# Block-sparse models in multi-modality : application to the inverse model in EEG/MEG

Fardin Afdideh

UNIVERSITÉ GRENOBLE ALPES

**THÈSE**

pour obtenir le grade de

**DOCTEUR DE L'UNIVERSITÉ DE GRENOBLE ALPES**

Spécialité : **Signal, Image, Parole, Télécoms (SIPT)**

Arrêté ministériel : 7 août 2006

Présentée par
**Fardin AFDIDEH**

Thèse dirigée par **Christian JUTTEN** et
co-encadrée par **Ronald PHLYPO**

préparée au sein du
**Grenoble Images Parole Signal Automatique laboratoire (GIPSA-Lab)**
dans **Ecole Doctorale d'Electronique, Electrotechnique, Automatique, Traitement du Signal (EEATS)**

# Block-sparse models in multi-modality: application to the inverse problem in EEG/MEG

Thèse soutenue publiquement le **12/10/2018**,
devant le jury composé de:

**Maureen CLERC**
Directeur de Recherche, INRIA, Rapporteur
**Charles SOUSSEN**
Professeur, CentraleSupélec, Rapporteur
**Saïd MOUSSAOUI**
Professeur, Ecole Centrale de Nantes, Examinateur
**David BRIE**
Professeur, Université de Lorraine, Examinateur, Président du jury
**Christian JUTTEN**
Professeur, Université de Grenoble Alpes, Directeur de thèse
**Ronald PHLYPO**
Maître de Conférences, Grenoble-INP, Co-encadrant de thèse

UNIVERSITÉ DE GRENOBLE ALPES

**Ecole Doctorale d'Electronique, Electrotechnique, Automatique, Traitement du Signal (EEATS)**

# T H È S E

pour obtenir le titre de

**docteur en sciences**

de l'Université de Grenoble Alpes

**Mention : Signal, Image, Parole, Télécoms (SIPT)**

Présentée et soutenue par

Fardin AFDIDEH

## Block-sparse models in multi-modality: application to the inverse problem in EEG/MEG

Thèse dirigée par Christian JUTTEN

préparée au Grenoble Images Parole Signal Automatique laboratoire (GIPSA-Lab)

soutenue le 12/10/2018

**Jury :**

| | | | |
|---|---|---|---|
| *Rapporteurs :* | Maureen CLERC | - | Directeur de Recherche, INRIA |
| | Charles SOUSSEN | - | Professeur, CentraleSupélec |
| *Directeur :* | Christian JUTTEN | - | Professeur, Université de Grenoble Alpes |
| *Co-encadrant :* | Ronald PHLYPO | - | Maître de Conférences, Grenoble-INP |
| *Président :* | David BRIE | - | Professeur, Université de Lorraine |
| *Examinateur :* | Saïd MOUSSAOUI | - | Professeur, Ecole Centrale de Nantes |

Communauté
UNIVERSITÉ Grenoble Alpes

# THÈSE

Pour obtenir le grade de

## DOCTEUR DE LA COMMUNAUTÉ UNIVERSITÉ GRENOBLE ALPES

Spécialité : SIGNAL IMAGE PAROLE TELECOMS

Arrêté ministériel : 25 mai 2016

Présentée par


## Fardin AFDIDEH

Thèse dirigée par **Christian JUTTEN**, Professeur, UGA
et codirigée par **Ronald Phlypo**

préparée au sein du **Laboratoire Grenoble Images Parole Signal Automatique**
dans **l'École Doctorale Electronique, Electrotechnique, Automatique, Traitement du Signal (EEATS)**


# Des modèles bloc-parcimonieux en multi-modalité : application au problème inverse en EEG/MEG

# Block-sparse Models for Multimodal Detection with Application to EEG/MEG Inverse Model

Thèse soutenue publiquement le **12 octobre 2018**,
devant le jury composé de :

**Monsieur CHRISTIAN JUTTEN**
PROFESSEUR, UNIVERSITE GRENOBLE ALPES, Directeur de thèse
**Madame MAUREEN CLERC**
DIRECTRICE DE RECHERCHE, INRIA CENTRE S. ANTIPOLIS - MEDITERRANEE, Rapporteur
**Monsieur CHARLES SOUSSEN**
PROFESSEUR, CENTRALE SUPELEC, Rapporteur
**Monsieur SAID MOUSSAOUI**
PROFESSEUR, ECOLE CENTRALE NANTES, Examinateur
**Monsieur DAVID BRIE**
PROFESSEUR, UNIVERSITE DE LORRAINE, Président
**Monsieur RONALD PHLYPO**
MAITRE DE CONFERENCES, GRENOBLE INP, Co-directeur de thèse

# Acknowledgements

I would like to express my sincere gratitude to my advisors, Prof. Christian JUTTEN and Ronald PHLYPO, for their continuous support of my Ph.D research, who their door of office was always open whenever I had a question about my research, in addition to their guidances in our regular meetings, which steered me in the right research direction.

This work would have been impossible without the financial support of Prof. Christian JUTTEN as the principal investigator of the granted fund of the European Research Council (ERC) for the research project of the Challenges in the Extraction and Separation of Sources (CHESS).

I am also grateful to each of the jury members, Prof. Maureen CLERC, Charles SOUSSEN, David BRIE, and Saïd MOUSSAOUI, for their very valuable comments on this thesis. Especially, I would like to thank Prof. Charles SOUSSEN who his invaluable and numerous comments had a significant impact on the quality of the thesis.

Moreover, many thanks go out to my post-doc supervisors in Lyon Neuroscience Research Center (CRNL), Mathilde BONNEFOND, Olivier BERTRAND, and Jérémie MATTOUT, for their patience and support I received during the extended months before the Ph.D defense.

It is also worth mentioning that my migration to France in order to undertake and start this Ph.D, which has been a truly life-changing experience for me, would definitely not have been feasible without the full support of Prof. Mohammad Bagher SHAMSOLLAHI (my Master's supervisor in Sharif University of Technology) and Christian JUTTEN, which makes me profoundly and life-lastingly grateful to them.

At last, my deep appreciation goes out to my family members, my parents, Aliakbar and Fahimeh, and my siblings, Farnaz and Pedram, for always believing in me and encouraging me to follow my dreams. And to my spiritual teachers especially Lida, for enormously helping me to first have the courage to embark on this journey, then confronting the consequent challenges, and finally to complete what I started.




# Contents

# List of Figures

# List of Tables

# Table of acronyms

- Multi-modality

  **EEG** *ElectroEncephaloGraphy*
  **MEG** *MagnetoEncephaloGraphy*
  **MRI** *Magnetic Resonance Imaging*

- (Block-)sparsity

  **(Block-)ERC** *(Block-sparse) Exact Recovery Condition(s)*
  **(Block-)MCC** *(Block) Mutual Coherence Constant*
  **(Block-)NSP** *(Block) Null Space Property*
  **(Block-)SL** *(Block-)Sparsity Level*
  **USLE** *Underdetermined System(s) of Linear Equations*

# Mathematical notations and nomenclature

**Mathematical notations** Throughout the manuscript, we represent matrices, vectors and scalars by boldface uppercase characters (e.g., $\boldsymbol{A}$), boldface lowercase italic characters (e.g., $\boldsymbol{a}$), and lowercase italic characters (e.g., $a$), respectively.

The mathematical notations used in the manuscript are shown in table 1. To better formulate the block-wise indexing, let $\mathbb{P}$ be a partitioning of indices of elements in vector $\boldsymbol{a}$ or indices of columns in matrix $\boldsymbol{A}$. Then suppose $\mathbb{P}_i$ indicates the $i^{th}$ partition and $\mathbb{P}_i(j)$ indexes the $j^{th}$ element in $\mathbb{P}_i$.

As an example of block-wise indexing, take vector $\boldsymbol{b}\in\mathbb{R}^5$, matrix $\boldsymbol{A}\in\mathbb{R}^{2\times5}$, and the partitioning $\mathbb{P}=\{\{1,2\},\{3,4\},\{5\}\}$. Then, $\boldsymbol{b}[2]$ is concatenation of the elements $b_3$ and $b_4$, whereas $\boldsymbol{A}[2]$ is concatenation of the columns $\boldsymbol{a}_3$ and $\boldsymbol{a}_4$, and $\boldsymbol{a}_1[2]$ is the column $\boldsymbol{a}_3$.

| | |
|---|---|
| Operators | |
| $\boldsymbol{A}^T$ | Transpose of matrix $\boldsymbol{A}$ |
| $\boldsymbol{A}^\dagger$ | Moore-Penrose pseudo-inverse of matrix $\boldsymbol{A}$ |
| $\boldsymbol{I}_d$ | Identity matrix in $\mathbb{R}^{d\times d}$ |
| $\boldsymbol{1}_{d_1\times d_2}$ | Matrix of ones in $\mathbb{R}^{d_1\times d_2}$ |
| Element-wise indexing | |
| $a_i$ | $i^{th}$ element in vector $\boldsymbol{a}$ |
| $\boldsymbol{a}_i$ | $i^{th}$ column in matrix $\boldsymbol{A}$ |
| $a_{i,j}$ | The element in the $i^{th}$ row and $j^{th}$ column in the matrix $\boldsymbol{A}$ |
| Block-wise indexing | |
| $\mathbb{P}$ | Partitioning of indices |
| $\boldsymbol{a}[i]$ | $i^{th}$ block of elements in vector $\boldsymbol{a}$, i.e., $\boldsymbol{a}[i]=[a_k \mid k\in\mathbb{P}_i]$ |
| $\boldsymbol{A}[i]$ | $i^{th}$ block of columns in matrix $\boldsymbol{A}$, i.e., $\boldsymbol{A}[i]=[\boldsymbol{a}_k \mid k\in\mathbb{P}_i]$ |
| $\boldsymbol{a}_j[i]$ | $j^{th}$ column in block $\boldsymbol{A}[i]$, i.e., $\boldsymbol{a}_j[i]=[\boldsymbol{a}_k \mid k\in\mathbb{P}_i(j)]$ |

Table 1: The mathematical notations used throughout the manuscript.

**Nomenclature** The nomenclature used throughout the manuscript can be divided into three categories as shown in table 2.

| | |
|---|---|
| | Structure and model |
| $\boldsymbol{y}$ | Measurement vector in $\mathbb{R}^m$ |
| $\boldsymbol{\Phi}$ | Dictionary (lead-field) matrix in $\mathbb{R}^{m\times n}$ |
| $\boldsymbol{\beta_0}$ | True representation vector in $\mathbb{R}^n$ |
| $\hat{\boldsymbol{\beta}}$ | Estimated representation vector |
| $d_k$ | Cardinality of the $k^{th}$ partition |
| $K$ | Number of blocks |
| $[\boldsymbol{A},\boldsymbol{B}]$ | Concatenation of matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ |
| $\langle \boldsymbol{a},\boldsymbol{b}\rangle$ | Inner product of vectors $\boldsymbol{a}$ and $\boldsymbol{b}$, i.e., $\boldsymbol{a}^T\boldsymbol{b}$ |
| | Optimisation problem |
| $\left\|\boldsymbol{a}\right\|_p$ | $\ell_p$ norm of vector $\boldsymbol{a}$ |
| $P_{p,q,\varepsilon}\left(\boldsymbol{a},\boldsymbol{b}\right)$ | $\min_{\boldsymbol{a}} \left\|\boldsymbol{a}\right\|_p \quad s.t. \quad \left\|\boldsymbol{b}\right\|_q \leq \varepsilon$ |
| $P_p$ | $\min_{\boldsymbol{\beta}} \left\|\boldsymbol{\beta}\right\|_p \quad s.t. \quad \boldsymbol{y}=\boldsymbol{\Phi\beta}$ |
| $\left\|\boldsymbol{a}\right\|_{p_1,p_2}$ | $\ell_{p_1,p_2}$ (pseudo-)mixed-norm of vector $\boldsymbol{a}$ |
| $P_{(p_1,p_2),q,\varepsilon}\left(\boldsymbol{a},\boldsymbol{b}\right)$ | $\min_{\boldsymbol{a}} \left\|\boldsymbol{a}\right\|_{p_1,p_2} \quad s.t. \quad \left\|\boldsymbol{b}\right\|_q \leq \varepsilon$ |
| $P_{p_1,p_2}$ | $\min_{\boldsymbol{\beta}} \left\|\boldsymbol{\beta}\right\|_{p_1,p_2} \quad s.t. \quad \boldsymbol{y}=\boldsymbol{\Phi\beta}$ |
| $\left\|\boldsymbol{a}\right\|_{\boldsymbol{w};p_1,p_2}$ | $\ell^{\boldsymbol{w}}_{p_1,p_2}$ weighted (pseudo-)mixed-norm of vector $\boldsymbol{a}$ |
| $P_{(\boldsymbol{w};p_1,p_2),q,\varepsilon}\left(\boldsymbol{a},\boldsymbol{b}\right)$ | $\min_{\boldsymbol{a}} \left\|\boldsymbol{a}\right\|_{\boldsymbol{w};p_1,p_2} \quad s.t. \quad \left\|\boldsymbol{b}\right\|_q \leq \varepsilon$ |
| $P_{\boldsymbol{w};p_1,p_2}$ | $\min_{\boldsymbol{\beta}} \left\|\boldsymbol{\beta}\right\|_{\boldsymbol{w};p_1,p_2} \quad s.t. \quad \boldsymbol{y}=\boldsymbol{\Phi\beta}$ |
| | Characterisations |
| $I\left(a\right)$ | Indicator function of scalar $a$, i.e., $I(a)\overset{\text{def}}{=}\begin{cases}1, & \text{if } \lvert a\rvert>0\\ 0, & \text{if } a=0\end{cases}$ |
| $\left\|\boldsymbol{A}\right\|_{q\to p}$ | $\ell_{q\to p}$ operator-norm of matrix $\boldsymbol{A}$, i.e., $\left\|\boldsymbol{A}\right\|_{q\to p}\overset{\text{def}}{=}\max_{\left\|\boldsymbol{a}\right\|_q\leq 1}\left\|\boldsymbol{Aa}\right\|_p$ |
| $S\left(\boldsymbol{\beta}\right)$ | Support of vector $\boldsymbol{\beta}$ |
| $S_b\left(\boldsymbol{\beta}\right)$ | Block support of vector $\boldsymbol{\beta}$ |
| $\left\lvert S\left(\boldsymbol{\beta}\right)\right\rvert$ | Cardinality of $S\left(\boldsymbol{\beta}\right)$ |
| $\boldsymbol{G}(\boldsymbol{\Phi})$ | Gram matrix of $\boldsymbol{\Phi}$ |
| $Ker(\boldsymbol{\Phi})$ | Kernel of $\boldsymbol{\Phi}$, i.e., $Ker(\boldsymbol{\Phi})\overset{\text{def}}{=}\{\boldsymbol{x}\in\mathbb{R}^n,\ \boldsymbol{\Phi x}=\boldsymbol{0}\}$ |
| $Spark(\boldsymbol{\Phi})$ | Spark of $\boldsymbol{\Phi}$, i.e., $Spark(\boldsymbol{\Phi})\overset{\text{def}}{=}\min_{\boldsymbol{x}\in Ker(\boldsymbol{\Phi})\setminus\{\boldsymbol{0}\}}\left\|\boldsymbol{x}\right\|_0$ |
| $Block{-}Spark(\boldsymbol{\Phi})$ | Block-Spark of $\boldsymbol{\Phi}$ |
| $\overline{M}\left(\boldsymbol{\Phi_1},\boldsymbol{\Phi_2}\right)$ | Basic mutual coherence constant of $\boldsymbol{\Phi_1}$ and $\boldsymbol{\Phi_2}$ |
| $M\left(\boldsymbol{\Phi}\right)$ | Mutual coherence constant of dictionary $\boldsymbol{\Phi}$ |
| $M\left(\boldsymbol{\Phi},k\right)$ | Cumulative mutual coherence constant of dictionary $\boldsymbol{\Phi}$ |
| $M_p\left(\boldsymbol{\Phi},k\right)$ | $\ell_p$-coherence function |
| $\overline{M}_{q,p}\left(\boldsymbol{\Phi_1},\boldsymbol{\Phi_2}\right)$ | $(q,p)$-basic block mutual coherence constant of $\boldsymbol{\Phi_1}$ and $\boldsymbol{\Phi_2}$ |
| $M_{q,p}\left(\boldsymbol{\Phi}\right)$ | $(q,p)$-block mutual coherence constant of dictionary $\boldsymbol{\Phi}$ |
| $M_{q,p}\left(\boldsymbol{\Phi},k\right)$ | $(q,p)$-cumulative block mutual coherence constant of dictionary $\boldsymbol{\Phi}$ |
| $\overline{M}^{Eldar}_{Inter}\left(\boldsymbol{\Phi_1},\boldsymbol{\Phi_2}\right)$ | Eldar et al.'s basic block-coherence constant of $\boldsymbol{\Phi_1}$ and $\boldsymbol{\Phi_2}$ |
| $M^{Eldar}_{Intra}\left(\boldsymbol{\Phi}\right)$ | Eldar et al.'s intra-block coherence constant of dictionary $\boldsymbol{\Phi}$ |
| $M^{Eldar}_{Inter}\left(\boldsymbol{\Phi}\right)$ | Eldar et al.'s inter-block coherence constant of dictionary $\boldsymbol{\Phi}$ |
| $M^{Eldar}_{Inter}\left(\boldsymbol{\Phi},k\right)$ | Cumulative inter-block coherence constant of dictionary $\boldsymbol{\Phi}$ |
| $Q_p\left(S_b\left(\boldsymbol{\beta}\right),\boldsymbol{\Phi}\right)$ | Characterisation of null space property |
| $Q_{p_1,p_2}\left(S_b\left(\boldsymbol{\beta}\right),\boldsymbol{\Phi}\right)$ | Characterisation of block null space property |
| $Q_{\boldsymbol{w};p_1,p_2}\left(S_b\left(\boldsymbol{\beta}\right),\boldsymbol{\Phi}\right)$ | Weighted characterisation of block null space property |

Table 2: The nomenclature used throughout the manuscript.

