the BF gets simplified as, $\mathbf{g} = \mathbf{V}_{max}(\widehat{\mathbf{h}}\widehat{\mathbf{h}}^H + R_{\tilde{\mathbf{h}}\tilde{\mathbf{h}}}) = \mathbf{C}\mathbf{V}_{max}(\mathbf{W})$, where, $\mathbf{W} = \mathbf{U}(\widehat{\mathbf{d}}\widehat{\mathbf{d}}^H + \tilde{\sigma}^2\mathbf{I}_L)\mathbf{U}^H + (\mathbf{U}-\mathbf{I})\mathbf{D}(\mathbf{U}-\mathbf{I})^H$ and $\widehat{\mathbf{d}} = \mathbf{C}^H\widehat{\mathbf{h}}_{LS} = \mathbf{d} + \tilde{\mathbf{d}}$. And at low SNR, the following simplifications can be done for $\mathbf{U}$ and $\mathbf{W}$ for LMMSE estimator

$$
\begin{aligned}
\mathbf{U}_L &= \left(\mathbf{I} + \tilde{\sigma}^2\mathbf{D}^{-1}\right)^{-1} \approx \tilde{\sigma}^{-2}\mathbf{D},\\
\text{So, } \mathbf{W}_L &= \tilde{\sigma}^{-2}\mathbf{D}\left(\tilde{\sigma}^{-2}\tilde{\mathbf{d}}\tilde{\mathbf{d}}^H + \mathbf{I}_L\right)\mathbf{D} + \mathbf{D}.
\end{aligned}
\tag{91}
$$

There is no signal concentration along $\mathbf{h}$ or $\mathbf{d}$, $\mathbf{V}_{max}(\mathbf{W}_L)$ remains a random projection in the channel subspace $\mathbf{C}$, if $\mathbf{D}$ is a multiple of identity. If $\mathbf{D}$ is not a multiple of identity, $\mathbf{V}_{max}(\mathbf{W}_L)$ is a function of $\tilde{\mathbf{d}}$, $\mathbf{D}$ and $\tilde{\sigma}^2$, independent of $\mathbf{d}$ which appears in $\mathbf{h}$. Further considering the signal part, $\mathrm{E}(|\mathbf{g}^H\mathbf{h}|^2) = \mathrm{tr}\{\mathbf{D}\,\mathrm{E}\mathbf{V}_{max}(\mathbf{W}_L)\mathbf{V}_{max}(\mathbf{W}_L)^H\}$, for example, in the extreme case, $\mathbf{D} = \mathrm{tr}\{\mathbf{D}\}\mathbf{e}_1\mathbf{e}_1^H$. Then $\mathbf{W}_L$ is proportional to $\mathbf{e}_1\mathbf{e}_1^H$, hence $\mathbf{g} = \mathbf{C}\mathbf{e}_1$. Then $\mathrm{E}(|\mathbf{g}^H\mathbf{h}|^2) = \mathrm{tr}\{\mathbf{D}\}$ but with $\|\mathbf{g}\| = 1$. For subspace channel estimator, $\mathbf{W}_S = \tilde{\mathbf{d}}\tilde{\mathbf{d}}^H + \tilde{\sigma}^2\mathbf{I}_L$ and $\mathbf{V}_{max}(\mathbf{W}_S) = \tilde{\mathbf{d}}$. Substituting for $\mathbf{g} = \mathbf{C}\tilde{\mathbf{d}}$ in the signal part and applying law of large numbers leads to $\mathrm{E}(|\mathbf{g}^H\mathbf{h}|^2) \xrightarrow[a.s]{L,M\to\infty} \tilde{\mathbf{d}}^H\mathbf{D}\tilde{\mathbf{d}} \xrightarrow[a.s]{L\to\infty} \tilde{\sigma}^2\mathrm{tr}\{\mathbf{D}\}$. Similarly computing $\|\mathbf{g}\|^2 = \tilde{\mathbf{d}}^H\tilde{\mathbf{d}} \xrightarrow[a.s]{L\to\infty} \mathrm{tr}\{\tilde{\sigma}^2\mathbf{I}_L\} = \tilde{\sigma}^2 L$. So, we conclude that the signal power

equals tr{**D**} for optimal EWSR BF with LMMSE instead of tr{**D**}/L for subspace channel estimator. This explains why LMMSE performs better than subspace estimator at low SNR, also illustrated by our simulations. For naive BF with subspace estimator, the BF has the same expression as that of the ESIP-WSR BF, since the $\mathbf{V}_{max}(\mathbf{W}_S)$ remains the same as $\widetilde{\mathbf{d}}$ and hence $\mathbf{g} = \mathbf{C}\widetilde{\mathbf{d}}$. Thus the performance at low SNR will be the same.

Further considering the naive BF, with the LMMSE channel estimate

$$\begin{aligned} \mathbf{V}_{max}(\mathbf{W}_L) &\propto \mathbf{D}\widetilde{\mathbf{d}}, \\ \mathbf{g}'' &= \mathbf{C}\mathbf{D}\widetilde{\mathbf{d}}, \\ |\mathbf{g}''^H\mathbf{h}|^2 &= \widetilde{\mathbf{d}}^H\mathbf{D}\mathbf{d}\mathbf{d}^H\mathbf{D}\widetilde{\mathbf{d}} \xrightarrow{M\to\infty} \widetilde{\sigma}^2\mathrm{tr}\{\mathbf{D}^3\}, \\ \left\|\mathbf{g}''\right\|^2 &= \widetilde{\mathbf{d}}^H\mathbf{D}\mathbf{D}\widetilde{\mathbf{d}} \xrightarrow[a.s]{M\to\infty} \widetilde{\sigma}^2\mathrm{tr}\{\mathbf{D}^2\}, \end{aligned} \tag{92}$$

Further we can write the SINR expressions as $\gamma_{k,L}^{(Opt)} - \mathrm{tr}\{\mathbf{D}_k\}p_k \xrightarrow[a.s]{M\to\infty} 0, \gamma_{k,S}^{(Opt)} = \gamma_{k,S}^{(N)}, \gamma_{k,S}^{(Opt)} - \frac{\mathrm{tr}\{\mathbf{D}_k\}}{L}p_k \xrightarrow[a.s]{M\to\infty} 0, \gamma_{k,L}^{(N)} - \frac{\mathrm{tr}\{\mathbf{D}_k^3\}p_k}{\mathrm{tr}\{\mathbf{D}_k^2\}} \xrightarrow[a.s]{M\to\infty} 0$. Note that $\frac{\mathrm{tr}\{\mathbf{D}_k^3\}p_k}{\mathrm{tr}\{\mathbf{D}_k^2\}} < \mathrm{tr}\{\mathbf{D}_k\}p_k$, so the naive BF with LMMSE performs slighly worse than the ESIP-WSR BF.

III BFs with LS Channel Estimate

From (87), at low SNR, since the interference power part is negligible, $\mathbf{g}_k \propto \widehat{\mathbf{h}}_{k,LS}$. The deterministic equivalent of the signal part simplifies as follows, $\mathbf{g}_k''^H\mathbf{h}_k = \widehat{\mathbf{h}}_{k,LS}^H\mathbf{h}_k \xrightarrow{M\to\infty} \mathrm{tr}\{\mathrm{E}(\mathbf{h}_k\mathbf{h}_k^H)\} = \mathrm{tr}\{\mathbf{D}_k\}$. Thus the signal power becomes $|\mathbf{g}_k''^H\mathbf{h}_k|^2 = \mathrm{tr}\{\mathbf{D}_k\}^2$. Similarly, the normalization part for $\mathbf{g}_k$, $\left\|\mathbf{g}_k''\right\|^2 = \widehat{\mathbf{h}}_{k,LS}^H\widehat{\mathbf{h}}_{k,LS} \xrightarrow[a.s]{M\to\infty} \mathrm{tr}\{\mathbf{D}_k + \widetilde{\sigma}^2\mathbf{I}_L\}$. Finally, we can write the SINR as

$$\gamma_{k,LS}^{(Opt)} = \gamma_{k,LS}^{(N)} = \frac{\mathrm{tr}\{\mathbf{D}_k\}^2 p_k}{\mathrm{tr}\{\mathbf{D}_k\} + \widetilde{\sigma}_k^2 M_{b_k}}. \tag{93}$$

Naive BF has the same expression since at high SNR BF expression computed above for ESIP-WSR case, $\mathbf{g}_k \propto \widehat{\mathbf{h}}_{k,LS}$ depends only on the LS channel estimator.

IV EWSMSE BF with LMMSE/Subspace Channel Estimate

For the simple case which we evaluate of varying channel attenuations, $\mathbf{D}_k = \frac{\eta_k}{L}I, \widetilde{\mathbf{d}}_k = \widetilde{d}_k\mathbf{I} = \widetilde{\sigma}^2\eta_k/(\widetilde{\sigma}^2 L + \eta_k)\mathbf{I}$, so $\mathbf{g}_k''$ becomes, $\mathbf{g}_k'' \propto \mathbf{\Gamma}_k^{-1}\mathbf{C}_k\mathbf{U}_k\widehat{\mathbf{d}}_k(1 - e_{b_k}/(e_{b_k} + \widetilde{d}_k^{-1})) = \mathbf{\Gamma}_k^{-1}\mathbf{C}_k\mathbf{U}_k\widehat{\mathbf{d}}_k$.

Now in the case of $\mathbf{D}_k = \frac{\eta_k}{L}\mathbf{I}$, the BF expression for ESIP-WSR BF is also $\mathbf{g}_k \propto \mathbf{\Gamma}_k^{-1}\mathbf{C}_k\mathbf{U}_k\widehat{\mathbf{d}}_k$. So the BFs are same for ESIP and EWSMSE at any SNR.

Further considering the case of distinct eigenvalues in $\mathbf{D}_k$, we use the fact that $e_{b_k} = \frac{1}{\mu_{b_k}}$ at low SNR. Also, from (7.6), we can approximate $(\mathbf{I} + \widetilde{\sigma}_k^2\mathbf{D}_k)^{-1} \approx \widetilde{\sigma}_k^{-2}\mathbf{D}_k$, $\widetilde{\mathbf{d}}_k \approx \mathbf{D}_k - \widetilde{\sigma}_k^{-2}\mathbf{D}_k^2 \approx \mathbf{D}_k$. Hence we can approximate, $\beta_k^{-1}\widetilde{\mathbf{d}}_k^{-1} + e_{b_k} \approx \beta_k^{-1}\widetilde{\mathbf{d}}_k^{-1}$ and we further obtain $\mathbf{g}_k \propto \mathbf{\Gamma}_k^{-1}\mathbf{C}_k(\mathbf{I} - \beta_k\widetilde{\mathbf{d}}_k e_{b_k})\mathbf{U}_k\widehat{\mathbf{d}}_k$. Since $e_{b_k} \to 0$ at low SNR, $\mathbf{g}_k \propto \mathbf{\Gamma}_k^{-1}\mathbf{C}_k\mathbf{U}_k\widehat{\mathbf{d}}_k$, which is the same expression as the ESIP-WSR BF and this explains the observed performance as seen in the simulations.

K High SNR Analysis ($\widetilde{\sigma}^2 \propto \frac{1}{P}$)

In Figure .3, we plot the comparison of various BF performance at high SNR when the channel estimation error is inversely proportional to the Tx SNR. Below we provide detailed high SNR analysis for the various BFs with different channel estimates and the SNR offsets are depicted in the table 8.1.