# Introduction

Many problems in different areas of engineering and science, such as inverse problems, can be reformulated as an underdetermined system of linear equations, i.e., the number of equations is less than the number of unknowns. In order to retain an appropriate unique solution from infinitely many solutions to a such system, problem-related constraints need to be applied. According to Tropp, for more than one century *sparsity* constraints have been studied and applied to numerous applications [Tro04b].

However, if a comprehensive measurement is not provided, attaining a desired solution will not be feasible, even considering optimal constraints. This is mainly due to the rich characteristics of the phenomenon of interest that makes it impossible for a single modality to project all aspects of the phenomenon. Therefore, the main research question of this dissertation is:

What is the added value of multi-modality, when solving inverse problems?

To answer the above-mentioned question, three main challenges will raise. First of all, it should be noticed that the desirable sparse solution can be extracted from a set of possible solutions through a proper constrained optimisation problem.

In order to ensure the uniqueness of the sparsest solution of the optimisation problem, some theoretical recovery conditions are proposed in literature. These conditions are based on the amount of similarity or coherence between the columns of coefficient matrix of the inverse problem. More precisely, the less coherence, the more relaxed are the conditions under which recovery is guaranteed to be successful.

For *vastly* underdetermined systems of linear equations, the columns of coefficient matrix are more likely to be coherent. This gives rise to the following first challenge:

**Challenge 1.** Many real-world inverse problems are vastly underdetermined, and classical sparse estimation techniques do no longer give acceptable recovery conditions. How can high-dimensional problems be adapted in favour of the coherence-based notion of conventional conditions?

Secondly, due to the inverse impact of the coherence on the recovery conditions, low coherence is favourable. Hence, the idea of clustering the coherent parts of the coefficient matrix seems to be promising. Although the first challenge has a rather straightforward approach, but this goes paired with the following technical challenge:

**Challenge 2.** In classical assumption, the recovery conditions are generated from the columns of coefficient matrix, but clustered coefficient matrix consists of some differently-sized blocks, and not necessarily columns. Therefore, the initial assumption of classical recovery

conditions does no longer hold true. How can appropriate recovery conditions be developed for block-structured problems?

Finally, assuming that the two mentioned challenges are addressed successfully, and returning to the main research question, the last challenge is related to the impact of multi-modality:

**Challenge 3.** Is joining multiple modalities always beneficial, knowing that each modality provides us with different properties of the same phenomenon? How can the added value of multi-modality be demonstrated?

In order to explore and address the three aforementioned challenges, this dissertation is organized in three main parts. Since Challenge 1 uses the results of Challenge 2, first Challenge 2 is presented in the dissertation, then Challenge 1, and ultimately in the last part a partial answer to Challenge 3 is provided through experiments.

The first part of this thesis offers general theoretical recovery conditions based on *block-sparsity* constraint, which ensure the uniqueness of the block-sparse solution of corresponding weighted (pseudo-)mixed-norm optimisation problem in an underdetermined system of linear equations. This part responds to Challenge 2.

The second part of the thesis suggests the clustering idea to improve the general theoretical recovery conditions proposed in the first part, which responds to Challenge 1.

Consequently, the third part of the thesis investigates Challenge 3, while considering the constraint in Challenge 2, i.e., block-sparsity. Our main application problem is distributed EEG/MEG source reconstruction.

Given a linear operator relating the current density to the observed potential/magnetic fields, which is called lead-field, the inverse problem in EEG/MEG source reconstruction solves a vastly underdetermined system of linear equations.

In EEG/MEG source reconstruction problem, two concepts are integrated: block-sparsity, and multi-modality. In the following, we elucidate the existence of the two mentioned concepts. There is the notion of block activity, because the activity of each brain source is related to a current dipole which can be represented by three magnitudes in $x$, $y$, and $z$ directions, respectively, and can be modelled by a vector of dimension three. Hence, the source vector will be a concatenation of groups or blocks of length three. Moreover, usually for a given brain task, a few regions of brain are activated, then it is consistent with the notion of block-sparsity.

Due to the fact that EEG and MEG measure electric and magnetic properties, respectively, they are complementary. Additionally, EEG and MEG measure properties of the same neuronal activity, induced by same current dipoles. Then, EEG and MEG information can be combined to form a multi-modal problem.

Accordingly, the real-world distributed EEG/MEG source reconstruction problem completely complies with the conditions of the three above-mentioned challenges, and will

be used as a leading example.

Considering the properties of the real-world distributed EEG/MEG source reconstruction problem, which require few number of active blocks in the source vector, in Chapter 2 a general framework for block-sparse recovery problem is proposed. In other words, this chapter gives a partial answer to "In an underdetermined system of linear equations, under which conditions a unique block-sparse solution can be recovered?".

In fact, in a wide range of problems, the value of a single coefficient inside a block of coefficients, whether zero or non-zero, is not important, but the status of the whole block which is determined by all of its coefficients is the atomic meaningful entity. Therefore, instead of the sparsity constraint which penalizes the non-zero coefficients, a block of coefficients should be penalized.

Moreover, in contrast to the conventional point of view, where, the number of non-zero values of coefficients determines the unique solution, in the block-structured problems the status of block of coefficients (active or inactive) has an effective role. In addition, depending on the definition of a block, an active block can also have zero coefficients.

The mentioned generality of the framework of Chapter 2 is in terms of the properties of the underdetermined system of linear equations, extracted characterisations, optimisation problems, and ultimately the recovery conditions. This chapter offers four main groups of recovery conditions based on:

1. Block-Spark,
2. block null space property,
3. block mutual coherence constant, and
4. cumulative coherence constant.

It is worth mentioning that utilising the block structure, besides the conformity with circumstances of some of the real-world problems, has the advantage of weakening the conventional sparse recovery conditions. Then, the benefit of block-sparsity assumption over conventional sparsity is proved theoretically, while it is shown that due to the natural generalisation, all the proposed materials reduce to the conventional findings for the unit block length. Finally, we prove the improvement of the proposed conditions over other recently proposed block-sparse recovery conditions.

As mentioned implicitly, the lead-field is a fat matrix. For instance, in an ordinary EEG/MEG source reconstruction problem with 30 sensors and 3000 sources, the lead-field would be a matrix of dimensions 30 by 9000 (3000 sources times 3 values in $x$, $y$, and $z$ directions).

Since the mentioned block mutual coherence constant introduced in Chapter 2 is a coherence characterisation extracted from the lead-field matrix, and due to the fact that

the low coherence is favourable, the idea of clustering coherent parts of lead-field is studied in Chapter 3. The clustering algorithm is applied in a hierarchical manner which enables us to estimate the number of clusters according to the resulted clustering tree.

To construct the similarity matrix in hierarchical clustering algorithm, the proposed coherence characterisation introduced in Chapter 2 is used. By applying clustering on the lead-field matrix we will have two main achievements:

1. improved block-sparse recovery conditions, and
2. brain source space segmentation.

To explain more about the brain source space segmentation, it should be mentioned that each consecutive block of three columns of the lead-field matrix corresponds to a single brain source, so by clustering the blocks of the lead-field matrix we are actually clustering brain sources.

It should be highlighted that in contrast to other existing brain source space segmentation scenarios, in this study this segmentation is done in the most general case, i.e., the clustering algorithm is blind to any information about the brain sources activity and EEG/MEG signals. In other words, the segmentation is not restricted to a special brain activity.

In Chapter 4, the goal is to show the effect of combining the information of two complementary modalities of EEG and MEG. We show that in multi-modal case more refined and precise brain regions appear. Hence, EEG and MEG multi-modality advantage will be proved.

Extensive experiments on synthetic and real data indicate a significant improvement over the state-of-the-art findings in our three main research orientations.

Chapter 1

# Background material

Contents



## 1.1 Introduction

This chapter presents a background on the methods to face two challenges of multi-modality, in Section 1.2, and block-sparsity. In order to review the recent studies on block-sparsity, we start from the conventional sparsity constraint in finding a unique solution of an USLE[1] in Section 1.3. Then, a review of recent studies on block-sparsity is presented in Section 1.4. At last, the role of block-sparsity in multi-modality is discussed in Section 1.5.

[1] *Underdetermined System(s) of Linear Equations*

## 1.2 Multi-modality

### 1.2.1 Introduction

Many natural phenomena consist of very complex structural and functional properties. For instance, considering the sophisticated example of human brain as the phenomenon, the electric, magnetic, functional and anatomical properties can be taken into account.

On the other hand, each modality or measurement instrument measures only a specific property of the phenomenon. Then, due to the rich characteristics of a natural multi-property phenomenon, a comprehensive information cannot be obtained by only a single modality. Therefore, the research domain of *multi-modality* revealed to jointly analyse the data collected from different modalities which are somehow complementary. The concept of multi-modality is shown in figure 1.1 in a simple language. As shown in figure 1.1, a natural multi-property phenomenon is represented by a multi-coloured object, where, each colour represents a specific property of the object. In figure 1.1, modalities are been represented by eyes that are observing the multispectral object. Then, the fact of shortcoming of a single modality to observe all the

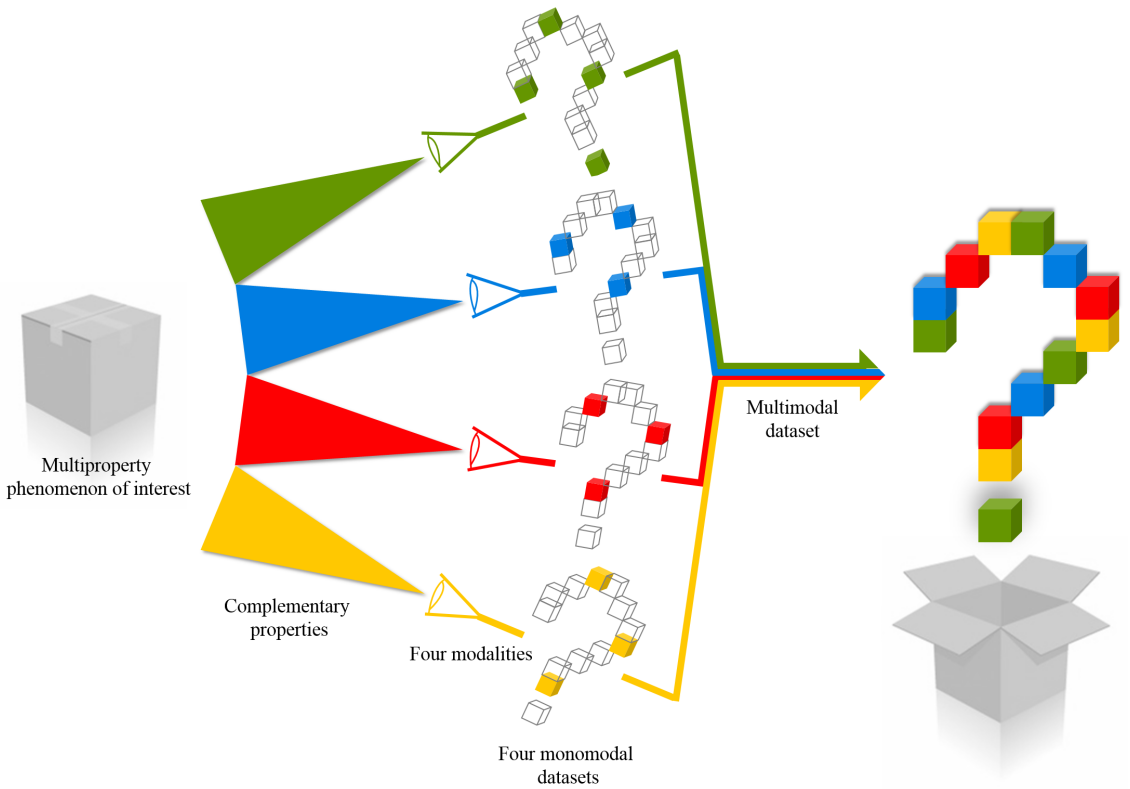



Figure 1.1: Mono-modal dataset cannot describe the multispectral phenomenon of interest, due to the monospectral detectability of each modality. While by combining mono-modal datasets, an acceptable description of the phenomenon can be recovered.

properties of the phenomenon is represented by the blindness of each of the eyes to a range of colours. Finally, the data observed from each of the eyes can be combined to build a correct image of the multi-coloured object of interest, which is the ultimate goal in multi-modality, i.e., compensating each modality's relative shortcomings by integrating other modality's dataset, in order to extract the maximum available information. In multi-modality, the combined dataset is called *multi-modal dataset* in contrast to the dataset of a single modality which is called mono-modal or uni-modal dataset. The interest in the multi-modality originates from the idea that the information in the multi-modal dataset is more than the sum of the information in each of individual mono-modal datasets.

In our study, the multi-property phenomenon of interest is the human brain activity and we are interested in its electromagnetic properties as represented in figure 1.2. In neurophysiology, a wide variety of modalities are used for bioelectric measurements, among which the EEG is the most interesting from the clinical applications point of view, thanks to the following four main advantages: (1) non-invasive, (2) high temporal resolution (in order of milliseconds), (3) relatively cost-effective, and (4) easy to use and portable.

In addition, for biomagnetic measurements the MEG is commonly used. As represented graphically in figure 1.2, fusion of EEG and MEG can be used to improve the quality of source estimation, due to the fact that multi-modality can also reduce the ill-posedness of problems and make them better determined [Vel+16]. In the Section 1.2.2, we review the electrophysiological properties of neuronal populations. In Section 1.2.3, the required steps to solve the EEG/MEG source reconstruction problem is discussed. Finally, Section 1.2.6 elucidates some complementarities of EEG and MEG.

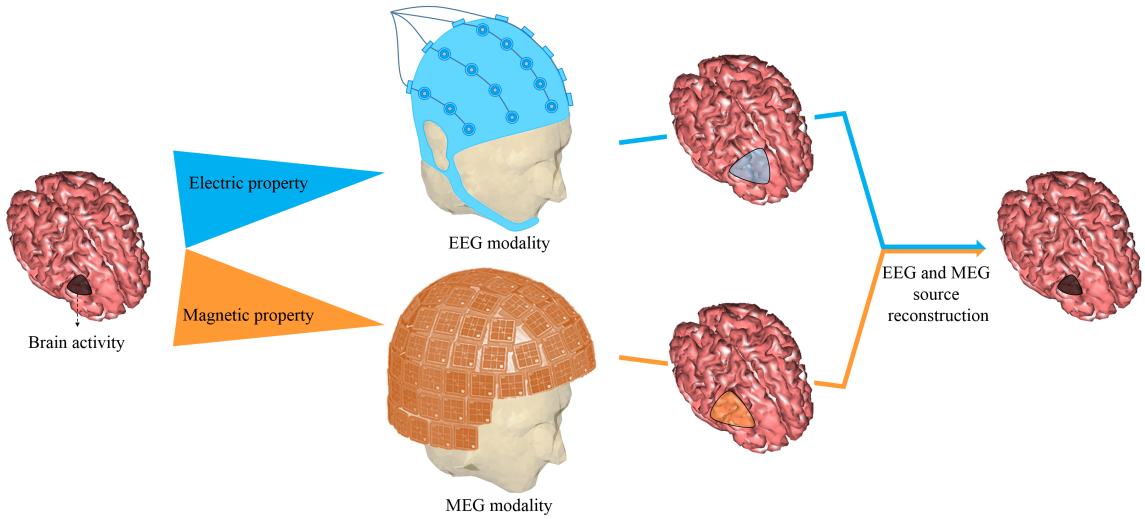



Figure 1.2: Electromagnetic properties of brain activity is measured by EEG and MEG modalities and can be used to reconstruct the brain activity. Source reconstruction using EEG and MEG multi-modality leads to improved results.

### 1.2.2 Physiological basis of EEG and MEG

The main direct neuronal source of EEG and MEG signals is pyramidal neuron. Task of a pyramidal neuron is to take synaptic inputs and produce patterned output of action potential [Oka93]. As can be seen in figure 1.3, the postsynaptic dendrite part is polarized due to the concentration discrepancy of ions. When the action current reaches the presynaptic axon terminal, the glands of neurotransmitters are released into the synapse and bind to the receptors on the postsynaptic dendrite, and this leads to a flow of ions of $Na^+$ from outside of cell to inside, and producing post synaptic current as shown in figure 1.3. As shown in figure 1.3, by spreading this ions, intracellular current (primary current) and extracellular current (secondary current) are produced. A small patch of brain cortex which comprises thousands of these simultaneously activated parallel pyramidal neurons can be represented by a current dipole [CH03].This relationship between the macroscopic representation and the corresponding microscopic neuronal activity is important [OWK97]. These pyramidal neurons are oriented perpendicular to the brain cortex, but since the cortex is folded, 'radial', 'tangential', and 'oblique' dipole orientations defined with respect to the local curvature of the skull, must be considered.