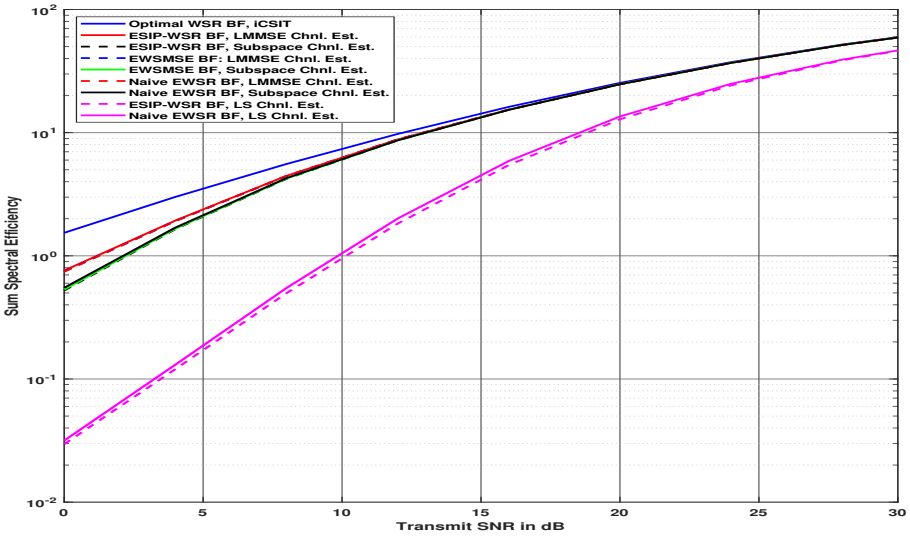


Figure .3: EWSR for $C = 1$ cell, $K_1 = K = 20$ users, $M = 64$, $L = 3$, $\tilde{\sigma}^2 \propto 1/SNR$.

I BFs with LS Channel Estimate

Considering each of the terms in the interference power part in (86) for the ESIP-WSR BF, the summation term $\sum_{i\neq k} \frac{p_i}{M_{b_i}} \frac{\text{tr}\{\mathbf{D}_{k,b_i}\}+\tilde{\sigma}_k^2 M_{b_i}}{(1+\beta_k \lambda^{(1)}_{k,b_i} e_{b_i})^2} \propto \frac{KL}{M}\frac{1}{SNR}$ and hence tends to zero. Thus the BF $\mathbf{g}_i$ does ZF to all the interfering LS channel estimates $\mathbf{h}_{k,b_i,LS}$ at high SNR. Also, we make use of the high SNR result that $e_{b_i} \propto 1/\mu_{b_i} \to \infty$. With $\mathbf{D}_{k,b_i} = \frac{\eta_{k,b_i}}{L}\mathbf{I}, M_{b_i} = M, \tilde{\sigma}_i^2 = \tilde{\sigma}^2, p_i = \frac{P}{K}, L_{k,b_i} = L, \forall i,k$, the SINR can be written as

$$\gamma^{(Opt)}_{k,LS} = \frac{\frac{P}{K}\left(1-\frac{K}{M}\right)\frac{\eta^2_{k,b_k}}{\eta_{k,b_k}+\tilde{\sigma}_k^2 M_{b_k}}}{\tilde{\sigma}^2 CP+1}, \tag{94}$$

where we made use of the fact that $x^{(LS)}_{b_k} \xrightarrow[a.s]{M\to\infty} \frac{K}{M}$ since the $(L-1)$ eigenvalues $\lambda^{(r)}_{i,b_k} \to 0$ and the largest eigenvalue $\lambda^{(r)}_{i,b_k} = \text{tr}\{\mathbf{D}_{i,b_k}\} + 2\tilde{\sigma}^2 M$ which makes the term $\frac{\beta_i^2 \lambda^{(r),2}_{i,c}}{(1+\beta_i \lambda^{(r)}_{i,c} e_c)^2} \approx 1$. Thus $x^{(LS)}_{b_k} = \frac{K}{M}, \forall k$. The signal and interference power parts for the naive BF remains the same. Also, note that $x^{(LS)}_{b_k}$ remains the same as that of the ESIP-WSR BF since the $(L-1)$ eigenvalues for $\lambda^{(r)}_{i,c}$ are zero and the largest eigenvalue $\lambda^{(1)}_{i,c}$ will be the same as that of the ESIP-WSR BF. Hence the SINR expression remains the same, $\gamma^{(Opt)}_{k,LS} = \gamma^{(N)}_{k,LS}$.

II ESIP-WSR/Naive BFs with LMMSE/Subspace Channel Estimate

Considering the simplifications at high SNR, from (61), we observe that e_c increases with SNR since μ_c converges to zero with high SNR and $e_c \gg 1$. So $\beta_i \tilde{d}^{(r)}_{i,c} e_c \to \infty$. Now considering each terms in the interference power part,

$$\frac{1}{M_{b_i}}\text{tr}\{\mathbf{D}_{k,b_i}\mathbf{B}^{-2}_{k,b_i}\} = \frac{1}{M_{b_i}}\left[\sum_{r=1}^{L_{k,b_i}} \frac{\eta^{(r)}_{k,b_i}}{\left(1+\beta_i \lambda^{(r)}_{k,b_i} e_{b_i}\right)^2}\right]. \tag{95}$$

As $\mu_c \to 0$, it is clear from (95) that, $\frac{1}{M_{b_i}}\text{tr}\{\mathbf{D}_{k,b_i}\mathbf{B}^{-2}_{k,b_i}\} \to 0$, since each of the summation term becomes proportional to $1/SNR$ or $\frac{1}{M_{b_i}}\text{tr}\{\mathbf{D}_{k,b_i}\mathbf{B}^{-2}_{k,b_i}\} \to \frac{K\sum_i L_{k,b_i}}{M}\frac{1}{SNR}$. At high SNR, for the LMMSE

channel estimate, the weight matrices $\mathbf{W}_L = (\mathbf{I} + \widetilde{\sigma}^2\mathbf{D}^{-1})^{-1} \approx \mathbf{I} = \mathbf{W}_S$. Hence the optimal or naive BF expression for both LMMSE or Subspace channel estimate can be represented as $\mathbf{g}_k'' \propto \mathbf{\Gamma}_k^{-1}\mathbf{C}_k\widehat{\mathbf{d}}_k$. Further considering the signal power part

$$
\begin{aligned}
|\mathbf{g}_k''^H\mathbf{h}_k|^2 &= \widehat{\mathbf{d}}_k^H\mathbf{C}_k^H\mathbf{\Gamma}_k^{-1}\mathbf{C}_k\mathbf{d}_k\mathbf{d}_k^H\mathbf{C}_k^H\mathbf{\Gamma}_k^{-1}\mathbf{C}_k\widehat{\mathbf{d}}_k \\
&= e_{b_k}^2\,\widehat{\mathbf{d}}_k^H\mathbf{d}_k\mathbf{d}_k^H\widehat{\mathbf{d}}_k \xrightarrow[a.s]{M\to\infty} e_{b_k}^2\,\mathrm{tr}\{\mathbf{D}_k\}\mathrm{tr}\{\mathbf{D}_k + \widetilde{\sigma}^2\mathbf{I}_M\}, \\
\left\|\mathbf{g}_k''^H\mathbf{g}_k''\right\|^2 &= e_{b_k}'\,\widehat{\mathbf{d}}_k^H\widehat{\mathbf{d}}_k \xrightarrow[a.s]{M\to\infty} e_{b_k}'\,\mathrm{tr}\{\mathbf{D}_k + \widetilde{\sigma}^2\mathbf{I}_M\}.
\end{aligned} \tag{96}
$$

Substituting for e_{b_k}', the SINR becomes

$$
\begin{aligned}
\gamma_{k,L}^{(Opt)} = \gamma_{k,S}^{(Opt)} &= \left(1 - x_{b_k}^{(L)}\right)\mathrm{tr}\{\mathbf{D}_k\}\frac{P}{K} \\
&= \left(1 - \frac{K}{M}\right)\mathrm{tr}\{\mathbf{D}_k\}.
\end{aligned} \tag{97}
$$

where we made the approximation that $x_{b_k}^{(L)} = \frac{K}{M}$.

III EWSMSE BF with LMMSE/Subspace Channel Estimate

Starting with the EWSMSE expression derived in Section IV, $\mathbf{g}_k'' = \mathbf{\Gamma}_k^{-1}\mathbf{C}_k\mathbf{U}_k\widehat{\mathbf{d}}_k - \mathbf{\Gamma}_k^{-1}\mathbf{C}_k(\widetilde{\mathbf{d}}_k^{-1} + e_{b_k}\mathbf{I})^{-1}e_{b_k}\mathbf{U}_k\widehat{\mathbf{d}}_k$. At high SNR, $\widetilde{\mathbf{d}}_k^{-1} = \widetilde{\sigma}^{-2}\mathbf{I}$. Thus we can approximate $\widetilde{\mathbf{d}}_k^{-1} + e_{b_k}\mathbf{I} \approx (\widetilde{\sigma}^2 + e_{b_k})\mathbf{I}$. Thus $\mathbf{g}_k'' \propto \mathbf{\Gamma}_k^{-1}\mathbf{C}_k\mathbf{U}_k\widehat{\mathbf{d}}_k$, which is same as that of optimal/naive BFs with the subspace and LMMSE channel estimates.

L Sum Rate Analysis with Constant Channel Estimation Error

In this section, to simplify the analysis, we consider identical system parameters, $M_c = M, L_{k,c} = L, \widetilde{\sigma}_k^2 = \widetilde{\sigma}^2, K_c = K/C, \forall k, c$.

I Naive BFs with LMMSE/Subspace Channel Estimate

The SINR expression remains the same as that derived in (69). Note that the terms $x_{b_k}^{(L,N)}, x_{b_k}^{(S,N)}$ does converge to $\frac{K}{M}$. This results due to the fact that $L-1$ of the eigenvalues of $\mathbf{W}_{i,b_k}$ are zero and the only nonzero eigenvalue remains a constant, $\mathrm{tr}\{\mathbf{U}_{i,b_k}^2(\mathbf{D}_{i,b_k} + \widetilde{\sigma}^2\mathbf{I})\}$. Also, from (61), we know that $e_{b_k} \to \infty$ at high SNR. Further substituting e_{b_k} in (62), $x_{b_k}^{(L,N)}, x_{b_k}^{(S,N)}$ converge to $\frac{K}{M}$.

II ESIP-WSR BFs with LMMSE/Subspace Channel Estimate

To simplify the analysis, we only consider the case, $\mathbf{D}_{k,b_i} = \frac{\eta_{k,b_i}}{L}\mathbf{I}_L$. In this case, $\mathbf{v}_{k,b_k} = \widehat{\mathbf{d}}_k$. Also the BF expression with LMMSE and subspace estimators are the same. We start with the deterministic equivalent of the signal power part

$$
\begin{aligned}
\mathbf{g}_k''^H\mathbf{h}_k &= \widehat{\mathbf{d}}_k^H\mathbf{C}_k^H\mathbf{\Gamma}_k^{-1}\mathbf{C}_k\mathbf{d}_k \\
&\xrightarrow[a.s]{M\to\infty} e_{b_k}\mathrm{tr}\{\mathbf{D}_k\}, \\
\left\|\mathbf{g}_k''\right\|^2 &= \widehat{\mathbf{d}}_k^H\mathbf{C}_k^H\mathbf{\Gamma}_k^{-2}\mathbf{C}_k\widehat{\mathbf{d}}_k \xrightarrow[a.s]{M\to\infty} e_{b_k}'\,\mathrm{tr}\{\mathbf{D}_k + \widetilde{\sigma}^2\mathbf{I}\}.
\end{aligned} \tag{98}
$$

Next we look at the interference power part

$$
\begin{aligned}
|\mathbf{g}_i''^H \mathbf{h}_{k,b_i}|^2 &= \widehat{\mathbf{d}}_i^H \mathbf{C}_i^H \mathbf{\Gamma}_i^{-1} \mathbf{C}_{k,b_i} \mathbf{d}_{k,b_i} \mathbf{d}_{k,b_i}^H \mathbf{C}_{k,b_i}^H \mathbf{\Gamma}_i^{-1} \mathbf{C}_i \widehat{\mathbf{d}}_i \\
&\xrightarrow[a.s]{M\to\infty} \widehat{\mathbf{d}}_i^H \widehat{\mathbf{d}}_i \frac{1}{M} \mathrm{tr}\{\mathbf{\Gamma}_i^{-2} \mathbf{C}_{k,b_i} \mathbf{d}_{k,b_i} \mathbf{d}_{k,b_i}^H \mathbf{C}_{k,b_i}^H\}.
\end{aligned} \tag{99}
$$

Here also, we apply the Lemma 4 twice to convert $\mathbf{\Gamma}_i^{-1}$ to $\mathbf{C}_i^{-1}$, where $\mathbf{\Xi}_i = \mathbf{\Gamma}_i - \beta_k \mathbf{C}_{k,b_i} \widehat{\mathbf{d}}_{k,b_i} \widehat{\mathbf{d}}_{k,b_i}^H \mathbf{C}_{k,b_i}^H$.

$$
\tfrac{1}{M} \mathrm{tr}\{\mathbf{\Gamma}_i^{-2} \mathbf{C}_{k,b_i} \mathbf{d}_{k,b_i} \mathbf{d}_{k,b_i}^H \mathbf{C}_{k,b_i}^H\} \xrightarrow[a.s]{M\to\infty} \tfrac{1}{M} \mathrm{tr}\{\mathbf{\Xi}_i^{-2} \mathbf{C}_{k,b_i} \mathbf{d}_{k,b_i} \mathbf{d}_{k,b_i}^H \mathbf{C}_{k,b_i}^H\} \tag{100}
$$

Further we apply Lemma 3 on each row of $\mathbf{C}_{k,b_i}^H \mathbf{\Xi}_i^{-1}$ such that $\mathbf{C}_{k,b_i}^H \mathbf{\Xi}_i^{-1} = \mathbf{B}_{k,b_i}^{-1} \mathbf{C}_{k,b_i}^H \mathbf{\Xi}_{i,\bar{k}}^{-1}$, where $\mathbf{\Xi}_{i,\bar{k}} = \mathbf{\Xi}_i - \beta_k \mathbf{C}_{k,b_i} \widetilde{\mathbf{D}}_{k,b_i} \mathbf{C}_{k,b_i}^H$. $\mathbf{B}_{k,b_i}$ is defined as, $\mathbf{B}_{k,b_i} = \mathbf{I} + \beta_k e_i \widetilde{\mathbf{D}}_{k,b_i}$. Finally we obtain

$$
\begin{aligned}
\frac{1}{M} \mathrm{tr}\{\mathbf{\Xi}_i^{-2} \mathbf{C}_{k,b_i} \mathbf{d}_{k,b_i} \mathbf{d}_{k,b_i}^H \mathbf{C}_{k,b_i}^H\} &\xrightarrow[a.s]{M\to\infty} \frac{1}{M} \mathrm{tr}\{\mathbf{\Xi}_{i,\bar{k}}^{-2} \mathbf{C}_{k,b_i} \mathbf{B}_{k,b_i}^{-1} \mathbf{d}_{k,b_i} \mathbf{d}_{k,b_i}^H \mathbf{B}_{k,b_i}^{-1} \mathbf{C}_{k,b_i}^H\} \\
&\xrightarrow[a.s]{M\to\infty} \frac{1}{M} \mathrm{tr}\{\mathbf{B}_{k,b_i}^{-2} \mathbf{D}_{k,b_i}\} e_i'
\end{aligned} \tag{101}
$$

Thus the interference power can be written as, $P_{I_k} = \frac{1}{M} \sum_{i \neq k}^{K} \frac{\mathrm{tr}\{\mathbf{B}_{k,b_i}^{-2} \mathbf{D}_{k,b_i}\}}{\mathrm{tr}\{\mathbf{D}_i + \widetilde{\sigma}^2 \mathbf{I}\}} p_i + 1$. Thus the SINR can be written as

$$
\begin{aligned}
\gamma_{k,L}^{(Opt)} = \gamma_{k,S}^{(Opt)} &= \frac{\left(1 - \frac{KL}{M}\right) (\mathrm{tr}\{\mathbf{D}_k\})^2}{\mathrm{tr}\{\left(\mathbf{D}_k + \widetilde{\sigma}^2 \mathbf{I}\right\}) \left(\frac{1}{M} \sum_{i \neq k}^{K} \frac{\mathrm{tr}\{\mathbf{B}_{k,b_i}^{-2} \mathbf{D}_{k,b_i}\}}{\mathrm{tr}\{\mathbf{D}_i + \widetilde{\sigma}^2 \mathbf{I}\}} p_i + 1 \right)} \\
&\overset{(a)}{=} \frac{\left(1 - \frac{KL}{M}\right) (\mathrm{tr}\{\mathbf{D}_k\})^2}{\mathrm{tr}\{(\mathbf{D}_k + \widetilde{\sigma}^2 \mathbf{I}\})},
\end{aligned} \tag{102}
$$

where (a) follows from: since $\mathrm{tr}\{\mathbf{D}_i + \widetilde{\sigma}^2 \mathbf{I}\}$ is a constant and $e_i \propto SNR$, each of the term in $\mathbf{B}_{k,b_i}^{-2} \propto \frac{1}{SNR^2}$. Since we consider $\frac{KL}{M}$ as a constant, the interference power $P_{I_k} \approx \frac{1}{SNR} + 1 = 1$. Thus ZF to all the interfering paths happen at high SNR for the case of ESIP-WSR BF with LMMSE/Subspace channel estimate. The SINR of the subspace estimate is the same as that of the LMMSE here since the BF expressions are equal when $\mathbf{D}_{k,b_i}$ is a multiple of identity.

III BFs with LS Channel Estimate

For the ESIP-WSR BF with LS channel estimate, we look at how the SINR expression derived in (87) evolves with SNR. In the constant channel estimation error case also, the summation term $\sum_{i \neq k} \frac{p_i}{M_{b_i}} \frac{\mathrm{tr}\{\mathbf{D}_{k,b_i}\} + \widetilde{\sigma}_k^2 M_{b_i}}{(1 + \beta_k \lambda_{k,b_i}^{(1)} e_{b_i})^2} \propto \frac{KL}{M} \frac{1}{SNR}$. However the second term $\widetilde{\sigma}^2 CP$ increases with SNR which explains the saturation at high SNR. Also, note that $x_{b_k}^{(LS)} = \frac{K}{M}$ due to the ZF to interfering LS channel estimates.

$$
\gamma_{k,LS}^{(Opt)} = \frac{\left(1 - \frac{KL}{M}\right) \frac{(\mathrm{tr}\{\mathbf{D}_k\})^2}{\mathrm{tr}\{\mathbf{D}_k\} + \widetilde{\sigma}_k^2 M_{b_k}} \frac{P}{K}}{\widetilde{\sigma}^2 CP + 1} \tag{103}
$$

Note that for the naive BF, there is only one eigenvalue for $\mathbf{W}$ and hence the number of ZF components is K (ZF to the interfering LS channel estimates which is rank one).

IV EWSMSE BF with LMMSE/Subspace Channel Estimator

To simplify further, considering the simplified case of $\mathbf{D}_{k,b_i} = \frac{\eta_{k,b_i}}{L}\mathbf{I}$, where $\widetilde{\mathbf{D}}_{k,b_i} = \widetilde{d}_{k,b_i}\mathbf{I}$, we obtain the simplification

$$\begin{aligned}\left(\mathbf{I} - e_{b_k}\mathbf{E}_k^{-1}\right)^2 &= \frac{1}{(1+\beta_k\widetilde{d}_k e_k)^2}\mathbf{I},\\ \text{So, } \left\|\mathbf{g}_k''\right\|^2 &= \frac{e'_{b_k}}{\left(1+\beta_k\widetilde{d}_k e_k\right)^2}\mathrm{tr}\{\mathbf{U}_k^2\left(\mathbf{D}_k + \widetilde{\sigma}_k^2\mathbf{I}\right)\}.\end{aligned} \tag{104}$$

Next, we look at further simplifications when $\mathbf{D}_{k,c} = \frac{\eta_{k,c}}{L_{k,c}}, \forall k, c$. The first term in the interference power, (75) becomes $(1+\beta_i\widetilde{d}_{i,b_i}e_i)^2\mathrm{tr}\{\mathbf{D}_{k,b_i}\mathbf{B}_{k,b_i}^{-2}\}$. The second term (76) can be simplified as $(1-x_{b_i}^{(L,E)})\beta_i\widetilde{d}_{i,b_i}\mathrm{tr}\{\mathbf{D}_{k,b_i}\mathbf{B}_{k,b_i}^{-2}\}$. And the third term (77) becomes, $\frac{(1-x_{b_i}^{(L,E)})(1+\beta_i\widetilde{d}_{i,b_i}e_i)}{e_{b_i}}\mathrm{tr}\{\mathbf{D}_{k,b_i}\mathbf{B}_{k,b_i}^{-2}\}$. Note that the fourth term is equivalent to the third term and thus have the same simplified expression. Finally, combining, we obtain the SINR expression as

$$\gamma_{k,L}^{(E)} - \frac{\left(1-x_{b_k}^{(L,E)}\right)\frac{\eta_{k,b_k}^2}{\left(\eta_{k,b_k}+\widetilde{\sigma}_k^2 L\right)}p_k}{\frac{1}{M}\sum_{i\neq k}p_i\mathrm{tr}\{\mathbf{D}_{k,b_i}\mathbf{B}_{k,b_i}^{-2}\}[\left(1+\beta_i\widetilde{d}_{i,b_i}e_{b_i}\right)^2+(1-x_{b_i}^{L,E})\beta_i\widetilde{d}_{i,b_i}-2\frac{\left(1-x_{b_i}^{(L,E)}\right)\left(1+\beta_i\widetilde{d}_{i,b_i}e_{b_i}\right)}{e_{b_i}}]+1} \xrightarrow[a.s]{M\to\infty} 0. \tag{105}$$