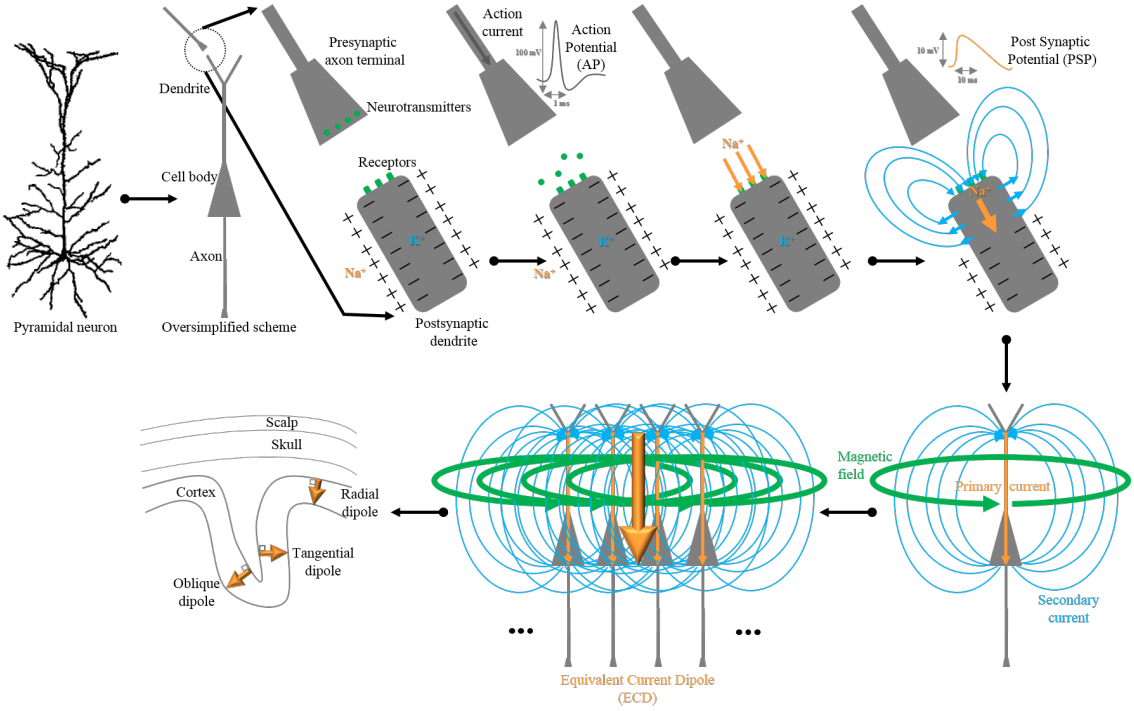



Figure 1.3: Action current leads to releasing neurotransmitters into the synapse and producing post synaptic current inside the cell. A current dipole is a representation of thousands of simultaneously firing neighbouring neurons and can have different orientations.

### 1.2.3 EEG/MEG forward and inverse problems

The goal in EEG/MEG source analysis is finding the appropriate sources based on the electric/magnetic measurements, which is also called the *inverse problem*. It is called an inverse problem because it exploits the information at the results stage and then leads to the causes.

Inverse problems are not specific to the EEG/MEG source analysis research field and are considered as one of the most important mathematical problems in science. The importance of the inverse problems is due to the fact that by solving the problems inversely, it would be possible to discover the latent parameters, which are not measurable directly.

In order to solve inverse problem, one must first solve the so-called *forward problem*, i.e., what would be the electromagnetic fields of a known source? This is the inverse of a inverse problem, which exploits the information of the causes and then leads to the results.

In figure 1.4, the forward problem is to determine EEG/MEG measurements in sensor space resulted from three active dipoles in source space, whereas, the inverse problem is to estimate the activity of dipoles in source space given EEG/MEG measurements in sensor space.

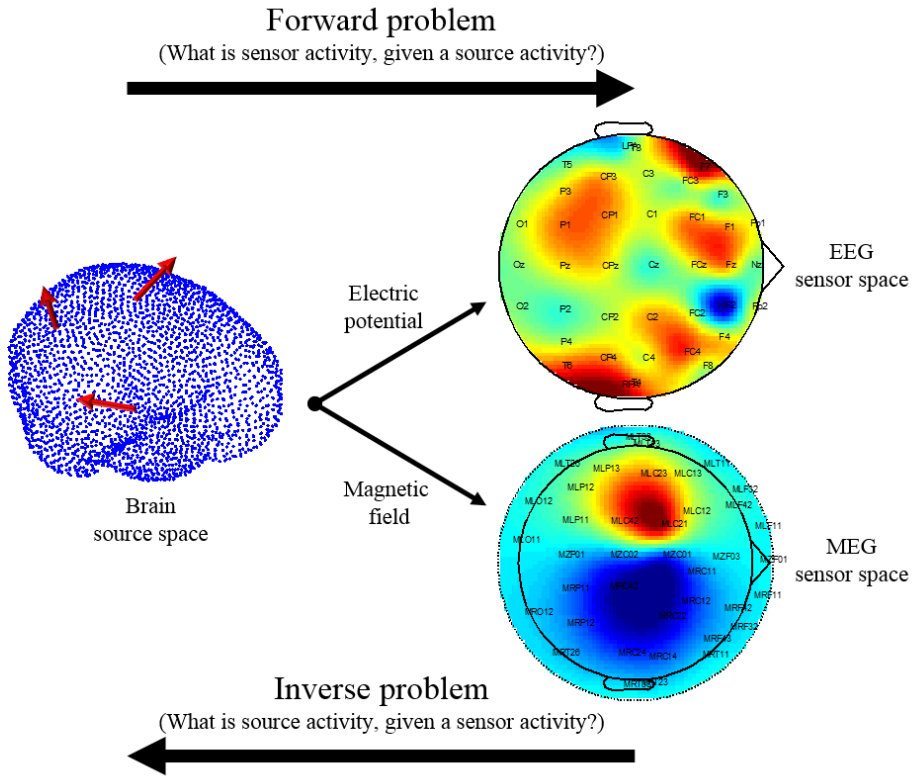



Figure 1.4: The forward and inverse problems in EEG/MEG source analysis.

### 1.2.4 Volume conduction head model

In bioelectromagnetism, the transmission of electromagnetic fields from a current dipole through the head tissues towards measurement sensors is called volume conduction head model, which contains the electrical conduction properties of the head. In frequencies below 1 kHz, the quasi-static approximation of Maxwell's equations can be used for modelling the volume conduction head model [PH67]. Therefore, the level of realism of the volume conduction head model determines the quality of EEG/MEG source analysis [Hau+14].

The volume conduction head model can be modelled analytically, or numerically. In other words, for the EEG/MEG, the head can be modelled analytically by a single sphere or three or even four concentric spheres corresponding to brain, cerebrospinal fluid, skull, and scalp [Hos+78]. The boundary element methods and finite element methods have been developed to numerically model the head and to better represent the realistic shape of the head [YNH91]; [AAG04]. Although utilising numerically-modelled head models lead to increased accuracy in the source localization problem, it is computationally heavy. Therefore, based on the requirements of the problem a trade-off between computational burden and source localization accuracy should be taken into account.

In figure 1.5, the three-layer analytical and numerical volume conduction head models are shown. The three layers correspond to scalp, skull, and brain, respectively. The head model in figure 1.5(a) is computed analytically, whereas the ones in (b) and (c) are computed numerically. The realistic brain layer in figure 1.5(b) is an inflated cortical sheet, whereas in figure 1.5(c) it is a highly-folded cortical sheet.

Naturally, in order to build realistic head models we need to structural information of head, which is provided by MRI[2].

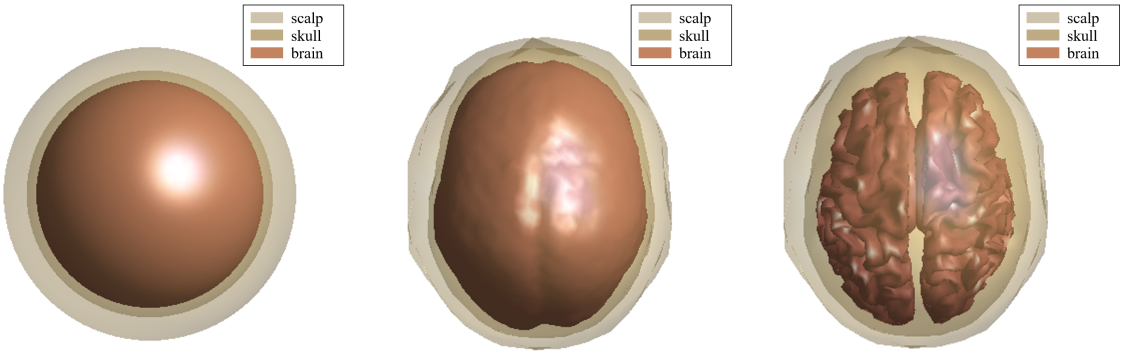



Figure 1.5: Three-layer (scalp, skull, and brain) head models modelled (a) analytically, and (b,c) numerically. In (b), the cortical sheet is inflated, whereas in (c) it is highly-folded.

[2] *Magnetic Resonance Imaging*

### 1.2.5 Lead-field

The transmission of electromagnetic fields from sources space through head tissues towards sensors space can be modelled with quasi-static approximations of Maxwell's equations [Sar87].

The quasi-static approximations of Maxwell's equations, which allows to ignore the time derivatives, can be divided into two groups of:

- quasi-electrostatics, which are describing electric field $\boldsymbol{E}$, and
- quasi-magnetostatics, which are describing magnetic field $\boldsymbol{B}$.

The mentioned quasi-static approximations with electric potential $\boldsymbol{V}$ and magnetic field $\boldsymbol{B}$ in distance $\boldsymbol{r}$ in an infinite homogeneous medium with permittivity $\varepsilon_0$ and permeability $\mu_0$ for a dipole $\boldsymbol{p}$ with electric charge density $\rho$ and electric current density $\boldsymbol{J}$ are shown in figure 1.6, and table 1.1.

The quasi-static approximations of Maxwell's equations in addition to the assumed discrete positions of the brain sources lead to a discretised and linearised relationship between electromagnetic signals and the source activities. The mentioned relationship is realised through the so-called *lead-field* matrix which contains the electromagnetic and geometrical properties of the head.

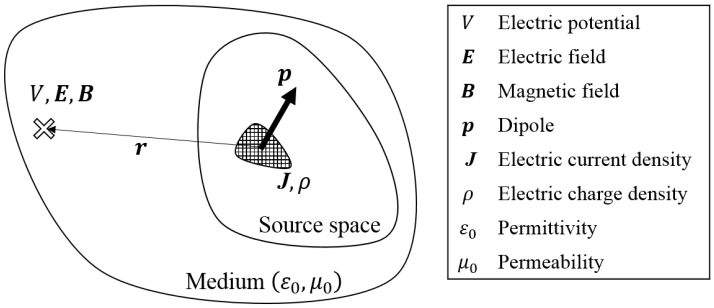



Figure 1.6: Electromagnetic fields ($\boldsymbol{E}$, $\boldsymbol{B}$) at distance $\boldsymbol{r}$ of dipole $\boldsymbol{p}$.

| Quasi-electrostatics | | Quasi-magnetostatics | |
|---|---|---|---|
| Gauss's law | $\nabla \cdot \boldsymbol{E} = \frac{\rho}{\varepsilon_0}$ | Gauss's law | $\nabla \cdot \boldsymbol{B} = 0$ |
| Faraday's law | $\nabla \times \boldsymbol{E} = \boldsymbol{0}$ | Ampère's law | $\nabla \times \boldsymbol{B} = \mu_0 \boldsymbol{J}$ |
| $V = \frac{1}{4\pi\varepsilon_0} \boldsymbol{p} \cdot \frac{\boldsymbol{r}}{\lvert\boldsymbol{r}\rvert^3}$ | | $\boldsymbol{B} = \frac{\mu_0}{4\pi} \boldsymbol{p} \times \frac{\boldsymbol{r}}{\lvert\boldsymbol{r}\rvert^3}$ | |

where, $\nabla\cdot$, $\nabla\times$, $\cdot$, and $\times$ are divergence, curl, dot and cross product operators, respectively.

Table 1.1: The quasi-static approximations of Maxwell's equations.

The lead-field matrix, which is also considered as the solution of the forward problem, is computed when the volume conduction head model, source model, and sensor model is given.

As shown in figure 1.7, knowing the volume conduction head model including scalp, skull, and brain models with $K$ source positions in source space and $m$ sensors in sensor space makes possible the computation of the lead-field matrix of dimension $m$ by $3K$.

Each block of *three columns* in lead-field matrix, consists of the response of all sensors to a specific probing dipole in the source space. Actually, it is a block of three canonical probing vectors.

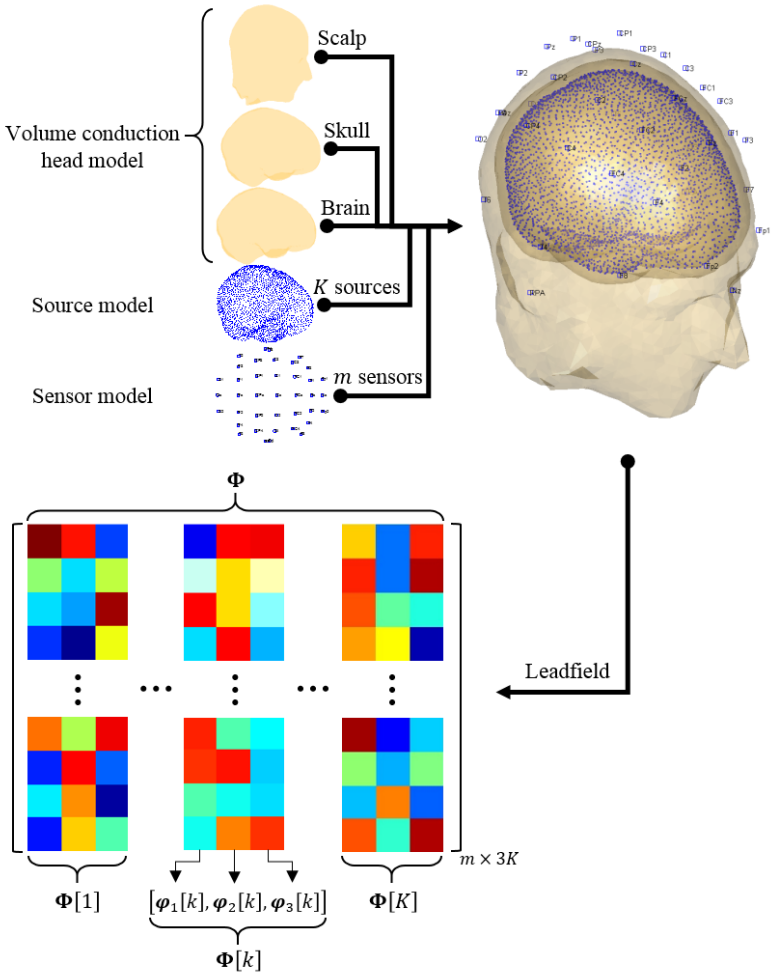



Figure 1.7: Forward problem is solved when the head (scalp, skull, and brain layers), source ($K$ source positions), and sensor ($m$ sensors) models are given. The lead-field matrix $\mathbf{\Phi}$ of dimension $m$ by $3K$ is resulted from concatenation of $K$ three-column matrices corresponding to $K$ source positions in source space.

### 1.2.6 EEG and MEG complementarities

In recent years, the partial independence of bioelectric and biomagnetic measurements has been studied and the added value of their combination under a multi-modal signal processing framework has been the subject of some researches. In this section, some relative merits of cerebral bioelectromagnetic measurements, i.e., EEG and MEG, are reviewed.

- **Dimension of measurement.** MEG measurement is vectorial, whereas EEG measurement is scalar.
- **Reference measurement.** MEG measures magnetic field, whereas EEG measures potential *differences*. In other words, the existence of reference electrode is essential in EEG contrary to MEG (figure 1.8(a)).
- **Sensor position.** MEG sensors are located outside but near the head, whereas EEG sensors are placed on the surface of the head. Therefore, MEG makes a contact-less recording indicating that it has the advantage of very low pre-recording preparation time and the disadvantage of necessity of existence of helmet for localization of the head position (figure 1.8(b)).
- **Sensitivity to cerebral currents.** MEG is mostly sensitive to primary currents, whereas EEG is sensitive to secondary currents (figure 1.8(c)). The cerebral currents were briefly discussed in Section 1.2.2.
- **Sensitivity to orientation of dipole.** MEG has very low sensitivity to radial dipoles and is zero for spherical head model [Ahl+10] (figure 1.8(d)).
- **Sensitivity to depth of dipole.** MEG has low sensitivity to deep sources, because deeper sources become more quasi-radial (figure 1.8(e)).
- **Sensitivity to head tissues.** In contrast to MEG, EEG is highly sensitive to the geometry and conductivity of media [AG03]; [GA04]. Therefore, EEG is distorted while passing through the brain tissues, especially the skull because of its low conductivity, i.e., high resistivity.

  On the other hand, MEG is far less sensitive to internal heterogeneities [HP17]. For instance in figure 1.8(f), for a same source model, which is realistic highly-folded cortical sheet, MEG is quite equal for the two head models, while EEG is significantly different.

  Because of the transparency of the skull to magnetic fields, MEG is able to measure the cerebral activity of smaller brain regions [Mal11].
- **Topography.** EEG and MEG topographies for a same source activity are almost orthogonal to each other (figure 1.8(g)).

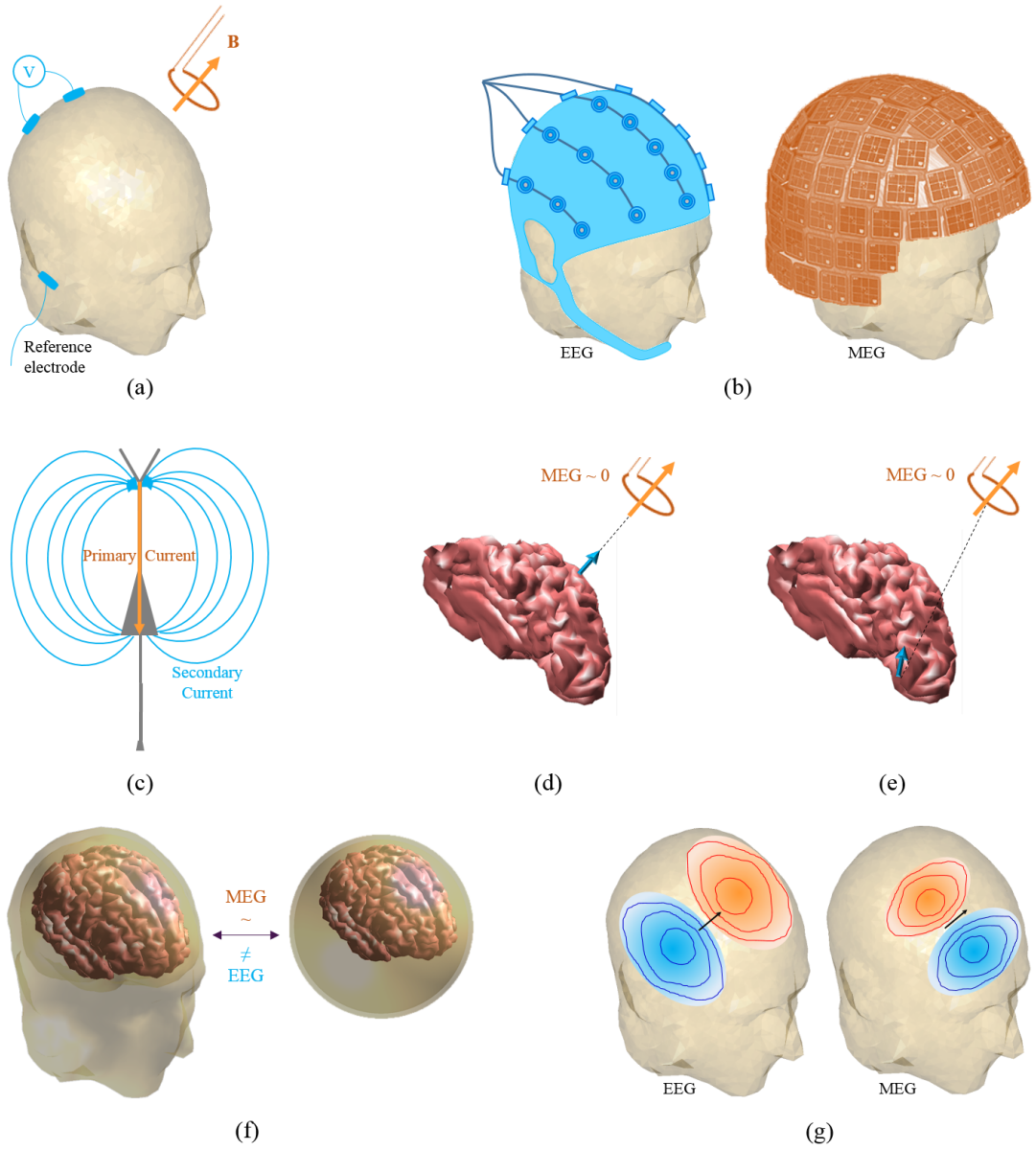



Figure 1.8: (a) MEG measures vectorial magnetic field, whereas EEG measures scalar potential differences, (b) MEG sensors are located outside the head, whereas EEG sensors are placed on the surface of the head, (c) MEG is mostly sensitive to primary currents, whereas EEG is sensitive to secondary currents, (d) MEG has very low sensitivity to radial dipoles, (e) MEG has low sensitivity to deep sources, (f) EEG is highly sensitive to the geometry and conductivity of media, and (g) EEG and MEG topographies for a same source activity are almost orthogonal to each other. .

## 1.3 Sparsity

### 1.3.1 Introduction

In science and technology, e.g., statistics [CT07], or biomedical engineering [GH04]; [ES05]; [LDP07], scientists and engineers end up with vastly USLE, which have an infinite number of solutions, if any. For instance, the research domain of inverse problem mainly deals with USLE. Because of these infinitely many solutions, the problem is said ill-posed [Had02]. According to some *a priori* knowledge about the nature of the data of interest, and consequently the solution, this eligible infinite number of solutions, which results in ambiguity in solution space, could be restricted to a smaller class of solutions or, pragmatically, to a unique solution, which is a good approximation to a true solution.

Commonly, the amount of *sparsity* of the true solution has been exploited as additional constraint to disappear the mentioned ill-posedness of the USLE Sparsity assumption is a generic constraint that its application in different domains can lead to solutions with different properties, e.g., smoothness of the solution. In this work, the sparsity assumption is utilised. In fact, the sparsity assumption is a very relevant constraint due to the practical observation indicating that many real world signals have approximately sparse representations.

If the coefficient matrix of USLE is union of orthonormal bases, the solution is *efficient*, where efficiency in solution means requiring very few significant coefficients in comparison to the dimension of the solution, so sparsity [CDS01]; [GN03b]. But practically it has been observed that because of the rich characteristics of the natural phenomena, if the coefficient matrix is a single orthonormal basis, usually the natural images or sounds do not necessarily have sparse solution [GN03a]. Therefore, the more generalised concept of coefficient matrix is introduced which is the so-called *dictionary* [MZ93]. Each column of a dictionary is called *atom*. For explanation of the atom-dictionary terminology, the interested reader is referred to [MZ93] and [CDS01]. The coefficient matrix appears in different names of design matrix, measurement matrix, coding matrix, codebook, or dictionary, based on its nature and the research community.

In the community of overcomplete signal representations, finding the sparse solution of a vastly USLE is translated as recovering a high-dimensional ideally sparse vector, which can be a digital signal or image. In addition, it can be translated as an efficient representation of a signal in an overcomplete dictionary. Overcompleteness of the dictionary implies that the linear system of equations is underdetermined, because the number of atoms in the dictionary is more than the number of entries in each of the atoms. However, despite of the fact that the problem is underdetermined, there are some heuristic and theoretical arguments on benefits of overcompleteness in theoretical neuroscience [OF97], approximation theory [CD02], signal processing [DCL97]; [BM99]; [CR02], and image processing [Huo99]; [SCD02]; [SED03].

Section 1.3.2 defines the sparse representation theory, next in Section 1.3.3, recovery conditions are reviewed.

### 1.3.2 Sparse representation theory

To elucidate the theory, this section is divided into three parts: (1) structure and model, (2) optimisation problem, and (3) characterisations:

**1) Structure and model** In general, the USLE can be formulated as [CT05]; [DET06]:

$$\boldsymbol{y} = \boldsymbol{\Phi}\boldsymbol{\beta_0} + \boldsymbol{e}, \tag{1.1}$$

where, $\boldsymbol{y}\in\mathbb{R}^m$ is the measurement vector, $\boldsymbol{\Phi}\in\mathbb{R}^{m\times n}$, $m<n$, is the dictionary, $\boldsymbol{\beta_0}\in\mathbb{R}^n$ is the true representation vector, or input variable, and $\boldsymbol{e}\in\mathbb{R}^m$ is noise, bounded by a known noise level, e.g., $\|\boldsymbol{e}\|_2<\epsilon$. if $\boldsymbol{e}=\boldsymbol{0}$, then the model is noiseless [DE03b]; [GN03b]. In this study, we will concentrate on the following noiseless model:

$$\boldsymbol{y} = \boldsymbol{\Phi}\boldsymbol{\beta_0}. \tag{1.2}$$

As mentioned before, the goal is to extract a representation vector $\hat{\boldsymbol{\beta}}$ which is the sparsest among all solutions, i.e., a representation with the fewest non-zero elements relative to its dimension. For example, suppose the following estimated representation vector $\hat{\boldsymbol{\beta}}$:

$$\hat{\boldsymbol{\beta}} = \left[\hat{\beta}_1, \cdots, \hat{\beta}_k, \cdots, \hat{\beta}_n\right]^T,$$

and following dictionary:

$$\boldsymbol{\Phi} = \left[\boldsymbol{\varphi}_1, \cdots, \boldsymbol{\varphi}_k, \cdots, \boldsymbol{\varphi}_n\right],$$

with $\boldsymbol{\varphi}_j\in\mathbb{R}^m$ and without loss of generality, it is assumed that atoms have unit Euclidean norm, i.e., $\forall j, \|\boldsymbol{\varphi}_j\|_2=1$. Then, the sparsity structure of the noiseless USLE is graphically represented in figure 1.9. Commonly, this problem is referred to as *atomic decomposition*, since the measurement vector $\boldsymbol{y}$ is decomposed into its fewest active atoms of the dictionary [CDS01]; [DH01]; [DE03a].

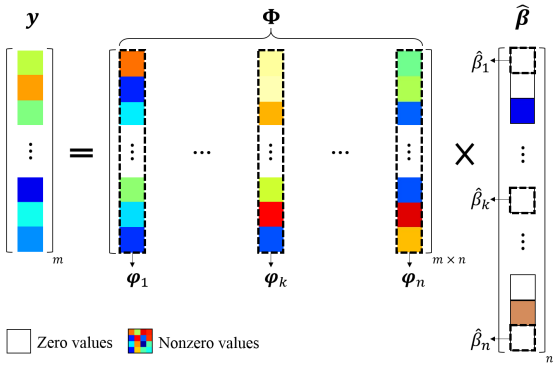



Figure 1.9: Sparsity structure of the noiseless USLE.

**2) Optimisation problem** The problem of finding a sparse solution $\hat{\boldsymbol{\beta}}$ in the noiseless case is called *exact (unique) signal recovery* whereas for the problem in the noisy case, it is called *stable (robust) signal recovery.*

In general, exact signal recovery provides deep insight into the problem of stable signal recovery [CT05]; [CT07]; [Don04b]; [Don04a]. Exact signal recovery remains invariant by scaling of the dictionary, $\boldsymbol{\Phi}\rightarrow k\boldsymbol{\Phi}$, whereas stable signal recovery has a direct relation with scaling of the dictionary [DG08]; [FL09].

Stablity in stable signal recovery means that bounded changes in the measurement vector $\boldsymbol{y}$ result in bounded changes in the estimated representation vector $\hat{\boldsymbol{\beta}}$ [CRT06]. Based on the amount of the bounded change either small or arbitrary, there are two main types of stability named local stability and global stability, respectively [DET06].

The best case in stable signal recovery is obtained when the size of the representation error is guaranteed to be bounded by the noise level, i.e., $\|\hat{\boldsymbol{\beta}}-\boldsymbol{\beta_0}\|_2<\epsilon$. In a real-world situation, usually the measurement is severely incomplete and also inaccurate and when the measurement vector is noisy and the true representation vector is not exactly sparse, the measurement is said to be imperfect [CRT06].

Norms play an essential role in quantifying distances between two vectors or amplitude of a single vector in the definition of cost functions for optimisation problems [Wal14]. Exact and stable signal recovery problems can be realised through a proper constrained optimisation problem. The straightforward optimisation problem for stable signal recovery is:

$$\min_{\boldsymbol{\beta}} \|\boldsymbol{\beta}\|_0 \qquad s.t. \qquad \|\boldsymbol{y}-\boldsymbol{\Phi\beta}\|_2 \leq \varepsilon, \tag{1.3}$$

where, for $\varepsilon$=0, the optimisation problem (1.3) solves an exact signal recovery problem. In a general framework in order to cover all the previous types of optimisation problems, consider the following *constrained $\ell_p$ norm minimisation problem*:

$$P_{p,q,\varepsilon}(\boldsymbol{a},\boldsymbol{b}) : \min_{\boldsymbol{a}} \|\boldsymbol{a}\|_p \qquad s.t. \qquad \|\boldsymbol{b}\|_q \leq \varepsilon, \tag{1.4}$$

where, $\boldsymbol{a}$ and $\boldsymbol{b}$ are vectors, and $\ell_p$ norm of vector $\boldsymbol{a}$ is defined:

$$\|\boldsymbol{a}\|_p \stackrel{\text{def}}{=} \begin{cases} \sum_i I(a_i), & \text{for } p=0 \\ \left(\sum_i |a_i|^p\right)^{\frac{1}{p}}, & \text{for } 0<p<+\infty \\ \max_i\{|a_i|\}, & \text{for } p=+\infty, \end{cases}$$

where, $I(\cdot)$ is the indicator function, i.e., $I(a) \stackrel{\text{def}}{=} \begin{cases} 1, & \text{if } |a|>0 \\ 0, & \text{if } a=0 \end{cases}$. The noise level $\epsilon$, which upper-bounds the Euclidean norm of noise $\boldsymbol{e}$ in the general model of (1.1) is bounded such that $0\leq\epsilon\leq\varepsilon$. Therefore, according to (1.4) the problem (1.3) can be represented as $P_{0,2,\varepsilon}(\boldsymbol{\beta},\boldsymbol{y}-\boldsymbol{\Phi\beta})$.

By defining $\boldsymbol{r}(\boldsymbol{\beta})=\boldsymbol{y}-\boldsymbol{\Phi\beta}$ as a *remainder* vector, it is assumed in this work as a consequence that $\|\boldsymbol{r}(\boldsymbol{\beta})\|_q=0$ is equivalent to $\boldsymbol{y}=\boldsymbol{\Phi\beta}$. For the sake of simplicity we refer to $\boldsymbol{r}(\boldsymbol{\beta})$ as $\boldsymbol{r}$. The estimated solution of the problem $P_{p,q,\varepsilon}(\boldsymbol{a},\boldsymbol{b})$ is called $\hat{\boldsymbol{\beta}}$. In the introduced general optimisation problem, the widely used case is when the minimisation is over $\boldsymbol{\beta}$ and the problem is an exact signal recovery, i.e., $P_{p,q,0}(\boldsymbol{\beta},\boldsymbol{r})$, which for the sake of simplicity we refer to it as $P_p$. Because regardless of choice of $q$, the $\|\boldsymbol{r}\|_q=0$ directly indicates to the noiseless model, so the only parameter to be determined is $p$.

A norm function $f$ has three following properties [GL13]; [Wal14]:

1. $f(\boldsymbol{a}) \geq 0$ (nonnegativity),
2. $f(\alpha\boldsymbol{a}) = |\alpha|f(\boldsymbol{a})$ (positive homogeneity or scalability),
3. $f(\boldsymbol{a} \pm \boldsymbol{b}) \leq f(\boldsymbol{a}) + f(\boldsymbol{b})$ (triangle inequality).

Some of the most important norms that their minimum induces sparsity in exact and stable signal recovery problems are:

- $\ell_0$ pseudo-norm (or semi-norm, or Hamming norm [HR09]),
- $\ell_1$ norm (or taxicab norm, or Manhattan norm, or grid norm [Wal14]),
- $\ell_2$ norm (or Euclidean norm), and
- $\ell_\infty$ norm (or maximum norm, or Chebyshev norm, or uniform norm).

**Remark 1.1.** Notice that, $\|\boldsymbol{a}\|_p$ for $p=0$ does not satisfy the required positive homogeneity property of a norm function, i.e., $\|\alpha\boldsymbol{a}\|_0\neq|\alpha|\|\boldsymbol{a}\|_0$. Therefore $\|\boldsymbol{a}\|_0$ is commonly referred to as a $\ell_0$ pseudo-norm instead of $\ell_0$ norm. $\|\boldsymbol{a}\|_0$ simply counts the number of non-zero elements in $\boldsymbol{a}$.

**Remark 1.2.** It should be noticed that, $\|\boldsymbol{a}\|_p$ for $0<p<1$ does not satisfy the required triangle inequality property of a norm function. Therefore $\|\boldsymbol{a}\|_p$ for $0<p<1$ is commonly referred to as a $\ell_p$ pseudo-norm instead of $\ell_p$ norm.

For a vector $\boldsymbol{\beta}$, the basic sparsity measure is the $\ell_0$ pseudo-norm [HR09], which is the number of non-zero entries of the vector $\boldsymbol{\beta}$. Therefore, $P_0$ which is an *equality-constrained* $\ell_0$ *pseudo-norm minimisation problem* can be used for the exact signal recovery problem [CT05]. Similarly, $P_{0,2,\varepsilon}(\boldsymbol{\beta},\boldsymbol{r})$ which is an $\ell_2$*-constrained* $\ell_0$ *pseudo-norm minimisation problem* can be used for the stable signal recovery problem [DET06], where, both $P_0$ and $P_{0,2,\varepsilon}(\boldsymbol{\beta},\boldsymbol{r})$ are in general NP-hard, i.e., they cannot be solved by a polynomial time algorithm [Nat95]; [AK98].

In finding the unique and sparse representations, some researchers focused on a more general cases of $P_p$, $0<p\leq1$ [Cha07]; [DG08]; [FL09]; [DG09b], and $P_p$, $0\leq p\leq1$ [GN03b]; [GN04]; [GN07] for the exact signal recovery problem and $P_{p,2,\varepsilon}(\boldsymbol{\beta},\boldsymbol{r})$, $0<p\leq1$ [FL09] for the stable signal recovery problem.

For $P_{p,q,\varepsilon}(\boldsymbol{\beta}, \boldsymbol{r})$, $0 \leq p < 1$, the optimisation problem is non-convex, combinatorial, and generally computationally intractable [DMA97], i.e., all possible combinations of the columns of the dictionary should be tested for finding the smallest subset of columns satisfying the optimisation constraint, and its complexity grows exponentially with $n$ the length of $\boldsymbol{\beta}$. But there is still hope, either by approximating the solution of $P_{p,q,\varepsilon}(\boldsymbol{\beta}, \boldsymbol{r})$, $0 \leq p < 1$, and relaxing the optimisation problem to $P_{1,q,\varepsilon}(\boldsymbol{\beta}, \boldsymbol{r})$, which can be viewed as a convexification of $P_{p,q,\varepsilon}(\boldsymbol{\beta}, \boldsymbol{r})$, $0 \leq p < 1$, or using numerical bypasses (shortcuts), such as *greedy algorithms*, e.g., matching pursuit [MZ93]; [Mal08], orthogonal matching pursuit [PRK93]; [MZ93]; [DMA97], and basis pursuit [CDS01], or *iterative thresholding algorithms*, e.g., iterative hard thresholding [FN03b]; [DDM08]; [BD08]; [BD09b], iterative soft thresholding, and iterative thresholding with inversion [Mal09]; [MD10].