From Section I, we know that $e_c \to \infty$ at high SNR, so we can simplify further the above equation at high SNR. The second term in the denominator $\frac{P}{KM}\sum_{i\neq k}(1-x_{b_i}^{L,E})\beta_i\widetilde{d}_{i,b_i}\mathrm{tr}\{\mathbf{D}_{k,b_i}\mathbf{B}_{k,b_i}^{-2}\}$ becomes proportional to $\frac{1}{SNR}$, similarly the third term $\frac{P}{KM}\sum_{i\neq k}\mathrm{tr}\{\mathbf{D}_{k,b_i}\mathbf{B}_{k,b_i}^{-2}\}2\frac{(1-x_{b_i}^{(L,E)})(1+\beta_i\widetilde{d}_{i,b_i}e_{b_i})}{e_{b_i}}$ also is proportional to $\frac{1}{SNR}$. Hence all terms except the first term in the interference power part goes to zero. The first term can be simplified as

$$\frac{CP}{KM}\sum_{i\neq k}\mathrm{tr}\{\mathbf{D}_{k,b_i}\mathbf{B}_{k,b_i}^{-2}\}\left(1+\beta_i\widetilde{d}_{i,b_i}e_{b_i}\right)^2 \xrightarrow[a.s]{M\to\infty} \frac{CP}{KM}\sum_{i\neq k}\mathrm{tr}\{\mathbf{D}_{k,b_i}\overline{\mathbf{B}}_{k,b_i}^{-2}\}\left(\beta_i\widetilde{d}_{i,b_i}\right)^2, \tag{106}$$

where

$$\overline{\mathbf{B}}_{k,b_i} = \mathrm{diag}\left(\beta_k\lambda_{k,b_i}^{(1)},....,\beta_k\lambda_{k,b_i}^{(r)}\right). \tag{107}$$

Finally we obtain

$$\gamma_{k,L}^{(E)} - \frac{\left(1-x_{b_k}^{(L,E)}\right)\frac{\eta_{k,b_k}^2}{\left(\eta_{k,b_k}+\widetilde{\sigma}_k^2 L\right)}p_k}{\frac{CP}{KM}\sum_{i\neq k}\mathrm{tr}\{\mathbf{D}_{k,b_i}\overline{\mathbf{B}}_{k,b_i}^{-2}\}\left(\beta_i\widetilde{d}_{i,b_i}\right)^2+1} \xrightarrow[a.s]{M\to\infty} 0. \tag{108}$$

Note that $\widetilde{d}_{i,b_i}$ is a constant since the channel estimation error is constant. Hence the term $(1+\beta_i\widetilde{d}_{i,b_i}e_{b_i})^2 \propto SNR^2$. The only term remaining as part of the interference power in the denominator, $\frac{1}{M}\sum_{i\neq k}p_i\mathrm{tr}\{\mathbf{D}_{k,b_i}\mathbf{B}_{k,b_i}^{-2}\}[(1+\beta_i\widetilde{d}_{i,b_i}e_{b_i})^2] \propto \frac{KL}{M}SNR$. Hence this term grows as SNR increases and this explains why the rate saturates at high SNR for EWSMSE BFs.

BIBLIOGRAPHY

[1] E. Björnson, E. G. Larsson, and T. L. Marzetta, " Massive MIMO: Ten Myths and One Critical Question ," *IEEE Comm's Mag.*, Feb. 2016.

[2] M. E. Tipping, "Sparse Bayesian learning and the relevance vector machine," *J. Mach. Learn. Res.*, vol. 1, pp. 211–244, 2001.

[3] M. Jouni, M. D. Mura, and P. Comon, "Classification of Hyperspectral Images as Tensors Using Nonnegative CP Decomposition ," *IEEE Transactions on Geoscience and Remote Sensing*, May 2019.

[4] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wire. Commun.*, Nov. 2010.

[5] X. Zhang, A. Molisch, and S. Kung, "Variable-phase-shift-based RF-baseband codesign for MIMO antenna selection," *IEEE Trans. on Sig. Process.*, vol. 53, no. 11, pp. 4091–4103, 2005.

[6] G. Caire and S. Shamai, "On the Achievable Throughput of a Multiantenna Gaussian Broadcast Channel," *IEEE Trans. Info. Theory*, vol. 49, pp. 1691–1706, July 2003.

[7] S. Vishwanath, N. Jindal, and A. Goldsmith, "Duality, Achievable Rates, and Sum-Rate Capacity of Gaussian MIMO Broadcast Channels," *IEEE Trans. Inform. Theory*, vol. 49, no. 10, Oct. 2003.

[8] M. Costa, "Writing on Dirty Paper," *IEEE Trans. Inform. Theory*, May 1983.

[9] S. S. Christensen, R. Agarwal, E. de Carvalho, and J. Cioffi, "Weighted sum-rate maximization using weighted MMSE for MIMO-BC beamforming design," *IEEE Trans. on Wireless Commun.*, December 2008.

[10] C. K. Thomas and D. Slock, "Mixed time scale weighted sum rate maximization for hybrid beamforming in multi-cell MU-MIMO systems," in *IEEE Globecom Wkshps.*, Singapore, December 2017.

[11] F. Negro, I. Ghauri, and D. T. M. Slock, "Deterministic annealing design and analysis of the noisy MIMO interference channel," in *Proc. IEEE Inf. Theo. and Applic. Workshop (ITA), San Diego, CA, USA*, 2011.

[12] A. M. Tulino and S. Verdú, "Random Matrix Theory and Wireless Communications," in *Now Publishers Inc*, 2004.

[13] R. Couillet and M. Debbah, "Random Matrix Methods for Wireless Communications," in *Cambridge University Press*, 2011.

[14] S. Wagner, R. Couillet, M. Debbah, and D. T. M. Slock, "Large system analysis of linear precoding in MISO broadcast channels with limited feedback," *IEEE Trans. Inf. Theory*, vol. 58, no. 7, pp. 4509–4538, July 2012.

[15] S. Wagner and D. Slock, "Weighted sum rate maximization of correlated MISO broadcast channels under linear precoding : a large system analysis," in *Proc. IEEE Int'l Workshop on Sig. Proc. Advances in Wireless Comm's (SPAWC), San Fransisco, USA*, 2011.

[16] W. Tabikh and D. Slock, "MIMO IBC beamforming with combined channel estimate and covariance CSIT," in *Proc. IEEE Int'l Symp. on Info. Theo. (ISIT)*, Aachen, Germany, 2017.

[17] Q. H. Spencer, A. L. Swindlehurst, and M. Haardt, "Zero-forcing methods for downlink spatial multiplexing in multi-user MIMO channels," *IEEE Trans. on Sig. Process.*, vol. 49, pp. 461–471, Feb. 2004.

[18] V. Stankovic and M. Haardt, "Generalized design of multi-user MIMO precoding matrices," *IEEE Trans. on Wire. Commun.*, vol. 7, pp. 953–961, Mar. 2008.

[19] W. Yu and J. M. Cioffi, " Sum Capacity of Gaussian Vector Broadcast Channels ," *IEEE Trans. on Info. Theo.*, Aug. 2004.

[20] S. M. Kay, "Fundamentals of statistical signal processing: Estimation theory," in *1st ed. Prentice Hall PTR*, April 1993.

[21] A. Adhikary, J. Nam, J. Y. Ahn, and G. Caire, "Joint spatial division and multiplexing: the large-scale array regime," *IEEE Trans. on Information Theory*, vol. 59, no. 10, pp. 6441—-6463, 2013.

[22] A. Alkhateeb, O. E. Ayach, G. Leus, and R. Heath, "Channel estimation and hybrid precoding for millimeter wave cellular systems," *IEEE J. for Sel. Topics in Sig Process.*, vol. 8, no. 5, pp. 831–846, October 2014.

[23] S. Park, J.Park, A. Yazdan, and R. W. Heath, "Exploiting spatial channel covariance for hybrid precoding in Massive MIMO systems," *IEEE Trans. on Sig. Process.*, vol. 65, no. 14, July 2017.

[24] F. Sohrabi and W. Yu, "Hybrid digital and analog beamforming design for large scale antenna arrays," *IEEE J. for Sel. Topics in Sig. Process.*, vol. 10, no. 3, pp. 501–513, April 2016.

[25] L. Liang, Y. Dai, W. Xu, and X. Dong, "How to approach zero-forcing under RF chain limitations in large mmWave multiuser systems ?" in *Proc. IEEE/CIC Intl. Conf. Commun, China*, October 2014.

[26] R. Mendez-Rial, C. Rusu, N. Gonzalez-Prelcic, A. Alkhateeb, and R. W. Heath, "Hybrid MIMO architectures for millimeter wave communications: Phase shifters or switches ?" *IEEE Access*, vol. 4, pp. 247–267, January 2016.

[27] P. Stoica and Y. Selén, "Cyclic Minimizers, Majorization Techniques, and the Expectation-Maximization Algorithm: A Refresher," *IEEE Sig. Proc. Mag.*, Jan. 2004.

[28] Q. Shi, M. Razaviyayn, Z. Q. Luo, and C. He, "An iteratively weighted mmse approach to distributed sum-utility maximization for a MIMO interfering broadcast channel," *IEEE Trans. on Sig. Process.*, vol. 59, no. 9, September 2011.

[29] F. Negro, I. Ghauri, and D. T. M. Slock, "Sum rate maximization in the noisy MIMO interfering broadcast channel with partial CSIT via the expected weighted MSE," in *Proc. Int'l Symposium on Wireless Communication Systems, ISWCS, Paris, France*, 2012.

[30] S. J. Kim and G. B. Giannakis, "Optimal resource allocation for MIMO ad-hoc cognitive radio networks," in *IEEE Trans. on Inf. Theory*, vol. 57, May 2011, pp. 3117–3131.

[31] K. B. Petersen and M. S. Pedersen, "The matrix cookbook," in *URL http://www2.imm.dtu.dk/pubdb/p.php?3274.*, November 2011.

[32] F. Negro, I. Ghauri, and D. Slock, "Deterministic Annealing Design and Analysis of the Noisy MIMO Interference Channel," in *Proc. IEEE Information Theory and Applications workshop (ITA)*, San Diego, CA, USA, Feb. 2011.

[33] W. Yu and T. Lan, "Transmitter optimization for the multi-antenna downlink with per-antenna power constraints," *IEEE Trans. Signal Proc.*, vol. 55, no. 6, pp. 2646 – 2660, June 2007.

[34] K. Karakayali, R. Yates, G. Foschini, and R. Valenzuela, "Optimum zero-forcing beamforming with per-antenna power constraints," in *Proc. IEEE Int'l Symp. on Info. Theo. (ISIT)*, Nice, France, 20107.

[35] T. E. Bogale and L. Vandendorpe, "Sum mse optimization for downlink multiuser mimo systems with per antenna power constraint: Downlink-uplink duality approach," in *IEEE 22nd Intl Symp. on PIMRC*, 2011.

[36] T. M. Pham, R. Farrell, J. Dooley, E. Dutkiewicz, D. N. Nguyen, and L.-N. Tran, "Efficient zero-forcing precoder design for weighted sum-rate maximization with per-antenna power constraint," *IEEE Trans. on Vehic. Techn.*, November 2017.

[37] R. Zhang, "Cooperative multi-cell block diagonalization with per-basestation power constraints," *IEEE J. Sel. Areas Commun.*, vol. 28, no. 9, p. 1435–1445, December 2010.

[38] J. S. Herd and M. D. Conway, "The Evolution to Modern Phased Array Architectures," *Proc. IEEE*, Mar. 2016.

[39] S. Boyd and L. Vandenberghe, "Convex optimization," in *Cambridge, U.K.: Cambridge Univ. Press*, 2004.