Matching pursuit as an iterative algorithm can be used for approximating the $P_0$, whereas basis pursuit leads to the exact solution to the $P_0$. At first, in practice it had been observed that $P_1$ and $P_0$ often lead to the same results [CDS01], later it turned out that under sufficient conditions, i.e., if $\boldsymbol{\beta}$ is sufficiently sparse, the solutions of the mentioned problems are indeed the same and unique [DH01]; [EB01]. Regarding the conditions for the relationship between the solutions of $P_1$ and $P_0$, Donoho and Elad have addressed the following two main questions [DE03b]; [DE03a]: Under which conditions,

Q1: the solution to $P_1$ is necessarily the sparsest representation (solution to $P_0$) (*Uniqueness*)?

Q2: the solution to $P_0$ is necessarily equal to the solution to $P_1$ (*Equivalence*)?

In general, $P_{1,q,\varepsilon}(\boldsymbol{a}, \boldsymbol{b})$ is convex, implying that there is no local minima problems in its numerical solution, and based on the optimisation theory, it can be recast as a linear programming problem [GMW90]; [Ber99]; [CDS01]. Therefore, it can be efficiently solved by search problems based on either the classical simplex method or recently popular interior point methods [Wyn79]; [BN03]; [BV04]. There are also some other fast algorithms [CM06]; [Gil+06]; [Gil+07]; [XH07]. Although $P_{1,q,\varepsilon}(\boldsymbol{a}, \boldsymbol{b})$ is convex, it is not strictly convex, so the range of unicity of the solution should be determined [GN03a]. Because of the mentioned convexity, researchers started using $P_1$ or *equality-constrained* $\ell_1$ *norm minimisation problem* or basis pursuit [CDS01]; [DE03b]; [CT05] for exact signal recovery problem, and in parallel $P_{1,2,\varepsilon}(\boldsymbol{\beta}, \boldsymbol{r})$ or $\ell_2$*-constrained* $\ell_1$ *norm minimisation problem* or basis pursuit denoising [CDS01]; [DET06]; [Tro06], for stable signal recovery problem.

There are two other widely used optimisation problems used for stable signal recovery: $P_{2,1,\varepsilon}(\boldsymbol{r}, \boldsymbol{\beta})$, and $P_{1,\infty,\varepsilon}(\boldsymbol{\beta}, \boldsymbol{\Phi}^T \boldsymbol{r})$. The former, $\ell_1$*-constrained* $\ell_2$ *norm minimisation problem* is LASSO[3], and is also known as $\ell_1$-regularized least squares [Tib94]; [Tro06]; [KKB07]; [MY09] and the latter, $\ell_\infty$*-constrained* $\ell_1$ *norm minimisation problem* is known as *Dantzig selector* [CT07]; [CXZ09]; [CWX10].

[3] *Least Absolute Shrinkage and Selection Operator*

**3) Characterisations** To define a framework for recovery or identifiability conditions, some properties and characterisations of the dictionary are introduced in literature, among which the widely studied are gathered in table 1.2. Most of the mentioned properties and characterisations of the dictionary are computationally unrealistic or intractable. Only the property of mutual coherence property or the characterisation of mutual coherence constant have the great advantage to be simple and tractable, but with the expense of making the recovery conditions more restrictive. Next, we recall some basic notations and characterisations which will be used in our work.

**Support and Cardinality:** The Support (or active set) and Cardinality of a vector $\boldsymbol{\beta}$ are defined as:

$$\begin{aligned} \text{Support:} \quad & S(\boldsymbol{\beta}) \stackrel{\text{def}}{=} \{k : \beta_k \neq 0\}, \\ \text{Cardinality:} \quad & |S(\boldsymbol{\beta})| \stackrel{\text{def}}{=} \|\boldsymbol{\beta}\|_0 . \end{aligned}$$

Indeed, the non-convex pseudo-norm $\|\boldsymbol{\beta}\|_0$, simply counts the non-zero elements of vector $\boldsymbol{\beta}$.

**Dictionary Kernel:** The Kernel (or null space) of a dictionary $\boldsymbol{\Phi}$ is defined as:

$$Ker(\boldsymbol{\Phi}) \stackrel{\text{def}}{=} \{\boldsymbol{x} \in \mathbb{R}^n, \boldsymbol{\Phi}\boldsymbol{x} = \boldsymbol{0}\}.$$

The Kernel of a dictionary plays an important role in establishing two different classes of recovery conditions, which will be investigated later.

| Properties | References |
|---|---|
| Null space property | [DH01]; [EB01]; [GN03b]; [FN03a]; [Zha05]; [SXH08]; [CDD09] |
| Robust null space property | [DG09a]; [FR13] |
| Mutual coherence property | [DH01]; [EB01]; [GN03b]; [Tro04a]; [Fuc04]; [Tro06]; [GN07] |
| Restricted isometry property | [CT05]; [Bar+08]; [Can08] |
| Uniform uncertainty principle | [CT05]; [CT06] |
| Exact reconstruction principle | [CT06] |
| Neighborliness of the projected polytopes | [VS92]; [Don05]; [DT05]; [DT06] |
| **Characterisations** | **References** |
| Spark | [GN03b]; [DE03b] |
| Cospark | [CT05] |
| Cumulative coherence | [DE03b]; [Tro04a] |
| Quantity invariant $\gamma_{2S}$ | [FL09] |
| Mutual coherence constant | [DH01]; [EB01]; [GN03b]; [Tro04a]; [Fuc04]; [Tro06]; [GN07] |
| Restricted isometry constant | [CT05]; [Bar+08]; [Can08] |
| Restricted orthogonality constant | [CT05] |
| Asymetric restricted isometry constant | [DG08]; [DG09b] |

Table 1.2: Properties and dictionary characterisations used in literature to define exact and stable recovery conditions.

**Spark:** The smallest number of columns of dictionary $\boldsymbol{\Phi}$ that are linearly dependent is called *Spark*.

Spark has a very important role in guaranteeing the uniqueness of the sparse representation. This importance has been already demonstrated by Gorodnitsky and Rao [GR97].

The Spark of a dictionary $\boldsymbol{\Phi}$ is defined mathematically as [GN03b]; [DE03b]; [DE03a]; [BDE09]:

$$Spark\left(\boldsymbol{\Phi}\right) \stackrel{\text{def}}{=} \min_{\boldsymbol{x}\in Ker(\boldsymbol{\Phi})\setminus\{\mathbf{0}\}} \|\boldsymbol{x}\|_0 \,. \tag{1.5}$$

By the definition of the $Spark(\boldsymbol{\Phi})$, any vector $\boldsymbol{x}$ in the null space of the dictionary, i.e., $\boldsymbol{\Phi x}=\mathbf{0}$, must satisfy $\|\boldsymbol{x}\|_0 \geq Spark(\boldsymbol{\Phi})$.

As it can be seen, Spark characterises the null space of a dictionary using the $\ell_0$ pseudo-norm.

Despite of the superficial similarity between Rank and Spark, they are entirely different in concept and computational complexity.

The Rank of a dictionary is defined as the largest number of the columns of $\boldsymbol{\Phi}$ which are linearly independent, and its computation is sequential (easy to compute), whereas Spark is the size of the smallest number of linearly dependent columns of $\boldsymbol{\Phi}$, and its computation is of combinatorial search with exponential complexity.

In general, without zero columns, the Spark of a dictionary $\boldsymbol{\Phi}\in\mathbb{R}^{m\times n}$ is bounded by [DE03b]:

$$\begin{aligned} 2 \leq Spark(\boldsymbol{\Phi}) &\leq \min\left\{Rank(\boldsymbol{\Phi})+1, n\right\} \\ &\leq \min\left\{m+1, n\right\} \\ &= m+1. \end{aligned} \tag{1.6}$$

A dictionary with maximal Spark, i.e., $Spark(\boldsymbol{\Phi})=m+1$, is called *full-Spark*.

For a dictionary with linearly dependent columns, at least two columns are linearly dependent, i.e., $2\leq Spark(\boldsymbol{\Phi})$, and for many randomly generated dictionaries, $Spark(\boldsymbol{\Phi})$ is equal to $Rank(\boldsymbol{\Phi})+1$.

On the other hand, we have $Rank(\boldsymbol{\Phi})\leq \min\{m,n\}$, and since the problem is underdetermined, i.e., $m<n$, $\min\{m,n\}$ is $m$ and also $\min\{m+1,n\}=m+1$.

As it can be seen in (1.5), the dictionary characterization $Spark(\boldsymbol{\Phi})$ is obtained by minimisation of the intractable non-convex $\ell_0$ pseudo-norm, which is not easy to compute. More precisely, given a matrix $\boldsymbol{\Phi}$, computing $Spark(\boldsymbol{\Phi})$ is NP-hard[4]. The computational complexity of $Spark(\boldsymbol{\Phi})$ is investigated in [TP13].

[4] *Non-deterministic Polynomial-hard*

**Mutual coherence constant:** The maximum pairwise absolute correlation between the normalized atoms of a dictionary can be used as a characterization of the dictionary, which is called Mutual Coherence Constant (MCC) [DH01]. From computational point of view, MCC can be considered as the maximum element of off-diagonal absolute Gram (Gramian) matrix $\boldsymbol{G}(\boldsymbol{\Phi}) \stackrel{\text{def}}{=} \boldsymbol{\Phi}^T\boldsymbol{\Phi}$ [DE03b]; [DE03a], i.e.,

$$\begin{aligned} M(\boldsymbol{\Phi}) \stackrel{\text{def}}{=} \max_{k,k'\neq k} |\langle \boldsymbol{\varphi}_k, \boldsymbol{\varphi}_{k'}\rangle| &= \max_{k,k'\neq k} \left|\boldsymbol{\varphi}_k^T \boldsymbol{\varphi}_{k'}\right| \\ &= \max_{k,k'\neq k} \left|\boldsymbol{G}_{k,k'}(\boldsymbol{\Phi})\right|, \end{aligned}$$

where, $\langle \boldsymbol{a}, \boldsymbol{b}\rangle$ computes the inner product of the vectors $\boldsymbol{a}$ and $\boldsymbol{b}$. Obviously, the main diagonal of $\boldsymbol{G}(\boldsymbol{\Phi})$ is composed of 1s, due to the $\ell_2$-normalization of the atoms of the dictionary. Since it is assumed that the atoms of the dictionary have unit $\ell_2$ norm, MCC satisfies $M(\boldsymbol{\Phi})\leq 1$, because of the Cauchy-Schwarz inequality $|\langle \boldsymbol{\varphi}_k, \boldsymbol{\varphi}_{k'}\rangle| \leq \|\boldsymbol{\varphi}_k\|_2 \|\boldsymbol{\varphi}_{k'}\|_2$. For a dictionary as an orthonormal basis, $M(\boldsymbol{\Phi})=0$, whereas for a dictionary consisting two or $m$ orthonormal bases, $M(\boldsymbol{\Phi})$ is bounded by [HSP06]:

$$\frac{1}{\sqrt{m}} \leq M(\boldsymbol{\Phi}) \leq 1. \tag{1.7}$$

For an equiangular tight frame deterministic dictionary, which satisfies the following three conditions [FR13]:

$$\begin{aligned} \|\boldsymbol{\varphi}_k\|_2 = 1, &\qquad \text{for } k = 1, \cdots, n, \\ |\langle \boldsymbol{\varphi}_k, \boldsymbol{\varphi}_{k'}\rangle| = c, &\qquad \forall k \neq k' \text{ and constant c}, \\ \frac{m}{n}\sum_{k=1}^{n} \langle \boldsymbol{\varphi}, \boldsymbol{\varphi}_k\rangle \boldsymbol{\varphi}_k = \boldsymbol{\varphi}, &\qquad \forall \boldsymbol{\varphi} \in \mathbb{R}^m. \end{aligned}$$

the lower-bound of $M(\boldsymbol{\Phi})$ or the *Welch bound* is $\sqrt{(n-m)/(m(n-1))}$ [Wel74]; [SJ03]. The Welch bound defined for equiangular tight frame deterministic dictionaries is even less than the lower-bound of $M(\boldsymbol{\Phi})$ for a general random dictionary in (1.7), unless $m=1$, for which the two lower-bounds are equal.

MCC was first computed for two orthonormal bases $\boldsymbol{\Phi_1}$ and $\boldsymbol{\Phi_2}$, which is called *basic MCC*:

$$\overline{M}(\boldsymbol{\Phi_1}, \boldsymbol{\Phi_2}) = \max_{k,k'} \left|\boldsymbol{\varphi}_{\mathbf{1}_k}^T \boldsymbol{\varphi}_{\mathbf{2}_{k'}}\right|.$$

Notice that for orthonormal bases $\boldsymbol{\Phi_1}$ and $\boldsymbol{\Phi_2}$ and their concatenation $[\boldsymbol{\Phi_1}, \boldsymbol{\Phi_2}]$, we have $\overline{M}(\boldsymbol{\Phi_1}, \boldsymbol{\Phi_2}) = M([\boldsymbol{\Phi_1}, \boldsymbol{\Phi_2}])$. Supposing the dimension of two orthonormal bases $\boldsymbol{\Phi_1}$ and $\boldsymbol{\Phi_2}$ is $m$ by $m$, it is proved that $1/\sqrt{m} \leq \overline{M}(\boldsymbol{\Phi_1}, \boldsymbol{\Phi_2}) \leq 1$ [DH01]; [EB01]; [EB02]. The lower-bound is achieved when the pair of orthonormal bases are spikes and sines [DH01] or Identity and Hadamard [EB02] matrices or any other orthonormal matrices corresponding to the orthonormal bases, whereas the upper-bound is achieved when at least one of the columns in each of two orthonormal bases is common.

**Cumulative mutual coherence constant:** MCC uses the maximum absolute off-diagonal element of the Gram matrix $\boldsymbol{G}(\boldsymbol{\Phi})$ as the characterization of the dictionary $\boldsymbol{\Phi}$, while by summing over any $k$ elements of $\boldsymbol{G}(\boldsymbol{\Phi})$, we would better characterize the dictionary. This kind of dictionary characterization is called *cumulative MCC*.

The conventional cumulative MCC, (or Babel function [Tro04a], or $\ell_1$-coherence function [FR13]) of a dictionary $\boldsymbol{\Phi}\in\mathbb{R}^{m\times n}$ is defined as:

$$M\left(\boldsymbol{\Phi},k\right)\stackrel{\text{def}}{=}\max_{|\Lambda|=k}\max_{j\notin\Lambda}\sum_{i\in\Lambda}\left|\langle\boldsymbol{\varphi}_i,\boldsymbol{\varphi}_j\rangle\right|, \tag{1.8}$$

where, $\Lambda$ represents $k$ different indices from $\{1,\cdots,n\}$.

Although cumulative MCC is computationally more difficult than MCC characterization, it leads to more weakened or relaxed recovery conditions, i.e., the solution of the corresponding optimisation problem is ensured to be unique even for less sparse representation vectors.

A straightforward extension of $\ell_1$-coherence function would be $\ell_p$-coherence function, defined for any $p>0$ [FR13]:

$$M_p\left(\boldsymbol{\Phi},k\right)\stackrel{\text{def}}{=}\max_{|\Lambda|=k}\max_{j\notin\Lambda}\left(\sum_{i\in\Lambda}\left|\langle\boldsymbol{\varphi}_i,\boldsymbol{\varphi}_j\rangle\right|^p\right)^{\frac{1}{p}}.$$

The $p$ parameter in $M_p(\boldsymbol{\Phi},k)$ controls the $\ell_p$ norm of $k$ off-diagonal elements of $\boldsymbol{G}(\boldsymbol{\Phi})$, which makes the $M_p(\boldsymbol{\Phi},k)$ to have the following properties:

- For $p=1$, $M_p(\boldsymbol{\Phi},k)$ reduces to the conventional cumulative MCC, i.e., $M_1(\boldsymbol{\Phi},k)\equiv M(\boldsymbol{\Phi},k)$.
- For $p=\infty$, for any value of $k$, $M_p(\boldsymbol{\Phi},k)$ reduces to the conventional MCC, i.e., $M_\infty(\boldsymbol{\Phi},k)\equiv M(\boldsymbol{\Phi})$, because $\ell_\infty$ norm of any vector is equal to the maximum absolute value of the vector.

The Welch bound mentioned for the MCC can be extended to the cumulative MCC, i.e., for $k<\sqrt{n-1}$, cumulative MCC is lower-bounded by $k\sqrt{(n-m)/(m(n-1))}$, where, again the lower-bound is achieved when the dictionary is an equiangular tight frame [SV08].

In a similar work for extracting the cumulative coherence of the dictionary, Donoho and Elad introduced $\mu_{1/2}$ and $\mu_1$ of the Gram matrix, which is the smallest $m$ off-diagonal entries in a single row or column of the Gram matrix $\boldsymbol{G}$, which sums at least to $1/2$ and 1, respectively.

There are the relationships $M^{-1}(\boldsymbol{\Phi})\leq\mu_1(\boldsymbol{G})<Spark(\boldsymbol{\Phi})$ and $\mu_{1/2}(\boldsymbol{G})\leq(1/2)\mu_1(\boldsymbol{G})$ [DE03b]; [DE03a].