[40] J. Chen and V. K. N. Lau, "Two-Tier Precoding for FDD Multi-Cell Massive MIMO Time-Varying Interference Networks," *IEEE J. Sel. Areas. Comm.*, June 2014.

[41] A. Alkhateeb and R. Heath, "Frequency selective hybrid precoding for limited feedback millimeter wave systems," *IEEE Trans. on Commun.*, vol. 64, no. 5, May 2016.

[42] K. Venugopal *et al.*, "Channel estimation for hybrid architecture-based wideband millimeter wave systems," *IEEE J. for Sel. Areas in Commun.*, vol. 35, no. 9, Sept. 2017.

[43] T. Cover and J. Thomas, "Elements of information theory," in *Wiley*, 1991.

[44] C. K. Thomas and D. Slock, "Hybrid beamforming design in multi-cell MU-MIMO systems with per-RF or per-antenna power constraints," in *IEEE 88th Veh. Tech. Conf.*, Chicago, USA, 2018.

[45] J. H. Brady and A. M. Sayeed, "Wideband communication with high-dimensional arrays: New results and transceiver architectures," in *Proc. IEEE Int. Conf. Commun. Workshop (ICCW)*, London, UK, 2015.

[46] C. K. Thomas and D. Slock, "Deterministic annealing for hybrid beamforming design in multi-cell MU-MIMO systems," in *Proc. IEEE Int. Wkshp on Sig. Proc. Adv. in Wirel. Commun.*, Kalamata, Greece, 2018.

[47] F. Sohrabi and W. Yu, "Hybrid analog and digital beamforming for mmWave OFDM large-scale antenna arrays," *IEEE J. Sel. Topics in Sig. Process.*, vol. 35, no. 7, July 2017.

[48] S. Li and R. D. Murch, "Full-Duplex Wireless Communication using Transmitter Output based Echo cancellation," in *IEEE GLOBECOM*, 2011.

[49] D. Bharadia, E. McMilin, and S. Katti, "Full Duplex Radios," in *ACM SIGCOMM Computer Communication Review*, vol. 43, no. 4, 2013.

[50] M. Duarte, C. Dick, and A. Sabharwal, "Experiment-Driven Characterization of Full-Duplex Wireless Systems," *IEEE Trans. on Wire. Commun.*, vol. 11, no. 12, 2012.

[51] T. Riihonen and R. Wichman, "Analog and Digital Self-Interference Cancellation in Full-Duplex MIMO-OFDM Transceivers with Limited Resolution in A/D Conversion," in *46th IEEE asilomar conference on signals, systems and computers (ASILOMAR)*, 2012.

[52] 3GPP Technical Report V14.1.0, "Study on scenarios and requirements for next generation access technologies," 2017.

[53] S. Huberman and T. Le-Ngoc, "MIMO full-duplex precoding: A joint beamforming and self-interference cancellation structure," *IEEE Trans. on Wire. Commun.*, vol. 14, no. 4, 2014.

[54] A. C. Cirik, R. Wang, , Y. Hua, and M. Latva-aho, "Weighted sum-rate maximization for full-fuplex MIMO interference channels," *IEEE Trans. on Commun.*, vol. 63, no. 3, Mar. 2015.

[55] P. Aquilina, A. C. Cirik, and T. Ratnarajah, "Weighted Sum Rate Maximization in Full-Duplex Multi-User Multi-Cell MIMO Networks," *IEEE Trans. on Commun.*, vol. 65, no. 4, Apr. 2017.

[56] K. Satyanarayana, M. El-Hajjar, , P.-H. Kuo, A. Mourad, and L. Hanzo, "Hybrid beamforming design for full-duplex millimeter wave communication," *IEEE Trans. on Veh. Techn.*, vol. 68, no. 2, Feb 2019.

[57] C. K. Thomas, C. K. Sheemar, and D. Slock, "Multi-Stage/Hybrid BF under Limited Dynamic Range for OFDM FD Backhaul with MIMO SI Nulling," in *ISWCS Workshop on Full Duplex Communications for 5G and Beyond 5G*, Oulu, Finland, 2019.

[58] O. Taghizadeh, V. Radhakrishnan, A. C. Cirik, R. Mathar, and L. Lampe, "Hardware impairments aware transceiver design for bidirectional full-fuplex MIMO OFDM systems," *IEEE Trans. on Vehic. Tech.*, vol. 67, no. 8, Aug. 2018.

[59] B. P. Day, A. R. Margetts, D. W. Bliss, and P. Schniter, "Full-Duplex bidirectional MIMO: achievable rates under limited dynamic range," *IEEE Trans. on Sig. Process.*, vol. 60, no. 7, July 2012.

[60] A. Hjorungnes and D. Gesbert, "Complex-valued matrix differentiation: Techniques and key results," *IEEE Trans. on Sig. Process.*, vol. 55, pp. 2740–2746, Jun. 2007.

[61] E. Björnson, J. Hoydis, and L. Sanguinetti, "Massive MIMO networks: Spectral, Energy, and Hardware Efficiency," in *Found. Trends. in Sig. Process.*, 2017.

[62] W. Tabikh, Y. Yuan-Wu, and D. Slock, "Beamforming design with combined channel estimate and covariance CSIT via random matrix theory," in *Proc. IEEE Int'l Conf. on Commun. (ICC)*, 2017.

[63] C. K. Thomas and D. Slock, "Rate maximization under partial CSIT for multi-stage/hybrid BF under limited dynamic range for OFDM full-duplex systems," in *IEEE 88th Veh. Tech. Conf.*, Antwerp, Belgium, 2020.

[64] ——, "A Massive MIMO Stochastic Geometry Analysis of Various Beamforming Designs with Partial CSIT," in *13th Wkshp. on Spat. Stoch. Mod. for Wire. Netw., WIOPT 2019*, Avignon, France, 2019.

[65] S. Ye and R. S. Blum, "Optimized Signaling for MIMO Interference Systems with Feedback," *IEEE Trans. on Sig. Process.*, vol. 51, no. 11, Dec. 2003.

[66] E. Larsson, O. Edfors, F. Tufvesson, and T. Marzetta, "Massive MIMO for Next Generation Wireless Systems," *IEEE Comm's Mag.*, Feb. 2014.

[67] Y. Lejosne, M. Bashar, D. Slock, and Y. Yuan-Wu, "From MU Massive MISO to Pathwise MU Massive MIMO," in *Proc. IEEE Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, Toronto, Canada, June 2014.

[68] ——, "Decoupled, Rank Reduced, Massive and Frequency-Selective Aspects in MIMO Interfering Broadcast Channels," in *Proc. IEEE Int'l Symp. on Communications, Control and Sig. Proc. (ISCCSP)*, Athens, Greece, May 2014.

[69] W. Tabikh, D. Slock, and Y. Yuan-Wu, "The Pathwise MIMO Interfering Broadcast Channel," in *Proc. IEEE Workshop on Information Theory and Applications (ITA)*, San Diego, CA, USA, Feb. 2015.

[70] F. Negro, S. Prasad Shenoy, I. Ghauri, and D. Slock, "On the MIMO Interference Channel," in *Proc. IEEE Information Theory and Applications workshop (ITA)*, San Diego, CA, USA, Feb. 2010.

[71] A. Medles and D. Slock, "Matched Filter Bounds without Channel Knowledge at the Receiver," in *Proc. IEEE Asilomar Conf. Signals, Systems and Computers*, Pacific Grove, CA, USA, Nov. 2003.

[72] M. Bashar, Y. Lejosne, D. Slock, and Y. Yuan-Wu, "MIMO Broadcast Channels with Gaussian CSIT and Application to Location based CSIT," in *Proc. IEEE Wkshp on Info. Theo. and Appl. (ITA)*, San Diego, CA, USA, Feb. 2014.

[73] W. Tabikh, D. Slock, and Y. Yuan-Wu, "Optimal Beamforming with Combined Channel and Path CSIT for Multi-Cell Multi-User MIMO," in *Proc. IEEE Workshop on Information Theory and Applications (ITA)*, San Diego, CA, USA, Feb. 2016.

[74] T. L. Marzetta, E. G. Larsson, H. Yang, and H. Q. Ngo, "Fundamentals of Massive MIMO," in *Cambridge University Press*, 2016.

[75] E. Björnson, J. Hoydis, and L. Sanguinetti, "Massive MIMO networks: Spectral, energy, and hardware efficiency," *Foundations and Trends® in Signal Processing*, vol. 11, no. 3-4, pp. 154–655, 2017.

[76] E. Björnson, L. Sanguinetti, H. Wymeersch, J. Hoydis, and T. L. Marzetta, "Massive MIMO is a Reality What is Next? Five Promising Research Directions for Antenna Arrays," *Dig. Sig. Process.*, Nov. 2019.

[77] E. Björnson, J. Hoydis, and L. Sanguinetti, "Massive MIMO Has Unlimited Capacity," *IEEE Trans. on Wire. Commun.*, Jan. 2018.

[78] L. Sanguinetti, E. Björnson, and J. Hoydis, "Toward massive MIMO 2.0: Understanding spatial correlation, interference suppression, and pilot contamination," *IEEE Trans. Commun.*, Jan 2020.

[79] B. Clerckx, H. Joudeh, C. Hao, M. Dai, and B. Rassouli, "Rate Splitting for MIMO Wireless Networks: a Promising PHY-Layer Strategy for LTE Evolution," *IEEE Comm. Mag.*, vol. 54, no. 5, May. 2016.

[80] H. Joudeh and B. Clerckx, "Sum-Rate Maximization for Linearly Precoded Downlink Multiuser MISO Systems With Partial CSIT: A Rate-Splitting Approach," *IEEE Trans. on Comm.*, Nov 2016.

[81] A. G. Davoodi and S. A. Jafar, "Aligned Image Sets under Channel Uncertainty: Settling Conjectures on the Collapse of Degrees of Freedom under Finite Precision CSIT," *IEEE Trans. on Info. Theo*, Oct 2016.

[82] E. Piovano and B. Clerckx, "Optimal DoF region of the K-user MISO BC with partial CSIT," *IEEE Commun. Lett.*, vol. 21, no. 11, Nov. 2017.

[83] H. Joudeh and B. Clerckx, "Robust Transmission in Downlink Multiuser MISO Systems: A Rate-Splitting Approach," *IEEE Trans. On Sig. Process.*, vol. 64, no. 23, Dec. 2016.

[84] Y. Mao, B. Clerckx, and V. O. K. Li, "Rate-Splitting Multiple Access for Downlink Communication Systems: Bridging, Generalizing, and Outperforming SDMA and NOMA," *EURASIP J. Wireless Commun. Netw.*, vol. 2018, no. 1, May 2018.

[85] Y. Mao and B. Clerckx, "Beyond Dirty Paper Coding for Multi-Antenna Broadcast Channel with Partial CSIT: A Rate-Splitting Approach," *https://arxiv.org/abs/1912.05409*, 2019.

[86] M. Dai, B. Clerckx, D. Gesbert, and G. Caire, "A Rate Splitting Strategy for Massive MIMO with Imperfect CSIT," *IEEE Trans. on Wire. Commun.*, vol. 15, no. 7, July 2016.

[87] A. Papazafeiropoulos, B. Clerckx, and T. Ratnarajah, "Rate-Splitting to Mitigate Residual Transceiver Hardware Impairments in Massive MIMO Systems," *IEEE Trans. on Veh. Tech.*, vol. 66, no. 9, Sept. 2017.

[88] H. Yang and T. L. Marzetta, "Performance of conjugate and zero-forcing beamforming in large-scale antenna systems," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 172–179, Feb. 2013.

[89] C. K. Thomas, B. Clerckx, L. Sanguinetti, and D. Slock, "A rate splitting strategy for mitigating intra-cell pilot contamination in massive MIMO," in *IEEE International Conference on Communications*, Dublin, Ireland, 2020.

[90] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 186 – 195, February 2014.

[91] H. Asgharimoghaddam, A. Tolli, and N. Rajatheva, "Decentralizing the optimal multi-cell beamforming via large system analysis," in *IEEE Int'l Conf. on Communications (ICC)*, 2014.

[92] W. Tabikh, Y. Yuan-Wu, and D. Slock, "Decentralizing multi-cell maximum weighted sum rate precoding via large system analysis ," in *EUSIPCO*, Budapest, Hungary, Sept. 2016.

[93] C. K. Thomas, W. Tabikh, D. Slock, and Y. Yuan-Wu, "Noncoherent multi-user MIMO communications using covariance CSIT," in *Asilomar Conference on Signals, Systems, and Computers,*, Pacific Grove, CA, USA, Nov. 2017.

[94] A. Muller, R. Couillet, E. Bjornson, S. Wagner, and M. Debbah, "Interference-aware rzf precoding for multi cell downlink systems," *IEEE Trans. Signal Proc.*, vol. 63, no. 15, pp. 3959 – 3973, August 2015.

[95] L. Sanguinetti, A. L. Moustakas, E. Bjornson, and M. Debbah, "Large system analysis of the energy consumption distribution in multi-user MIMO systems with mobility," *IEEE Trans. on Wirel. Commun.*, vol. 14, no. 3, pp. 1730–1745, March 2015.

[96] P. Billingsley, "Probability and measure." in *Hoboken, NJ: Wiley*, 1995.

[97] G. Caire and K. R. Kumar, "Information theoretic foundations of adaptive coded modulation," *Proc. IEEE*, vol. 95, no. 12, pp. 2274–2298, December 2007.

[98] H. Joudeh and B. Clerckx, "Sum-rate maximization for linearly precoded downlink multiuser MISO systems with partial CSIT: A rate-splitting approach," *IEEE Transactions on Communications*, vol. 64, no. 11, pp. 4847–4861, November 2016.

[99] C. K. Thomas and D. Slock, "Reduced-order zero-forcing beamforming vs optimal beamforming and dirty Paper coding and massive MIMO Analysis," in *10th IEEE Sens. Arr. and Mul.chnl. Sig. Process. Wkshp.*, Sheffield, UK, 2018.

[100] ——, "Massive MISO IBC beamforming - a multi-antenna stochastic geometry perspective," in *IEEE Globecom Wkshps*, Abu Dhabi, UAE, 2018.

[101] M. M´edard, "The Effect upon Channel Capacity in Wireless Communications of Perfect and Imperfect Knowledge of the Channel," *IEEE Trans. on Info. Theo.*, vol. 46, no. 3, 2000.

[102] J. Hoydis, S. ten Brink, and M. Debbah, "Massive MIMO in the UL/DL of Cellular Networks: How Many Antennas Do We Need?" *IEEE Journ. on Select. Areas in Commun.*, vol. 31, no. 2, pp. 160–171, 2013.

[103] S. Lakshminarayana, M. Assaad, and M. Debbah, "Coordinated multicell beamforming for massive MIMO: A random matrix approach," *IEEE Trans. Information Theory*, vol. 61, no. 6, pp. 3387 – 3412, June 2015.

[104] L. Sanguinetti, A. Kammoun, and M. Debbah, "Theoretical Performance Limits of Massive MIMO with Uncorrelated Rician Fading Channels ," *IEEE Trans. on Commun.*, Nov. 2018.

[105] A. Kammoun *et al.*, "Asymptotic Analysis of RZF in Large-Scale MU-MIMO Systems over Rician Channels ," *Trans. on Info. Theo.*, 2019.

[106] C. K. Thomas and D. Slock, "Massive MISO IBC beamforming - a multi-antenna stochastic geometry perspective," in *IEEE Globecom Wkshps.*, Abu dhabi, UAE, 2018.

[107] ——, "Massive MISO IBC Reduced Order Zero Forcing Beamforming - a Multi-Antenna Stochastic Geometry Perspective," in *Intl. Conf. on Comp., Netwrk. and Commun. (ICNC)*, Honolulu, Hawaii, USA, 2019.

[108] N. Jindal, "MIMO Broadcast Channels With Finite-Rate Feedback ," *IEEE Trans. on Info. Theo.*, Nov. 2006.

[109] G. Caire *et al.*, "Multiuser MIMO Achievable Rates With Downlink Training and Channel State Feedback ," *IEEE Trans. on Info. Theo.*, Jun. 2010.

[110] M. Haenggi, J. G. Andrews, , F. Baccelli, O. Dousse, and M. Franceschetti, "Stochastic Geometry and Random Graphs for the Analysis and Design of Wireless Networks ," *IEEE Jrnl. on Sel. Areas in Commun.*, Sept. 2009.

[111] G. George, R. K. Mungara, A. Lozano, and M. Haenggi, "Ergodic Spectral Efficiency in MIMO Cellular Networks," *IEEE Trans. on Wireless Communications*, May 2017.

[112] Z. Chen and E. Björnson, "Channel Hardening and Favorable Propagation in Cell-Free Massive MIMO with Stochastic Geometry," *IEEE Trans. on Comm.*, vol. 66, no. 11, 2018.

[113] A. Klein *et al.*, "Direction-of-Arrival of Partial Waves in Wideband Mobile Radio Channels for Intelligent Antenna Concepts ," *in Proceedings of Vehic. Tech. Conf. (VTC)*, 1996.

[114] Y. Zhou *et al.*, "Experimental Study of MIMO Channel Statistics and Capacity via the Virtual Channel Representation ," *Univ. Wisconsin-Madison, Madison, WI, USA, Tech. Rep*, 2007.

[115] H. Tataria *et al.*, "Channel Correlation Diversity in MU-MIMO Systems-Analysis and Measurements ," *arXiv preprint arXiv:1904.07726*, Apr. 2019.

[116] H. Q. Ngo, M. Matthaiou, and E. G. Larsson, "Performance Analysis of Large Scale MU-MIMO with Optimal Linear Receivers ," *IEEE Swedish Communication Technologies Workshop (Swe-CTW)*, 2012.

[117] X. Li *et al.*, "Massive MIMO with Multi-Cell MMSE Processing: Exploiting all Pilots for Interference Suppression ," *EURASIP Journal on Wireless Communications and Networking*, 2017.

[118] J. Nam, G. Caire, and J. Ha, "On the Role of Transmit Correlation Diversity in Multiuser MIMO Systems ," *IEEE Trans. on Info. Theo.*, Jan. 2017.

[119] L. Sanguinetti, E. Björnson, and J. Hoydis, "Towards Massive MIMO 2.0: Understanding Spatial Correlation, Interference Suppression, and Pilot Contamination ," *arXiv preprint arXiv:1904.03406*, Apr. 2019.

[120] T. L. Marzetta *et al.*, "Fundamentals of Massive MIMO," in *Cambridge University Press*, 2016.

[121] O. Ozdogan, E. Björnson, and E. G. Larsson, "Massive MIMO with Spatially Correlated Rician Fading Channels ," *IEEE Trans. on Commun.*, Jan. 2019.

[122] J. Nam, G. Caire, M. Debbah, and H. V. Poor, "Capacity Scaling of Massive MIMO in Strong Spatial Correlation Regimes ," *IEEE Trans. On Inf. Theo.*, May 2020.

[123] R. B. Ertel *et al.*, "Overview of Spatial Channel Models for Antenna Array Communication systems ," *IEEE Pers. Commun.*, Feb. 1998.

[124] C. Qian, X. Fu, and N. D. Sidiropoulos, "Algebraic Channel Estimation Algorithms for FDD Massive MIMO systems ," *arXiv preprint arXiv:1903.08938*, 2019.

[125] C. K. Thomas and D. Slock, "Space Alternating Variational Estimation and Kronecker Structured Dictionary Learning," in *IEEE Intl. Conf. on Acous. Spee. and Sig. Process. (ICASSP)*, May 2019.

[126] K. Gopala and D. Slock, "A Refined Analysis of the Gap Between Expected Rate for Partial CSIT and the Massive MIMO Rate Limit," in *IEEE Proc. IEEE Int'l Conf. Acoustics Speech and Sig. Proc. (ICASSP)*, Calgary, Canada, 2018.

[127] T. Marzetta *et al.*, *Fundamentals of Massive MIMO*. Cambridge U. Press, 2016.

[128] L. Liu, C. Oestges, J. Poutanen, K. Haneda, P. Vainikainen, F. Quitin, F. Tufvesson, and P. D. Doncker, "The COST 2100 MIMO Channel Model ," *IEEE Wire. Commun.*, 2012.

[129] X. Gao *et al.*, "Massive MIMO Performance Evaluation Based on Measured Propagation Data ," *IEEE Trans. Wireless Commun.*, Jul. 2015.

[130] M. Shafi *et al.*, "Microwave vs. Millimeter-Wave Propagation Channels: Key Differences and Impact on 5G Cellular systems ," *IEEE Commun. Mag.*, Dec. 2018.

[131] H. Yin *et al.*, "A Coordinated Approach to Channel Estimation in Large-Scale Multiple-Antenna Systems ," *IEEE Jrnl. on Sel. Areas in Commun.*, Feb. 2013.

[132] C. Qian, X. Fu, N. D. Sidiropoulos, and Y. Yang, "Tensor-based parameter estimation of double directional massive MIMO channel with dual-polarized antennas," in *ICASSP*, 2018.

[133] Z. Yang, L. Xie, and C. Zhang, "Off-Grid Direction of Arrival Estimation using Sparse Bayesian Inference," *IEEE Trans. On Sig. Process.*, vol. 61, no. 1, 2013.

[134] I. F. Gorodnitsky, J. S. George, and B. D. Rao, "Neuromagnetic Source Imaging with FOCUSS: a Recursive Weighted Minimum Norm Algorithm," *J. Electroencephalog. Clinical Neurophysiol.*, vol. 95, no. 4, 1995.

[135] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit ," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, December 2007.

[136] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit ," *SIAM J. Sci. Comput.*, vol. 20, no. 1, pp. 33–61, 1998.

[137] D. Wipf and S. Nagarajan, "Iterative reweighted l_1 and l_2 methods for finding sparse solutions ," *IEEE J. Sel. Topics Sig. Process.*, vol. 4, no. 2, pp. 317–329, April 2010.

[138] R. Giri and B. D. Rao, "Type I and type II bayesian methods for sparse signal recovery using scale mixtures," *IEEE Trans. on Sig Process.*, vol. 64, no. 13, pp. 3418–3428, 2018.

[139] D. P. Wipf and B. D. Rao, "Sparse Bayesian Learning for Basis Selection ," *IEEE Trans. on Sig. Process.*, vol. 52, no. 8, pp. 2153–2164, August 2004.

[140] T. Park and G. Casella, "The Bayesian Lasso," *J. Amer. Statist. Assoc.*, vol. 103, no. 482, Nov. 2008.

[141] M. E. Tipping and A. C. Faul, "Fast marginal likelihood maximisation for sparse Bayesian models," in *AISTATS*, January 2003.