In another study, based on the manner of identifying the Support set of a sparse signal in the OMP[5] algorithm, *union cumulative coherence* is proposed and denoted by $M_U(\boldsymbol{\Phi}, k)$ [Dos05]; [Zha+15]:

$$M_U\left(\boldsymbol{\Phi}, k\right) \stackrel{\text{def}}{=} \max_{|\Lambda|=k} \left\{ \max_{j \notin \Lambda} \sum_{i \in \Lambda} |\langle \boldsymbol{\varphi}_i, \boldsymbol{\varphi}_j \rangle| + \max_{l \in \Lambda} \sum_{i \in \Lambda \setminus \{l\}} |\langle \boldsymbol{\varphi}_i, \boldsymbol{\varphi}_l \rangle| \right\},$$

where, $\Lambda$ represents $k$ different indices from $\{1, \cdots, n\}$.

$M_U(\boldsymbol{\Phi}, k)$ even better characterizes the dictionary $\boldsymbol{\Phi}$ in comparison to conventional cumulative MCC, i.e., $M(\boldsymbol{\Phi}, k)$. Because, a part from the first common term in $M_U(\boldsymbol{\Phi}, k)$, which seeks the maximum of sum of pairwise absolute *inter-set* correlations between the columns $i \in \Lambda$ and $j \notin \Lambda$, the second term in $M_U(\boldsymbol{\Phi}, k)$ seeks the same characterization for *intra-set* columns $i \in \Lambda$ and $l \in \Lambda$.

Although $M_U(\boldsymbol{\Phi}, k)$ is computationally more complicated than $M(\boldsymbol{\Phi}, k)$, it leads to more accurate analysis of the reconstruction capacity of the orthogonal matching pursuit [Zha+15].

Tropp has shown the following properties for cumulative MCC [Tro04a]:

- $M(\boldsymbol{\Phi}, 0) = 0$ (by convention),
- $M(\boldsymbol{\Phi}, 1) = M(\boldsymbol{\Phi})$,
- $M(\boldsymbol{\Phi}, k) \leq k\, M(\boldsymbol{\Phi})$,
- $M(\boldsymbol{\Phi}, k+1) - M(\boldsymbol{\Phi}, k) \geq 0$,
- $M(\boldsymbol{\Phi}, k+2) - 2M(\boldsymbol{\Phi}, k+1) + M(\boldsymbol{\Phi}, k) \leq 0, \forall k \geq 0$, and
- For orthonormal basis, $M(\boldsymbol{\Phi}, k) = 0, \forall k \geq 0$.

In a more general case, for $1 \leq k_1, k_2 \leq n-1$ with $k_1 + k_2 \leq n-1$, we have [FR13]:

$$\begin{aligned} \max\left\{M(\boldsymbol{\Phi}, k_1), M(\boldsymbol{\Phi}, k_2)\right\} &\leq M(\boldsymbol{\Phi}, k_1 + k_2) \\ &\leq M(\boldsymbol{\Phi}, k_1) + M(\boldsymbol{\Phi}, k_2). \end{aligned}$$

As it can be learned from the above-mentioned different formula of cumulative version of the ordinary coherence measure, this type of dictionary characterisation is more informative and general. Since instead of the first maximum off-diagonal absolute value of the Gram matrix $\boldsymbol{G}(\boldsymbol{\Phi})$, $k$ of them characterise the dictionary $\boldsymbol{\Phi}$.

[5] *Orthogonal Matching Pursuit*

### 1.3.3 Exact recovery condition

One of the fundamental contributions of the overcomplete signal representations community is theoretically-proved necessary and sufficient conditions for exact or stable signal recovery problems. In other words, if based on some regularity conditions for sparsity assumption the representation vector is sufficiently sparse, then it can be recovered exactly or stably [VPF15]. Therefore, ill-posedness issue of the USLE can be solved [DET06].

The importance of the *Exact Recovery Condition(s) (ERC)* or *stable recovery condition* is in guaranteeing the uniqueness or faithful approximation of the solution of the model through the sparse optimisation problem, otherwise, different optimisation algorithms may return different or significantly different solutions.

The theoretical ERC and stable recovery conditions guarantee that the solution can be found independent of the algorithm used, i.e., supposing the existence of a sparse enough solution, it is possible to derive necessary and sufficient conditions for the recovery of the desired solution, regardless of the recovery algorithm used.

As mentioned earlier, in this study we concentrate on the ERC, i.e., the noiseless model in (1.2), i.e., $\boldsymbol{y}=\boldsymbol{\Phi}\boldsymbol{\beta_0}$.

The aforementioned sufficient sparsity condition, which sometimes can be applied explicitly on the representation vector, is determined by a so-called *Sparsity Level (SL)* or *sparsity bound*, which is the upper-bound for the number of non-zero entries of the representation vector, and is derived from the dictionary. The sparsity level of a dictionary is represented by $SL(\boldsymbol{\Phi})$.

In other words, when the representation vector is very sparse, the sparsity level is low and vice versa [DH01]; [CT05]; [Cha07]; [CWX10].

A representation is said to be *k-sparse* if it has at most $k$ non-zero entries, which can be arbitrarily placed anywhere in the representation. In other words, $k$ is less than sparsity level, so for a $k$-sparse representation $\boldsymbol{\beta_0}$, we have $\|\boldsymbol{\beta_0}\|_0 \leq k < SL(\boldsymbol{\Phi})$.

In this section, we briefly explain the main following ERC:

- ERC based on Spark,
- ERC based on null space property,
- ERC based on mutual coherence constant, and
- ERC based on cumulative coherence constant.

**ERC based on Spark** In literature, to approach the problem of determining the sufficient conditions for unique sparse recovery, i.e., ERC, a different problem inspired by the concept of uncertainty principle is considered [GR97]; [DS89].

Consider the problem $P_0$ and suppose $\boldsymbol{\beta_0}$ and $\boldsymbol{\beta_1}$ are two distinct representations of the non-zero signal $\boldsymbol{y}$, in the dictionary $\boldsymbol{\Phi}$, i.e., $\boldsymbol{y}=\boldsymbol{\Phi\beta_0}$ and $\boldsymbol{y}=\boldsymbol{\Phi\beta_1}$. The uncertainty principle of redundant solutions states that a non-zero signal cannot have multiple highly sparse representations. In other words, in a given dictionary $\boldsymbol{\Phi}$ there is a limit on the sparsity level of the representations $\boldsymbol{\beta_0}$ and $\boldsymbol{\beta_1}$:

$$\|\boldsymbol{\beta_0}\|_0 + \|\boldsymbol{\beta_1}\|_0 \geq Spark\,(\boldsymbol{\Phi})\,. \tag{1.9}$$

The mentioned uncertainty principle has been proved based on the definition of Spark in (1.5), i.e., $Spark(\boldsymbol{\Phi})\stackrel{\text{def}}{=}\min_{\boldsymbol{x}\in Ker(\boldsymbol{\Phi})\setminus\{\boldsymbol{0}\}}\|\boldsymbol{x}\|_0$, indicating that for $\boldsymbol{\beta_0}-\boldsymbol{\beta_1}$ in the Kernel of the dictionary $\boldsymbol{\Phi}$, i.e., $\boldsymbol{\Phi}(\boldsymbol{\beta_0}-\boldsymbol{\beta_1})=\boldsymbol{0}$, we have $\|\boldsymbol{\beta_0}-\boldsymbol{\beta_1}\|_0\geq Spark(\boldsymbol{\Phi})$. Then, triangle inequality induces (1.9).

The uncertainty principle in (1.9) has been stated and demonstrated for different cases of dictionaries. At first, the dictionary was considered as a concatenation of two orthonormal bases [DH01]; [EB01]; [EB02]. Then, this uncertainty principle was generalised to dictionaries which arise from the union of more than two orthonormal bases [GN03b]. Finally, it was generalised to dictionaries which can be the concatenation of less structured blocks [DE03b]; [DE03a]; [GN03a].

Using the aforementioned uncertainty principle in different cases of dictionary, and the simple criterion of Spark, the uniqueness of the sparse solution can be demonstrated.

If $\boldsymbol{\beta_0}$ is a candidate solution of the $P_0$ problem for a general dictionary $\boldsymbol{\Phi}$, and meets

$$\|\boldsymbol{\beta_0}\|_0 < \frac{Spark\,(\boldsymbol{\Phi})}{2}, \tag{1.10}$$

then according to the mentioned uncertainty principle in (1.9), any other solution must be denser. Therefore the solution which is sufficiently sparse according to (1.10), is unique and the sparsest.

For a dictionary $\boldsymbol{\Phi}\in\mathbb{R}^{m\times n}$, considering the upper bound of Spark in (1.6), i.e., $m+1$, and the above-mentioned conventional Spark-based condition in (1.10), the admitted sparsity level is at most $(1+m)/2$.

Although among different ERC, ERC based on Spark is the most relaxed condition, i.e., the highest sparsity level, the computation of the Spark characterisation is not tractable [TP13].

**ERC based on null space property** Conventional *Null Space Property (NSP)* provides necessary and sufficient conditions for the exact recovery of $k$-sparse representation via $P_1$, in other words the equivalence of $P_0$ and $P_1$, i.e., the solution to $P_0$ is necessarily equal to the solution to $P_1$.

For a $k$-sparse representation $\boldsymbol{\beta}\in\mathbb{R}^n$, i.e., $|S(\boldsymbol{\beta})|\leq k$, assuming that the Support $S$ of the true solution lies within $S(\boldsymbol{\beta})\subset\{1,\cdots,n\}$, i.e., $S(\boldsymbol{\beta_0})\subset S(\boldsymbol{\beta})$, the mentioned traditional condition states that if

$$Q_p\left(S\left(\boldsymbol{\beta}\right),\boldsymbol{\Phi}\right)\overset{\text{def}}{=}\max_{\boldsymbol{x}\in Ker(\boldsymbol{\Phi})\setminus\{\mathbf{0}\}}\frac{\sum\limits_{i\in S(\boldsymbol{\beta})}|x_i|^p}{\sum\limits_{i}|x_i|^p}<\frac{1}{2},$$

then $\boldsymbol{\beta_0}$ is the unique solution to the problem $P_p$.

NSP has been demonstrated by starting to prove that for all $\boldsymbol{x}\in Ker(\boldsymbol{\Phi})$, we have $\|\boldsymbol{\beta_0}\|_p^p<\|\boldsymbol{\beta_0}+\boldsymbol{x}\|_p^p$. Then using a variant of quasi-triangle inequality, i.e., $|a+b|^p-|a|^p\geq-|b|^p$, and calculating $\ell_p$ norm over on-Support ($\in S(\boldsymbol{\beta})$) and off-Support ($\notin S(\boldsymbol{\beta})$) parts, $Q_p(S(\boldsymbol{\beta}),\boldsymbol{\Phi})$ is achieved.

NSP was first stated for a dictionary which is concatenation of two orthonormal matrices corresponding to orthonormal bases and for $p$ equal to one [DH01]; [EB01]; [EB02]; [FN03a], then was proved for arbitrary nonorthogonal dictionaries [DE03b]; [GN03a]; [Zha05]; [SXH08]; [CDD09]. Later, it was more generalised to $0\leq p\leq 1$ and for dictionaries being a union of orthonormal bases [GN03a]; [GN03b] and for general arbitrary dictionaries [GN04]; [GN07].

In order to investigate the relationship between the ERC based on NSP and the previously mentioned ERC based on Spark in (1.10), let $p=0$ in $Q_p(S(\boldsymbol{\beta}),\boldsymbol{\Phi})$:

$$\begin{aligned}Q_0\left(S\left(\boldsymbol{\beta}\right),\boldsymbol{\Phi}\right)\overset{\text{def}}{=}\max_{\boldsymbol{x}\in Ker(\boldsymbol{\Phi})\setminus\{\mathbf{0}\}}\frac{|S\left(\boldsymbol{\beta}\right)|}{\|\boldsymbol{x}\|_0}=\frac{|S\left(\boldsymbol{\beta}\right)|}{\min\limits_{\boldsymbol{x}\in Ker(\boldsymbol{\Phi})\setminus\{\mathbf{0}\}}\|\boldsymbol{x}\|_0}=\frac{|S\left(\boldsymbol{\beta}\right)|}{Spark\left(\boldsymbol{\Phi}\right)}\\<\frac{1}{2}.\end{aligned}$$

Then, $|S(\boldsymbol{\beta})|<Spark(\boldsymbol{\Phi})/2$.

On the other hand, due to the assumption $S(\boldsymbol{\beta_0})\subset S(\boldsymbol{\beta})$, we have $|S(\boldsymbol{\beta_0})|\leq|S(\boldsymbol{\beta})|$, which leads to the ERC based on Spark. Therefore, the ERC based on NSP is a general property, which in a special case of $p=0$ reduces to the ERC based on Spark in (1.10).

Another stable variant of NSP is also introduced in literature, which is called robust NSP [DG09a]; [FR13]. From algorithmic point of view, robust NSP is used for stable signal recovery via basis pursuit [FR13].

**ERC based on mutual coherence constant** In general, Spark and NSP are computationally unrealistic, in other words, it is computationally intractable in polynomial time to check the identifiability of the model through the recovery conditions, specially when the number of atoms in the dictionary is high. For detailed information about the computational complexity of Spark and NSP, the interested reader is referred to [TP13] and [TP14].

To overcome this shortcoming, another characterisation of the dictionary called *Mutual Coherence Constant* (MCC) was exploited in literature with the expense of making the recovery conditions more restrictive, i.e., lowering the sparsity level.

MCC, which is defined on page 22, i.e., $M(\boldsymbol{\Phi}) \overset{\text{def}}{=} \max_{k,k' \neq k} |\langle \boldsymbol{\varphi}_k, \boldsymbol{\varphi}_{k'} \rangle|$, is a simple approach for characterising the proximity or similarity between the atoms of the dictionary.

MCC was first introduced by Mallat and Zhang to heuristically evaluate the performance of the MP[6] algorithm [MZ93].

Like Spark, to approach the problem of exact signal recovery based on mutual coherence, uncertainty principle was used in literature. Suppose $\boldsymbol{\beta_1}$ and $\boldsymbol{\beta_2}$ are two distinct representations of the non-zero signal $\boldsymbol{y}$ in two orthonormal bases $\boldsymbol{\Phi_1}$ and $\boldsymbol{\Phi_2}$, respectively, i.e., $\boldsymbol{y}=\boldsymbol{\Phi_1}\boldsymbol{\beta_1}=\boldsymbol{\Phi_2}\boldsymbol{\beta_2}$. The basic or classic uncertainty principle states that a non-zero signal cannot have multiple sparse representations in two distinct orthonormal bases, if both bases are mutually incoherent [DS89]; [DH01]; [EB01]; [EB02].

Therefore, there is a limit on the sparsity level of the representations $\boldsymbol{\beta_1}$ and $\boldsymbol{\beta_2}$:

$$\|\boldsymbol{\beta_1}\|_0 + \|\boldsymbol{\beta_2}\|_0 \geq \frac{2}{\overline{M}\left(\boldsymbol{\Phi_1}, \boldsymbol{\Phi_2}\right)}, \tag{1.11}$$

where, $\overline{M}(\boldsymbol{\Phi_1}, \boldsymbol{\Phi_2}) = \max_{k,k'} |\boldsymbol{\varphi}_{\mathbf{1}_k}^T \boldsymbol{\varphi}_{\mathbf{2}_{k'}}|$ is the basic MCC.

In an attempt to extend the basic uncertainty principle to non-orthonormal bases but still square and non-singular matrices $\boldsymbol{\Phi_1}$ and $\boldsymbol{\Phi_2}$, $\|\boldsymbol{\beta_1}\|_0 + \|\boldsymbol{\beta_2}\|_0 \leq (1/2)\tilde{M}^{-1}(\boldsymbol{\Phi_1}, \boldsymbol{\Phi_2})$ is developed as recovery condition for the uniqueness of the solutions of $P_0$ and $P_1$, and their equivalence, where, $\tilde{M}(\boldsymbol{\Phi_1}, \boldsymbol{\Phi_2}) = \max\{\max_{i,j} |\boldsymbol{\Phi_1}^{-1}\boldsymbol{\Phi_2}|_{i,j}, \max_{i,j} |\boldsymbol{\Phi_2}^{-1}\boldsymbol{\Phi_1}|_{i,j}\}$ [DH01].

In fact, the definition of coherence in $\tilde{M}(\boldsymbol{\Phi_1}, \boldsymbol{\Phi_2})$, which computes the $\ell_\infty$ norm (maximum entry in a vector) of $\ell_{1,\infty}$ norm (maximum absolute entry in a matrix) of matrices $\boldsymbol{\Phi_1}^{-1}\boldsymbol{\Phi_2}$ and $\boldsymbol{\Phi_2}^{-1}\boldsymbol{\Phi_1}$, implicitly indicates to the definition of coherence of blocks in a special case, which will be explained in Chapter 2.

Returning back from basic two orthornomal bases $\boldsymbol{\Phi_1}$ and $\boldsymbol{\Phi_2}$ to the dictionaries $\boldsymbol{\Phi}$, and in order to shift the recovery conditions based on Spark to MCC, first we need to find their relationship.

MCC for an orhtogonal matrix is zero, i.e., no constraint. It should be mentioned that, Spark is lower bounded by a function of inverse of MCC, i.e., $Spark(\boldsymbol{\Phi}) \geq f(M^{-1}(\boldsymbol{\Phi}))$.

[6] *Matching Pursuit*

Therefore, we can approximate the intractable characterisation Spark with a computationally tractable characterisation MCC.