[142] D. L. Donoho, A. Maleki, and A. Montanari, "Message-passing algorithms for compressed sensing ," *PNAS*, vol. 106, no. 45, pp. 18 914–18 919, November 2009.

[143] S. Rangan, "Generalized approximate message passing for estimation with random linear mixing," in *Proc. IEEE Int. Symp. Inf. Theory*, Saint Petersburg, Russia, August 2011, p. 2168–2172.

[144] S. Rangan, P. Schniter, and A. Fletcher, "On the convergence of approximate message passing with arbitrary matrices," in *Proc. IEEE Int. Symp. Inf. Theory*, 2014.

[145] S. Rangan, P. Schniter, and A. K. Fletcher, "Vector Approximate Message Passing," *IEEE Trans. On Info. Theo.*, vol. 65, no. 10, Oct. 2019.

[146] M. J. Beal, "Variational algorithms for approximate Bayesian inference," in *Thesis, Univeristy of Cambridge, UK*, May 2003.

[147] D. G. Tzikas, A. C. Likas, and N. P. Galatsanos, "The variational approximation for Bayesian inference," *IEEE Sig. Process. Mag.*, vol. 29, no. 6, pp. 131–146, November 2008.

[148] D. Shutin, T. Buchgraber, S. R. Kulkarni, and H. V. Poor, "Fast variational sparse bayesian learning with automatic relevance determination for superimposed signals," *IEEE Trans. on Sig. Process*, vol. 59, no. 12, December 2011.

[149] C. Couvreur and Y. Bresler, "Decomposition of a mixture of Gaussian AR processes," *IEEE Intl. Conf. on Acous., Speech, and Sig. Process.*, vol. 3, pp. 1605–1608, 1995.

[150] W. Gao, S. Tsai, and J. S. Lehnert, "Diversity combining for DS/SS systems with time-varying, correlated fading branches," *IEEE Trans. on Commun.*, vol. 51, no. 2, pp. 284–295, Feb 2003.

[151] M. Feder and E. Weinstein, "Parameter estimation of superimposed signals using the em algorithm," *IEEE Transactions on Acous., Speech and Sig. Process.,*, vol. 36, no. 4, pp. 477–489, Apr 1988.

[152] R. D. Bass, V. D. Norum, and L. Swartz, "Optimal multichannel nonlinear filtering," *J. Mufh. Anal. Appl.*, vol. 16, pp. 152 – 164, 1966.

[153] A. H. Jazwinski, *Stochastic Processes and Filtering Theory*. Dover Publications, 1970.

[154] R. Henriksen, "The truncated second-order nonlinear filter revisited," *IEEE Transactions on Automatic Control*, vol. 27, no. 1, pp. 247 – 251, feb 1982.

[155] M. Athans, R. Wishner, and A. Bertolini, "Suboptimal state estimation for continuous-time nonlinear systems from discrete noisy measurements," *IEEE Transactions on Automatic Control*, vol. 13, no. 5, pp. 504 – 514, oct 1968.

[156] J. Villares and G. Vazquez, "The quadratic extended Kalman filter," in *Sens. Arr. and Multichnl. Sig. Process. Wkshp. (SAM), 2004*, july 2004, pp. 480 – 484.

[157] R. Mehra, "On the identification of variances and adaptive kalman filtering," *IEEE Transactions on Automatic Control*, vol. 15, no. 2, pp. 175 – 184, apr 1970.

[158] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *JOURNAL OF THE ROYAL STATISTICAL SOCIETY, SERIES B*, vol. 39, no. 1, pp. 1–38, 1977.

[159] L. Ljung, "Asymptotic behavior of the extended kalman filter as a parameter estimator for linear systems," *IEEE Transactions on Automatic Control*, vol. 24, no. 1, pp. 36 – 50, feb 1979.

[160] P. Tichavsky, C. Muravchik, and A. Nehorai, "Posterior cramer-rao bounds for discrete-time nonlinear filtering," *Signal Processing, IEEE Transactions on*, vol. 46, no. 5, pp. 1386 – 1396, may 1998.

[161] H. L. V. Trees and K. L. Bell, *Bayesian Bounds for Parameter Estimation and Nonlinear Filtering/Tracking*. Wiley-IEEE Press, 2007.

[162] R. Giri and B. Rao, "Type I and Type II Bayesian Methods for Sparse Signal Recovery Using Scale Mixtures," *IEEE Trans. On Sig. Process.*, Jul. 2016.

[163] S. Theodoridis, "Machine Learning, a Bayesian and Optimization Perpsective," in *Elsevier*, 2015.

[164] X. Tan and J. Li, "Computationally Efficient Sparse Bayesian Learning via Belief Propagation ," *IEEE Trans. on Sig. Proc.*, vol. 58, no. 4, Apr. 2013.

[165] D. Zachariah and P. Stoica, "Online Hyperparameter-Free Sparse Estimation Method," *IEEE Trans. On Sig. Process.*, vol. 63, no. 13, July 2015.

[166] M. Al-Shoukairi, P. Schniter, and B. D. Rao, "GAMP-based low complexity sparse bayesian learning algorithm," *IEEE Trans. on Sig. Process.*, vol. 66, no. 2, January 2018.

[167] H. Duan, L. Yang, and H. Li, "Fast inverse-free sparse bayesian learning via relaxed evidence lower bound maximization," *IEEE Sig. Process. Letters*, vol. 24, no. 6, June 2017.

[168] C. K. Thomas and D. Slock, "SAVE - space alternating variational estimation for sparse Bayesian learning," in *Data Science Workshop*, 2018.

[169] J. M. Ortega, "Numerical Analysis: A Second Course. ," in *SIAM*, Philadelphia, 1990.

[170] J. K. Johnson *et al.*, "Fixing Convergence of Gaussian Belief Propagation ," in *IEEE Intl. Symp. on Info. Theo.*, 2009.

[171] A. Kammoun, A. Muller, E. Bjornson, and M. Debbah, "Linear precoding based on polynomial expansion: Large-scale multi-cell MIMO systems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 861–875, 2014.

[172] T. Sadiki and D. Slock, ""Bayesian Adaptive Filtering: Principles and Practical Approaches"," in *Proc. 12th European Signal Processing Conf. (EUSIPCO)*, Vienna, Austria, Sept. 2004.

[173] M. Lenardi and D. Slock, "Estimation of Time-Varying Wireless Channels and Application to the UMTS W-CDMA FDD Downlink," in *Proc. European Wireless (EW)*, Florence, Italy, Feb. 2002.

[174] Consortium Partners, "Deliverable D2.3 "Hybrid Localization Techniques"," EC FP7 project WHERE, Tech. Rep., May 2010, URL: http://www.kns.dlr.de/where/public_documents_deliverables.php.

[175] S. Bensaid, A. Schutz, and D. Slock, "Single Microphone Blind Audio Source Separation Using EM-Kalman Filter and Short+Long Term AR Modeling," in *Proc. Int'l Conf. on Latent Var. Analy. and Sig. Sep. (LVA-ICA)*, Saint-Malo, France, Sept. 2010.

[176] R. Herbrich, "Minimising the Kullback-Leibler Divergence," in *Microsoft Research*, August 2015.

[177] B. Ait-El-Fquih and I. Hoteit, "Fast Kalman-like filtering for large-dimensional linear and Gaussian state-space models," *IEEE Trans. on Sig. Process.*, vol. 63, no. 21, Nov. 2015.

[178] S. Bensaid and D. Slock, "Comparison of Various Approaches for Joint Wiener/Kalman Filtering and Parameter Estimation with Application to BASS," in *IEEE 45th Asilomar Conference on Sig., Sys. and Comp.*, Pacific Grove, CA, USA, 2011.

[179] E. Riegler *et al.*, "Merging Belief Propagation and the Mean Field Approximation: a Free Energy Approach ," *IEEE Trans. on Info. Theo.*, vol. 59, no. 1, Jan. 2013.

[180] J. J. Boutros and G. Caire, "Iterative multiuser joint decoding: Unified framework and asymptotic analysis," *IEEE Trans. on Info. Theo.*, 2012.

[181] D. Wipf, B. D. Rao, and S. Nagarajan, "Latent variable bayesian models for promoting sparsity," *IEEE Trans. on Info. Theo.*, Sept. 2011.

[182] C. K. Thomas and D. Slock, "Generalized swept approximate message passing based kalman filtering for dynamic sparse Bayesian learning," in *EUSIPCO*, Amsterdam, Netherlands, 2020.

[183] J. S. Yedidia *et al.*, "Constructing free-energy approximations and generalized belief propagation algorithms ," *IEEE Trans. on Info. Theo.*, vol. 51, no. 7, June 2005.

[184] C. K. Thomas and D. Slock, "Space alternating variational Bayesian learning for LMMSE filtering," in *EUSIPCO*, 2018.

[185] T. Sadiki and D. T. Slock, "Bayesian adaptive filtering: principles and practical approaches," in *EUSIPCO*, 2004.

[186] B. H. Fleury, M. Tschudin, R. Heddergott, D. Dahlhaus, and K. I. Pedersen, "Channel Parameter Estimation in Mobile Radio Environments Using the SAGE Algorithm ," *IEEE J. on Sel. Areas in Commun.*, vol. 17, no. 3, pp. 434–450, March 1999.

[187] C. K. Thomas and D. Slock, "Gaussian Variational Bayes Kalman Filtering for Dynamic Sparse Bayesian Learning," in *ITISE*, 2018.

[188] T. Minka, "A family of algorithms for approximate bayesian inference," in *Ph.D. dissertation, Mass. Inst. Technol., Cambridge, MA, USA*, 2001.

[189] S. Fortunati *et al.*, "Performance bounds for parameter estimation under misspecified models," *IEEE Sig. Proc. Mag.*, Nov. 2017.

[190] R. E. Kass and A. E. Raftery, "Bayes factors ," *J. Am. Stat. Assoc.*, vol. 90, 1995.

[191] P. Tichavský, C. H. Muravchik, and A. Nehorai, "Posterior Cramer Rao Bounds for Discrete-Time Nonlinear Filtering," *IEEE Trans. On Sig. Process.*, vol. 46, May 1998.

[192] A. E. Gelfand and S. K. Sahu, "Identifiability, Improper Priors, and Gibbs Sampling for Generalized Linear Models," *Journ. of the Americ. Stat. Assoc.*, Mar. 1999.

[193] J. Du *et al.*, "Convergence Analysis of Distributed Inference with Vector-Valued Gaussian Belief Propagation," *Jrnl. of Mach. Learn. Res.*, April 2018.

[194] D. M. Malioutov, J. K. Johnson, and A. S. Willsky, "Walk-Sums and Belief Propagation in Gaussian Graphical Models ," *Jrnl. of Mach. Learn. Res.*, Oct. 2006.

[195] S. Rangan *et al.*, "On the Convergence of Approximate Message Passing With Arbitrary Matrices," *IEEE Trans. on Info. Theo.*, Sept. 2019.

[196] M. Bayati and A. Montanari, "The Dynamics of Message Passing on Dense Graphs, with Applications to Compressed Sensing ," *IEEE Trans. on Inf. Theory*, vol. 57, no. 2, pp. 764–785, February 2011.

[197] J. A. Fessler and A. Hero, "Space-alternating generalized expectation-maximization algorithm." *IEEE Trans. on Sig. Process.*, Oct. 1994.