For a general dictionary, the smallest value of the MCC is of interest, in other words the tightest lower bound of the Spark, because as mentioned before in (1.10), the sparsity level is defined as half of the Spark, i.e., $\|\boldsymbol{\beta_0}\|_0<Spark(\boldsymbol{\Phi})/2$.

In addition to theoretic results, from algorithmic point of view, generally the smaller the MCC, the better the performance of recovery algorithms. [FR13] has justified the claim for OMP[7], BP[8], and basic thresholding algorithms.

If the dictionary is the concatenation of two orthonormal bases, [DH01] proved that for guaranteeing the uniqueness of the solution of $P_0$ and equivalence of $P_1$, it is sufficient for $f(M^{-1}(\boldsymbol{\Phi}))=1+M^{-1}(\boldsymbol{\Phi})$.

Later, Elad and Bruckstein improved the condition by getting $f(M^{-1}(\boldsymbol{\Phi}))=2M^{-1}(\boldsymbol{\Phi})$ (also proved in [DE03b] and [DE03a] as a special case) and $f(M^{-1}(\boldsymbol{\Phi}))=(2\sqrt{2}-1)M^{-1}(\boldsymbol{\Phi})$ for guaranteeing the uniqueness of the solution of $P_0$ and equivalence of $P_1$, respectively [EB01]. Feuer and Nemirovski proved that the latter is the maximum lower bound for Spark that can be achieved in the $P_1$ problem, i.e., the bound is tight [FN03a].

Supposing that the dictionary arises from the union of $L$ orthonormal bases ($L\geq 2$), Gribonval and Nielsen proved $f(M^{-1}(\boldsymbol{\Phi}))=L/(L-1)M^{-1}(\boldsymbol{\Phi})$ and $f(M^{-1}(\boldsymbol{\Phi}))=(2\sqrt{2}-2+1/(L-1))M^{-1}(\boldsymbol{\Phi})$ for guaranteeing the uniqueness of the solution of $P_0$ and equivalence of $P_1$, respectively [GN03b]; [GN03a]. Tropp also proved the recovery condition resulted from the latter $f(M^{-1}(\boldsymbol{\Phi}))$, for the equivalence of OMP and BP algorithms [Tro04a].

Later, Donoho, Elad, Gribonval, Nielsen, and Bruckstein demonstrated that the previous results of Donoho and Huo, i.e., $f(M^{-1}(\boldsymbol{\Phi}))=1+M^{-1}(\boldsymbol{\Phi})$, can be generalised from a union of two orthonormal bases to a dictionary in a general case, which it can be the concatenation of less structured blocks (in addition to orthonormal bases) [DE03b]; [DE03a]; [GN03a]; [BDE09].

For $\boldsymbol{\beta_0}$ and $\boldsymbol{\beta_1}$ as two distinct representations of the non-zero signal $\boldsymbol{y}$, in the dictionary $\boldsymbol{\Phi}$, i.e., $\boldsymbol{y}=\boldsymbol{\Phi\beta_0}$ and $\boldsymbol{y}=\boldsymbol{\Phi\beta_1}$, and by combining $f(M^{-1}(\boldsymbol{\Phi}))=1+M^{-1}(\boldsymbol{\Phi})$ with (1.9), we have

$$\|\boldsymbol{\beta_0}\|_0 + \|\boldsymbol{\beta_1}\|_0 \geq Spark\left(\boldsymbol{\Phi}\right) \geq 1 + M^{-1}\left(\boldsymbol{\Phi}\right). \tag{1.12}$$

Therefore, for any general dictionary, the $k$-sparse representation vector $\boldsymbol{\beta_0}$ is the unique solution of the $P_0$ and $P_1$ problems, if

$$\|\boldsymbol{\beta_0}\|_0 \leq k < \frac{1 + M^{-1}\left(\boldsymbol{\Phi}\right)}{2}. \tag{1.13}$$

According to the lower bound of MCC for a general random dictionary $\boldsymbol{\Phi}$ in (1.7), i.e., $1/\sqrt{m}$, and the above-mentioned conventional MCC-based condition in (1.13), the sparsity

[7] *Orthogonal Matching Pursuit*
[8] *Basis Pursuit*

level is at most $(1+\sqrt{m})/2$, although for equiangular tight frame deterministic dictionaries it can goes up until $(1+\sqrt{(m(n-1))/(n-m)})/2$.

Gribonval and Nielsen demonstrated that under the condition in (1.13), and for an arbitrary dictionary, the problems $P_p$ with $0 \leq p < 1$ and $P_1$ are equivalent [GN07].

From algorithmic point of view, Tropp proved the same condition of (1.13) for representation recovery through some greedy recovery algorithms, i.e., equivalence of OMP and BP algorithms [Tro04a].

In another algorithmic study, Maleki demonstrated slightly stronger sufficient recovery conditions than (1.13) via iterative thresholding algorithms of iterative thresholding with inversion, namely IHT[9] and IST[10]. He showed that, supposing $\boldsymbol{\beta_0}$ is sorted in descending order of its absolute values [Mal09]:

1. if $k < (1/3)M^{-1}(\boldsymbol{\Phi})$, then iterative thresholding with inversion recovers the true solution, i.e., $\boldsymbol{\beta_0}$, in at most $k$ iterations.
2. if $k < (1/3.1)M^{-1}(\boldsymbol{\Phi})$ and $|\beta_{0_i}|/|\beta_{0_{i+1}}| < 3^{\ell_i - 4}, 1 \leq i < k$, then iterative hard thresholding will recover Support of the true solution, i.e., $S(\boldsymbol{\beta_0})$, in at most $\sum_{i=1}^{k} \ell_i + k$ iterations, and after this number of iterations, without changing the Support, the error of $\boldsymbol{\beta_0}$ recovery will be eliminated exponentially. $\ell_i$ is the step after which $\beta_{0_{i+1}}$ will get into the Support.
3. if $k < (1/4.1)M^{-1}(\boldsymbol{\Phi})$ and $|\beta_{0_i}|/|\beta_{0_{i+1}}| < 3^{\ell_i - 5}, 1 \leq i < k$, then iterative soft thresholding will recover Support of the true solution, i.e., $S(\boldsymbol{\beta_0})$, in at most $\sum_{i=1}^{k} \ell_i + k$ iterations, and after this number of iterations, without changing the Support, the error of $\boldsymbol{\beta_0}$ recovery will be eliminated exponentially. $\ell_i$ is the step after which $\beta_{0_{i+1}}$ will get into the Support.

In another algorithm-based condition, it has been shown that for $k < M^{-1}(\boldsymbol{\Phi})/4$ every $k$-sparse representation can be exactly recovered via iterative hard thresholding [FR13].

In addition, Donoho et al. demonstrated the same condition of (1.13) for a stable solution in a stable signal recovery problem of $P_{0,2,\varepsilon}(\boldsymbol{\beta}, \boldsymbol{r})$ [DET06].

[9] *Iterative Hard Thresholding*
[10] *Iterative Soft Thresholding*

**ERC based on cumulative coherence constant** The coherence characterisation of MCC [DH01] represents only the most extreme correlation between the atoms of the dictionary and does not offer a comprehensive description of the dictionary. In order to extract more information and to better characterise the dictionary, researchers focused on cumulative coherence characterisations. Although it is computationally more difficult than the conventional coherence characterisation, it leads to weakened recovery conditions; i.e., it provides sharper results for the equivalence of $P_0$ and $P_1$ optimisation problems.

The characterization $\mu_{1/2}(\boldsymbol{G})$ introduced by Donoho and Elad as a cumulative coherence characterisation is the smallest $m$ off-diagonal entries in a single row or column of the Gram matrix $\boldsymbol{G}$, which sums at least to 1/2. Based on $\mu_{1/2}(\boldsymbol{G})$, if $\|\boldsymbol{\beta_0}\|_0 < \mu_{1/2}(\boldsymbol{G})$, then $\boldsymbol{\beta_0}$ is the unique solution of the $P_1$ optimisation problem [DE03b].

From algorithmic point of view, considering cumulative MCC defined in (1.8), i.e., $M(\boldsymbol{\Phi}, k) \stackrel{\text{def}}{=} \max_{|\Lambda|=k} \max_{j \notin \Lambda} \sum_{i \in \Lambda} |\langle \boldsymbol{\varphi}_i, \boldsymbol{\varphi}_j \rangle|$, Tropp demonstrated that, if

$$M(\boldsymbol{\Phi}, k) + M(\boldsymbol{\Phi}, k-1) < 1, \tag{1.14}$$

then $k$-sparse representation vector $\boldsymbol{\beta_0}$ can be recovered correctly from orthogonal matching pursuit and basis pursuit algorithms, which is a sufficient condition [Tro04a].

The ERC based on cumulative MCC, i.e., (1.14), is weaker than the ERC based on MCC, i.e., the inequality of (1.13). Because by rearranging (1.13) we have $2\,k\,M(\boldsymbol{\Phi}) - M(\boldsymbol{\Phi}) < 1$, which can be reformulated such as: $kM(\boldsymbol{\Phi}) + (k-1)M(\boldsymbol{\Phi}) < 1$. Now we can compare the reformulated condition of (1.13), i.e., $kM(\boldsymbol{\Phi}) + (k-1)M(\boldsymbol{\Phi}) < 1$, with condition (1.14). Considering the $M(\boldsymbol{\Phi}, k) \leq k\,M(\boldsymbol{\Phi})$ and $M(\boldsymbol{\Phi}, k-1) \leq (k-1)\,M(\boldsymbol{\Phi})$ properties, we conclude that the left-hand side of the just obtained reformulated inequality is greater than or equal to the left-hand side of the inequality in (1.14), i.e., $M(\boldsymbol{\Phi}, k) + M(\boldsymbol{\Phi}, k-1) \leq kM(\boldsymbol{\Phi}) + (k-1)M(\boldsymbol{\Phi})$. Then, higher values of $k$ could meet the condition in (1.14) compared to the reformulated inequality of (1.13). Hence, the condition based on cumulative MCC is weaker than the condition based on MCC.

The results were improved more by a characterisation called union cumulative coherence: $M_U(\boldsymbol{\Phi}, k) \stackrel{\text{def}}{=} \max_{|\Lambda|=k} \{\max_{j \notin \Lambda} \sum_{i \in \Lambda} |\langle \boldsymbol{\varphi}_i, \boldsymbol{\varphi}_j \rangle| + \max_{l \in \Lambda} \sum_{i \in \Lambda \setminus \{l\}} |\langle \boldsymbol{\varphi}_i, \boldsymbol{\varphi}_l \rangle|\}$ [Dos05]; [Zha+15]. If $M_U(\boldsymbol{\Phi}, k) < 1$, then the $k$-sparse representation vector can be exactly recovered by the orthogonal matching pursuit algorithm. The ERC using $M_U(\boldsymbol{\Phi}, k)$ is weaker than (1.14), since $M_U(\boldsymbol{\Phi}, k) < M(\boldsymbol{\Phi}, k) + M(\boldsymbol{\Phi}, k-1)$, and so higher values of $k$ can be selected in $M_U(\boldsymbol{\Phi}, k)$-based condition. There is also a stable recovery condition via orthogonal matching pursuit using the characterisation $M_U(\boldsymbol{\Phi}, k)$ [Zha+15].

In another work, it is been demonstrated that if $M(\boldsymbol{\Phi}, k) + M(\boldsymbol{\Phi}, k-1) < \min_{i \in T} |\beta_{0_i}| / \max_{i \in T} |\beta_{0_i}|$, where, $|T| = k$, then the $k$-sparse representation vector $\boldsymbol{\beta_0}$ can be recovered exactly via basic thresholding algorithm [FR13]. Moreover, it has been shown that if $2M(\boldsymbol{\Phi}, k) + M(\boldsymbol{\Phi}, k-1) < 1$ and $M(\boldsymbol{\Phi}, 2k) < 1/2$, then the $k$-sparse representation is exactly recovered via hard thresholding pursuit and iterative hard thresholding algorithms, respectively [FR13].

## 1.4 Block-sparsity

### 1.4.1 Introduction

According to Tropp, for more than a century, sparsity constraint has been utilised to recover the sparsest unique solution of a proper optimisation problem and to derive the corresponding recovery conditions [Tro04b]. As described in Section 1.3, sparsity constraint penalizes the non-zero coefficients; however, in some real-world engineering problems and applications, which some of them are gathered in table 1.3, applying penalty on non-zero coefficients does not lead to a proper solution compatible to the real-world conditions.

In these kinds of problems, in addition to sparsity, structural, geometrical, or categorical constraints or a priori knowledge may be available that leads to block sparsity [EKB10a], group sparsity [RRN12] and joint sparsity [FR08]. Therefore, a block of coefficients need to be penalized and recovery conditions determine the maximum number of active blocks, and not coefficients, to ensure the uniqueness of the block-sparse solution of the corresponding optimisation problem.

On demonstrating the benefits of the block structure constraint, Kwon and Rao showed that for exact Support recovery in the presence of noise, block-sparsity leads to the reduced number of measurements by the factor of at least the inverse of block length, compared to the conventional element-wise sparse recovery [KR12]. If each block of columns in the dictionary corresponding to each block in the representation vector is composed of linearly dependent columns, the blocks are said to be *redundant* blocks, whereas if the columns are not linearly dependent, the blocks are referred to as *non-redundant* blocks [EV12]. Section 1.4.2 defines the block-sparse representation theory, next in Section 1.4.3, recovery conditions are reviewed.

| Applications | References |
|---|---|
| Radar | [MZ06] |
| Sensor arrays | [MCW05] |
| Signal sampling | [CK04]; [LD08]; [BD09a]; [ME09]; [Eld09]; [ME10]; [GE10] |
| Face recognition | [Wri+09]; [EV11]; [EV12]; [PL15] |
| MRI pulse design | [Zel+08b]; [Zel+08a] |
| Speech recognition | [GC08a]; [GC08b] |
| Music segmentation | [Krs+05]; [Dau06]; [KOP13] |
| Gene expression levels | [ES05]; [Par+08] |
| Multiple measurement vector | [Cot+05]; [CH06]; [ME08]; [FR08]; [BF10]; [ER10]; [LL11]; [LWY13] |
| Sparse communication channels | [CR02] |
| Finding datasets' representatives | [ESV12] |
| Multi-band signal reconstruction | [Lan67]; [ME09]; [Mis+11]; [ME10] |
| EEG/MEG source reconstruction | [GGR95] |
| Data clustering on multiple subspaces | [EV09]; [EV10]; [EV13] |

Table 1.3: Some engineering problems that have structural constraints.

### 1.4.2 Block-sparse representation theory

To elucidate the theory, this section is divided into the following three parts: (1) structure and model, (2) optimisation problem, and (3) characterisations:

**1) Structure and model** As mentioned before, the goal is to extract a representation vector $\hat{\boldsymbol{\beta}}$ which is the block-sparsest among all solutions, i.e., a representation with the fewest active blocks relative to its dimension. For instance, suppose the following estimated representation vector $\hat{\boldsymbol{\beta}}$ which consists of nonoverlapping equally-sized blocks of length $d$, i.e., $\forall k, \hat{\boldsymbol{\beta}}[k] \in \mathbb{R}^d$:

$$\hat{\boldsymbol{\beta}} = \left[\hat{\boldsymbol{\beta}}^T[1], \cdots, \hat{\boldsymbol{\beta}}^T[k], \cdots, \hat{\boldsymbol{\beta}}^T[K]\right]^T,$$

where, $Kd{=}n$, and the $k^{th}$ block is:

$$\hat{\boldsymbol{\beta}}[k] = \left[\hat{\beta}_1[k], \cdots, \hat{\beta}_d[k]\right]^T.$$

Then, the dictionary is represented by:

$$\boldsymbol{\Phi} = [\boldsymbol{\Phi}[1], \cdots, \boldsymbol{\Phi}[k], \cdots, \boldsymbol{\Phi}[K]],$$

where, the $k^{th}$ blocks of atoms of the dictionary is:

$$\boldsymbol{\Phi}[k] = [\boldsymbol{\varphi}_1[k], \cdots, \boldsymbol{\varphi}_d[k]],$$

with $\boldsymbol{\varphi}_j[k] \in \mathbb{R}^m$ and without loss of generality, it is assumed that atoms have unit Euclidean norm, i.e., $\forall j, k, \|\boldsymbol{\varphi}_j[k]\|_2{=}1$. Then, assuming that the true solution $\boldsymbol{\beta_0}$ is exactly recovered, i.e., $\hat{\boldsymbol{\beta}}{=}\boldsymbol{\beta_0}$, the block-sparsity structure of the estimated representation vector $\hat{\boldsymbol{\beta}}$ in the noiseless linear model is represented graphically in figure 1.10. The model can be either noiseless, i.e., $\boldsymbol{y}{=}\boldsymbol{\Phi\beta_0}$, or in general case, noisy, i.e., $\boldsymbol{y}{=}\boldsymbol{\Phi\beta_0}{+}\boldsymbol{e}$, which are defined in (1.2) and (1.1), respectively.

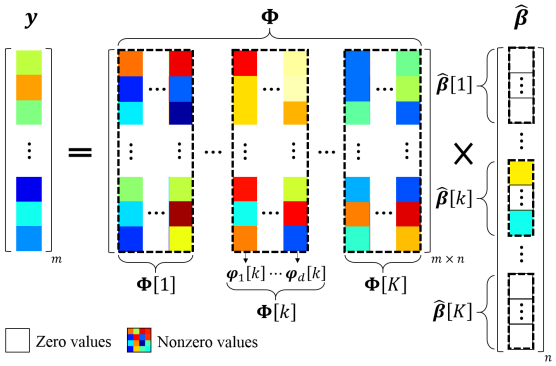



Figure 1.10: Block-sparsity structure of the exactly recovered representation vector $\hat{\boldsymbol{\beta}}$ in the noiseless linear model of $\boldsymbol{y}{=}\boldsymbol{\Phi\beta_0}$.