[198] C. K. Thomas and D. Slock, "Low Complexity Static and Dynamic Sparse Bayesian Learning Combining BP, VB and EP Message Passing," in *Asilomar Conf. on Sig., Sys., and Comp.*, CA, USA, 2019.

[199] ——, "Space Alternating Variational Estimation and Kronecker Structured Dictionary Learning," in *ICASSP*, 2019.

[200] ——, "SAVED - Space Alternating Variational Estimation for Sparse Bayesian Learning with Parametric Dictionaries," in *Asilomar Conf. on Sig., Sys., and Comp.*, 2018.

[201] D. L. Donoho, A. Javanmard, and A. Montanari, "Information-theoretically optimal compressed sensing via spatial coupling and approximate message passing," in *IEEE Intl. Symp. on Info. Theo.*, 2012.

[202] F. Caltagirone, F. Krzakala, and L. Zdeborová, "On the Convergence of Approximate Message Passing," in *IEEE Intl. Symp. on Info. Theo.*, 2014.

[203] J. Vila, P. Schniter, S. Rangan, F. Krzakala, and L. Zdeborová, "Adaptive damping and mean removal for the generalized approximate message passing algorithm," in *In IEEE Intl. Conf. on Acous., Spee. and Sig. Process. (ICASSP)*, April 2015.

[204] A. Manoel, F. Krzakala, E. W. Tramel, and L. Zdeborová, "Swept Approximate Message Passing for Sparse Estimation," in *Proc. of the 32nd Intl. Conf. on Mach. Learn.*, 2015.

[205] F. Krzakala, A. Manoel, E. W. Tramel, and L. Zdeborova, "Variational Free Energies for Compressed Sensing," in *IEEE Intl. Sympo. Info. Theo.*, Honolulu, HI, USA, Jun. 2014.

[206] C. K. Thomas and D. Slock, "Nonlinear MMSE using Linear MMSE Bricks and Application to Compressed Sensing and Adaptive Kalman Filtering ," *Special Issue in Communications in Information and Systems, International Press*, 2020.

[207] D. Nion and N. D. Sidiropoulos, "Tensor algebra and multidimensional harmonic retrieval in signal processing for MIMO radar ," *IEEE Trans. on Sig. Process.*, vol. 58, no. 11, Nov. 2010.

[208] R. A. Harshman, "Foundations of the PARAFAC procedure: Models and conditions for an "explanatory" multi-modal factor analysis," in *UCLA Working Papers in Phonetics, Available at http://publish.uwo.ca/~harshman/wpppfac0.pdf.*, 1970.

[209] L. R. Tucker, "Implications of factor analysis of three-way matrices for measurement of change ," *Probl. Meas. Change*, 1963.

[210] M. Sørensen, L. D. Lathauwer, P. Comon, S. Icart, and L. Deneire, "Canonical polyadic decomposition with a columnwise orthonormal factor matrix ," *SIAM J. Matrix Anal. Appl.*, vol. 33, no. 4, 2010.

[211] M. Haardt, F. Roemer, and G. D. Galdo, "Higher-order SVD based subspace estimation to improve the parameter estimation accuracy in multi-dimensional harmonic retrieval problems," *IEEE Trans. on Sig. Process.*, vol. 56, pp. 3198–3213, Jul. 2008.

[212] M. F. Duarte and R. G. Baraniuk, "Kronecker compressive sensing ," *Proceedings of the IEEE*, vol. 21, no. 2, Feb. 2012.

[213] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications ," *SIAM Review*, vol. 51, no. 2, Aug. 2009.

[214] M. S. Lewicki and T. J. Sejnowski, "Learning overcomplete respresentations ," *Neural Computation*, vol. 12, no. 2, 2000.

[215] K. Skretting and K. Engang, "Recursive least squares dictionary learning algorithm ," *IEEE Trans. on Sig. Process.*, vol. 58, Apr. 2010.

[216] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation ," *IEEE Trans. on Sig. Process.*, vol. 54, no. 11, Nov. 2006.

[217] F. Roemer, G. D. Galdo, and M. Haardt, "Tensor-based algorithms for learning multidimensional separable dictionaries," in *IEEE Intl. Conf. on Acous., Speech and Sig. process. (ICASSP)*, 2014.

[218] X. Ding, W. Chen, and I. J. Wassell, "Joint sensing matrix and sparsifying dictionary optimization for tensor compressive sensing. ," *IEEE Trans. on Sig. Process.*, vol. 65, no. 4, Jul. 2017.

[219] C. K. Thomas and D. Slock, "BP-VB-EP based static and dynamic sparse Bayesian learning with kronecker structured dictionaries," in *IEEE Intl. Conf. on Acous. Spee. and Sig. Process. (ICASSP)*, May 2020.

[220] ——, "Space alternating variational Bayesian learning for LMMSE filtering," in *EUSIPCO*, 2018.

[221] ——, "Gaussian variational Bayes Kalman filtering for dynamic sparse Bayesian learning," in *Intl. Conf. on Time Ser. and Forec.*, 2018.

[222] N. D. Sidiropoulos *et al.*, "Tensor decomposition for signal processing and machine learning ," *IEEE Trans. on Sig. Process.*, vol. 65, no. 13, July 2017.

[223] C. K. Thomas and D. Slock, "Variational Bayesian Learning for Channel Estimation and Transceiver Determination," in *Info. Theo. and Appl. Wkshp*, San Diego, USA, February 2018.

[224] C. Qian, X. Fu, N. D. Sidiropoulos, and Y. Yang, "Tensor based parameter estimation of double directional massive MIMO channel with dual- polarized antennas," in *Proc. IEEE Int'l Conf. Acoustics Speech and Sig. Proc. (ICASSP)*, 2018.

[225] Z. Shakeri, A. D. Sarwate, and W. U. Bajwa, "Identifiability of Kronecker-structured dictionaries for tensor data ," *IEEE Journ. Sel. Top. in Sig. Process.*, vol. 12, no. 5, Oct. 2018.

[226] A. K. Gupta and D. K. Nagar, "Matrix variate distributions," in *Boca Raton FL, USA: CRC Press*, 1999.

[227] K. P. Murphy, "Machine learning: A probabilistic perspective," in *MA, USA: MIT press*, 2012.

[228] M. Boizard, R. Boyer, G. Favier, and P. Comon, "Performance estimation for tensor CP decomposition with structured factors," in *IEEE Intl. Conf. on Acous., Speech and Sig. Process. (ICASSP)*, Brisbane, Australia, 2015.

[229] M. Sorensen and L. D. Lathauwer, "Blind Signal Separation via Tensor Decomposition With Vandermonde Factor: Canonical Polyadic Decomposition ," *IEEE Trans. on Sig. Process.*, vol. 61, no. 22, 2013.

[230] Z. Shakeri, W. U. Bajwa, and A. D. Sarwate, "Minimax lower bounds on dictionary learning for tensor data ," *IEEE Trans. Info. Theory*, vol. 64, no. 4, Apr. 2018.

[231] C. D. Richmond and L. L. Horowitz, "Parameter bounds on estimation accuracy under model misspecification ," *IEEE Trans. on Sig. Process.*, vol. 63, no. 9, May. 2015.

[232] F. Roemer and M. Haardt, "A semi-algebraic framework for approximate CP decompositions via Simultaneous Matrix Diagonalizations (SECSI)," *Else. Sig. Process.*, vol. 93, pp. 2722–2738, Sept. 2013.

[233] C. K. Thomas and D. Slock, "Sparse Bayesian Learning for a Bilinear Calibration Model and Mismatched CRB," in *IEEE EUSIPCO*, Sept. 2019.

[234] S. Fortunati, F. Gini, M. Greco, and C. Richmond, "Performance Bounds for Parameter Estimation under Misspecified Models [Fundamental findings and applications]," *IEEE Sig. Proc. Mag.*, Nov. 2017.

[235] C. Shepard, H. Yuand, N. Anand, E. Li, T. Marzetta, R. Yang, and L. Zhong, "Argos: Practical many-antenna base stations," in *ACM Intern. Conf. Mobile Computing and Netw. (Mobicom)*, Istanbul, Turkey, Aug. 2012.

[236] X. Jiang, A. Decurninge, K. Gopala, F. Kaltenberger, M. Guillaud, D. Slock, and L. Deneire, "A Framework for Over-the-air Reciprocity Calibration for TDD Massive MIMO Systems," *IEEE Trans. Wireless Commun.*, Sept. 2018.

[237] K. Gopala and D. Slock, "Optimal algorithms and CRB for reciprocity calibration in massive MIMO," in *IEEE ICASSP*, Apr. 2018.

[238] E. de Carvalho and D. Slock, "Semi–Blind Methods for FIR Multichannel Estimation," in *IEEE SPAWC*, Apr. 2000.

[239] E. de Carvalho, S. Omar, and D. Slock, "Performance and complexity analysis of blind FIR channel identification algorithms based on deterministic maximum likelihood in SIMO systems," *Cir., Sys., and Sig. Process.*, vol. 34, no. 4, Aug. 2012.

[240] E. de Carvalho and D. Slock, "Blind and semi-blind FIR multichannel estimation: (Global) identifiability conditions," *IEEE Trans. Sig. Proc.*, Apr. 2004.

[241] ——, "Cramér-Rao bounds for blind multichannel estimation," *arXiv:1710.01605 [cs.IT]*, 2017.

[242] E. de Carvalho, J. M. Cioffi, and D. Slock, "Cramér-Rao bounds for blind multichannel estimation," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, CA, USA, Nov. 2020.

[243] V. Smídl and A. Quinn, "The variational Bayes method in signal processing." in *New York: Springer-Verlag*, 2005.

[244] M. Opper and O. Winther, "Expectation Consistent Approximate Inference," *J. Mach. Learn. Res.*, vol. 6, Dec. 2005.

[245] J. Vieira, F. Rusek, O. Edfors, S. Malkowsky, L. Liu, and F. Tufvesson, "Reciprocity Calibration for Massive MIMO: Proposal, Modeling and Validation," *IEEE Trans. On Wire. Commun.*, May. 2017.

[246] J. M. Kantor, C. D. Richmond, B. C. Jr., and D. W. Bliss, "Prior Mismatch in Bayesian Direction of Arrival Estimation for Sparse Arrays," in *IEEE Proc. RadarCon*, 2015.

[247] W. James and C. M. Stein, "Estimation with Quadratic Loss," in *Breakthroughs in Statistics, Springer, New York, NY*, 1992.

[248] R. Rogalin, O. Y. Bursalioglu, H. Papadopoulos, G. Caire, A. F. Molisch, A. Michaloliakos, V. Balan, and K. P. and, "Scalable synchronization and reciprocity calibration for distributed multiuser MIMO," *IEEE Trans. Wire. Commun.*, vol. 13, no. 4, Apr. 2014.

[249] A. L. Yuille and A. Rangarajan., "The concave-convex procedure." *Neural computation*, vol. 15, no. 4, pp. 915–936, 2003.

[250] K.Gregor and Y. LeCun, "Learning fast approximations of sparse coding," in *Proc. Int. Conf. Mach. Learn.*, 2010.

[251] M. Borgerding, P. Schniter, and S. Rangan, "Amp-inspired deep networks for sparse linear inverse problems," *IEEE Transactions on Signal Processing*, vol. 65, no. 16, Aug. 2014.