**2) Optimisation problem** Similar to the conventional element-wise sparse recovery, the problem of finding a solution $\hat{\boldsymbol{\beta}}$, in the noiseless case of (1.2), i.e., $\boldsymbol{y}{=}\boldsymbol{\Phi}\boldsymbol{\beta_0}$, is called *block-sparse exact (unique) signal recovery* whereas in the other case of (1.1), i.e., $\boldsymbol{y}{=}\boldsymbol{\Phi}\boldsymbol{\beta_0}{+}\boldsymbol{e}$, is called *block-sparse stable (robust) signal recovery.*

Block-sparse exact and stable signal recovery problems can be realised through a proper constrained optimisation problem. The straightforward optimisation problem for block-sparse stable signal recovery is:

$$\min_{\boldsymbol{\beta}} \|\boldsymbol{\beta}\|_{2,0} \qquad s.t. \qquad \|\boldsymbol{y}-\boldsymbol{\Phi}\boldsymbol{\beta}\|_2 \leq \varepsilon, \tag{1.15}$$

where, for $\varepsilon{=}0$, the optimisation problem (1.15) solves a block-sparse exact signal recovery problem.

In a general framework, consider the following *constrained $\ell_{p_1,p_2}$ (pseudo-)mixed-norm minimisation problem*:

$$P_{(p_1,p_2),q,\varepsilon}\left(\boldsymbol{a},\boldsymbol{b}\right) : \min_{\boldsymbol{a}} \|\boldsymbol{a}\|_{p_1,p_2} \quad s.t. \quad \|\boldsymbol{b}\|_q \leq \varepsilon, \tag{1.16}$$

where, $\boldsymbol{a}$ and $\boldsymbol{b}$ are vectors, $\varepsilon$ bounds the noise level, i.e., $0{\leq}\epsilon{\leq}\varepsilon$, and $\ell_{p_1,p_2}$ (pseudo-)mixed-norm of block-structured vector $\boldsymbol{a}{=}\left[\boldsymbol{a}^T\left[1\right],\cdots,\boldsymbol{a}^T\left[k\right],\cdots,\boldsymbol{a}^T\left[K\right]\right]^T$, where, $\boldsymbol{a}\left[k\right]{=}\left[a_1\left[k\right],\cdots,a_d\left[k\right]\right]^T$, is defined:

$$\|\boldsymbol{a}\|_{p_1,p_2} \stackrel{\text{def}}{=} \begin{cases} \sum_k I\left(\|\boldsymbol{a}\left[k\right]\|_{p_1}\right), & \text{for } \; p_2 = 0 \\ \left(\sum_k \|\boldsymbol{a}\left[k\right]\|_{p_1}^{p_2}\right)^{\frac{1}{p_2}}, & \text{for } \; 0 < p_2 < +\infty \\ \max_k\left\{\|\boldsymbol{a}\left[k\right]\|_{p_1}\right\}, & \text{for } \; p_2 = \infty, \end{cases}$$

where, $I(a)$ is the indicator function, i.e., $I(a) \stackrel{\text{def}}{=} \begin{cases} 1, & \text{if } |a|{>}0 \\ 0, & \text{if } a{=}0 \end{cases}$.

According to (1.16) the problem (1.15) can be represented as $P_{(2,0),2,\varepsilon}(\boldsymbol{\beta},\boldsymbol{y}{-}\boldsymbol{\Phi}\boldsymbol{\beta})$.

Although the notation $P_{(p_1,p_2),q,\varepsilon}(\boldsymbol{a},\boldsymbol{b})$ can cover all the previous types of optimisation problems, the widely used case is when the minimisation is over $\boldsymbol{\beta}$ and the problem is a block-sparse exact signal recovery, i.e., $P_{(p_1,p_2),q,0}(\boldsymbol{\beta},\boldsymbol{r})$, which for the sake of simplicity we refer to as $P_{p_1,p_2}$. Because regardless of the choice of $q$, the $\|\boldsymbol{r}\|_q{=}0$ directly indicates to the noiseless model, i.e., $\boldsymbol{y}{=}\boldsymbol{\Phi}\boldsymbol{\beta_0}$, so the only parameters to be determined are $p_1$ and $p_2$.

Since the optimisation problem in block-sparsity domain is a generalisation of the conventional optimisation problem in sparsity domain, for $d{=}1$, the two optimisation problems are equivalent, i.e., $P_{p_1,p_2}{\equiv}P_{p_2}$.

The straightforward configuration of the mentioned optimisation problem in a block-sparse exact or stable signal recovery problem is $P_{(p,0),q,\varepsilon}(\boldsymbol{\beta},\boldsymbol{r})$, which simply minimise the number of active blocks, i.e., blocks with non-zero $\ell_p$ norm.

However, the $P_{(p,0),q,\varepsilon}(\boldsymbol{\beta},\boldsymbol{r})$ optimisation problem is in general non-convex and finding its solution is NP-hard[11] [Nat95]; [AK98]. Therefore, there is a need for convexification, relaxation or approximation of $P_{(p,0),q,\varepsilon}(\boldsymbol{\beta},\boldsymbol{r})$. Similar to sparsity domain, in block-sparsity domain the following two main questions are addressed:

Q1: Is the solution to $P_{2,1}$ necessarily the block-sparsest representation (solution to $P_{2,0}$) (*Uniqueness*)?

Q2: Is the solution to $P_{2,0}$ necessarily equal to the solution to $P_{2,1}$ (*Equivalence*)?

It has been shown that under sufficient conditions, $P_{2,1}$ and $P_{2,0}$ are equivalent [EB09]; [EM09]; [SPH09]; [EKB10b]; [EKB10a], and in a more general case the equivalence of $P_{p,1}$ and $P_{p,0}$ for $p>0$ and for both non-redundant and redundant blocks is investigated [EV11]; [EV12].

Subspace matching pursuit is a greedy algorithm, which is used for a block-sparse exact and stable signal recovery problems of $P_{(2,0),q,0}(\boldsymbol{\beta}_{\boldsymbol{\Phi}},\boldsymbol{r})$ and $P_{(2,0),2,\varepsilon}(\boldsymbol{\beta}_{\boldsymbol{\Phi}},\boldsymbol{r})$, respectively, where, $\boldsymbol{\beta}_{\boldsymbol{\Phi}}=\left[(\boldsymbol{\Phi}[1]\boldsymbol{\beta}[1])^T,\cdots,(\boldsymbol{\Phi}[k]\boldsymbol{\beta}[k])^T,\cdots,(\boldsymbol{\Phi}[K]\boldsymbol{\beta}[K])^T\right]^T$, and the $\boldsymbol{\Phi}[k]$'s are assumed to be full rank matrices [GZM09].

In addition, subspace base pursuit and its modified version is used for a block-sparse exact and stable signal recovery problems of $P_{(2,1),q,0}(\boldsymbol{\beta}_{\boldsymbol{\Phi}},\boldsymbol{r})$ and $P_{(2,1),2,\varepsilon}(\boldsymbol{\beta}_{\boldsymbol{\Phi}},\boldsymbol{r})$, respectively, and under sufficient conditions, $P_{(2,1),q,0}(\boldsymbol{\beta}_{\boldsymbol{\Phi}},\boldsymbol{r})$ and $P_{(2,0),q,0}(\boldsymbol{\beta}_{\boldsymbol{\Phi}},\boldsymbol{r})$ are equivalent [GZM09].

Moreover, there are theoretical conditions for ensuring the equivalence of $P_{(p,1),q,0}(\boldsymbol{\beta}_{\boldsymbol{\Phi}},\boldsymbol{r})$ and $P_{(p,0),q,0}(\boldsymbol{\beta}_{\boldsymbol{\Phi}},\boldsymbol{r})$ for $p>0$, while the blocks could be non-redundant or redundant [EV11]; [EV12].

In fact, unlike $P_{p,0}$ that extracts a representation with the fewest active blocks $\boldsymbol{\beta}[k]$, the $P_{(p,0),q,0}(\boldsymbol{\beta}_{\boldsymbol{\Phi}},\boldsymbol{r})$ extracts a representation with the fewest active reconstructed vectors $\boldsymbol{\Phi}[k]\boldsymbol{\beta}[k]$ [Elh12]. When the blocks $\boldsymbol{\Phi}[k]$'s are non-redundant, the solution to $P_{(p,0),q,0}(\boldsymbol{\beta}_{\boldsymbol{\Phi}},\boldsymbol{r})$ has also the fewest active blocks [Elh12].

There are also some recovery conditions of $P_{2,1}$ via the mixed $\ell_2/\ell_1$-optimisation program [EM09]; [SPH09], also known in the statistics literature as group LASSO[12] [YL06]; [LBW11], BMP (Block Matching Pursuit) and BOMP (Block Orthogonal Matching Pursuit) algorithms [EKB10a].

In addition, there are block-sparse stable recovery conditions for block orthogonal matching pursuit and block-thresholding algorithms [BHE11]. Also $\ell_2/\ell_1$-optimisation program [EM09], block versions of CoSaMP[13] and IHT[14] [Bar+10] algorithms (which are the extensions of iterative sparse coding algorithms to model-based sparse representation) lead to robust recoveries in the presence of noise.

[11] *Non-deterministic Polynomial-hard*
[12] *Least Absolute Shrinkage and Selection Operator*
[13] *Compressive Sampling Matched Pursuit*
[14] *Iterative Hard Thresholding*

**3) Characterisations** Similar to the conventional sparse representation theory, in order to define block-sparse recovery conditions, some characterisations of the dictionary are required.

In the following, the main characterisations extended from the conventional notions of Spark and mutual coherence are provided:

- Subspace Spark,
- Block-coherence and sub-coherence,
- Extended block-coherence,
- Fusion coherence,
- Mutual subspace coherence, and
- Cumulative mutual subspace coherence constant.

**Subspace Spark:** In a related work for defining block-sparse recovery conditions, Spark of the set of subspaces $S_i{=}span(\boldsymbol{\Phi}[i]){\in}\mathbb{R}^d$, denoted by $Spark_S(\boldsymbol{\Phi})$, is defined as the smallest integer $r$ such that $\sum_{i=1}^{r} s_i{=}0$, where, $s_i{\neq}0$, $s_i{\in}S_i$ and $i{\in}\{1,\cdots,K\}$ [GZM09].

**Block-coherence and sub-coherence:** Eldar and her colleagues introduced two different measures of the block-wise coherence of the dictionary, named block-coherence and *sub-coherence* [EB09]; [EKB10b]; [EKB10a].

Block-coherence measures the global coherence properties of the dictionary, in other words, quantifies the *inter-block* coherence through:

$$M_{Inter}^{Eldar}\left(\boldsymbol{\Phi}\right) \stackrel{\text{def}}{=} \max_{k,k'\neq k} \frac{1}{d} \left\| \boldsymbol{\Phi}^T\left[k\right] \boldsymbol{\Phi}\left[k'\right] \right\|_{2\to 2}, \tag{1.17}$$

where, $\forall k, \boldsymbol{\Phi}[k]{\in}\mathbb{R}^{m\times d}$, and $d$ is the size of all equally-sized blocks, i.e., $d_1{=}\cdots{=}d_K{=}d$. For a typical matrix $\boldsymbol{A}$, the $\ell_{2\to 2}$ operator-norm, i.e., $\|\boldsymbol{A}\|_{2\to 2}{\stackrel{\text{def}}{=}}\max_{\|\boldsymbol{a}\|_2\leq 1}\|\boldsymbol{A}\boldsymbol{a}\|_2$, is equal to its maximum singular value. For $d{=}1$, the block-coherence is equivalent to the conventional MCC, i.e., $M_{Inter}^{Eldar}(\boldsymbol{\Phi}){\equiv}M(\boldsymbol{\Phi})$.

On the other hand, sub-coherence measures the local coherence properties of the dictionary by computing the *intra-block* coherence through:

$$M_{Intra}^{Eldar}\left(\boldsymbol{\Phi}\right) \stackrel{\text{def}}{=} \max_{\substack{i,j\neq i \\ k}} \left| \boldsymbol{\varphi}_i^T\left[k\right] \boldsymbol{\varphi}_j\left[k\right] \right|. \tag{1.18}$$

The sub-coherence characterisation for dictionaries with intra-block orthonormality is zero, i.e., for $1{\leq}k{\leq}K$, $\boldsymbol{\Phi}^T[k]\boldsymbol{\Phi}[k]{=}\boldsymbol{I}_d$, we have $M_{Intra}^{Eldar}(\boldsymbol{\Phi}){=}0$.

**Extended block-coherence:** As an extension of the studies of Eldar et al., Ziaei and his colleagues study the same problem with different block sizes [Zia+10]. For establishing their findings, they introduced a new version of the block-coherence of Eldar et al., i.e., $M_{Inter}^{Eldar}(\mathbf{\Phi})$ defined in (1.17), replacing $d$ by $\max_{i,j\neq i}\sqrt{d_i d_j}$, where, $d_i$ is the length of block $i$ [Zia+10].

**Fusion coherence:** Fusion coherence in the context of fusion frames is defined as $M_F(\mathbf{\Phi})=\max_{i\neq j}\{|\langle\boldsymbol{\varphi}_i,\boldsymbol{\varphi}_j\rangle|\cdot||\boldsymbol{P}_i\boldsymbol{P}_j||_{2\to 2}\}$, where, $\boldsymbol{P}_i$ denotes the orthogonal projection onto the fusion frame subspace $W_i$ [BKR11]; [Aya14]; [ADR16].

Fusion coherence in the special setting, where the projection matrices are equal to one or all the subspaces have dimension one, reduces to MCC, i.e., $M_F(\mathbf{\Phi})\equiv M(\mathbf{\Phi})$ [BKR11], whereas for the equally-sized subspaces, reduces to the block-coherence defined by Eldar et al., i.e., $M_F(\mathbf{\Phi})\equiv M_{Inter}^{Eldar}(\mathbf{\Phi})$ [BKR11].

**Mutual subspace coherence:** In a related work, mutual subspace coherence has been introduced by Ganesh et al. as a measure of inter-block coherence, which measures the smallest angle between any two subspaces that do not intersect with each other, i.e., disjoint subspaces [GZM09]:

$$M_S\left(\mathbf{\Phi}\right)\stackrel{\text{def}}{=}\max_{i,j\neq i}\max_{\substack{\boldsymbol{x}\in S_i\\ \boldsymbol{y}\in S_j}}\frac{\left|\boldsymbol{x}^T\boldsymbol{y}\right|}{\|\boldsymbol{x}\|_2\,\|\boldsymbol{y}\|_2},$$

where, $S_i{=}span(\mathbf{\Phi}[i])\in\mathbb{R}^d$, under the assumption of linearly independent columns of each block, i.e., non-redundant blocks.

In another work, the concept of mutual subspace coherence has been used for establishing recovery conditions with more relaxed assumption of redundant blocks, where, the columns of each block can be linearly dependent [EV11]; [EV12]. When the blocks are non-redundant and orthonormal, the inter-block coherence proposed by Eldar et al. is equivalent to mutual subspace coherence, i.e., $M_{Inter}^{Eldar}(\mathbf{\Phi})\equiv M_S(\mathbf{\Phi})$ [EV12]. For $d$ equal to 1, the mutual subspace coherence is equivalent to the conventional MCC, i.e., $M_S(\mathbf{\Phi})\equiv M(\mathbf{\Phi})$ [EV12].

**Cumulative mutual subspace coherence constant:** In literature on block sparsity, there is another cumulative characteristic based on subspaces, named cumulative mutual subspace coherence, which is defined as [EV12]:

$$M_S\left(\mathbf{\Phi},k\right)\stackrel{\text{def}}{=}\max_{|\Lambda|=k}\max_{j\notin\Lambda}\sum_{i\in\Lambda}\max_{\substack{\boldsymbol{x}\in S_i\\ \boldsymbol{y}\in S_j}}\frac{\left|\boldsymbol{x}^T\boldsymbol{y}\right|}{\|\boldsymbol{x}\|_2\,\|\boldsymbol{y}\|_2},$$

where, $\Lambda$ represents $k$ different indices from $\{1,\cdots,K\}$. For $d$ equal to 1, the cumulative mutual subspace coherence is equivalent to the conventional cumulative coherence, i.e., $M_S(\mathbf{\Phi},k)\equiv M(\mathbf{\Phi},k)$ [EV12]. Similar properties of conventional cumulative coherence hold true for cumulative mutual subspace coherence, i.e., $M_S(\mathbf{\Phi},1)$ is equal to $M_S(\mathbf{\Phi})$, and $M_S(\mathbf{\Phi},k)\leq kM_S(\mathbf{\Phi})$ [EV12].

### 1.4.3 Block-sparse exact recovery condition

Recently, there has been a huge surge of interest in developing recovery conditions, which ensure the uniqueness or robustness of the block-sparse representation of the USLE. In addition, block-sparsity can be used in dictionary learning, where there is a joint block-sparse representation of signals [ZMRE12].

In addition to the aforementioned practical interest of the block-sparse representation, i.e., compatibility with some real world problems, from a mathematical point of view, assuming the block-wise structure for the representation leads to weakened recovery conditions [EKB10a]; [Zia+10]; [BHE11]. By weakened recovery conditions we mean that for the same number of non-zero elements in the representation, assuming the block structure guarantees the uniqueness of the representation with a higher sparsity level.

Similarly, conditions for guaranteeing the uniqueness or faithful approximation of the solution with the models of (1.2), i.e., $\boldsymbol{y}=\boldsymbol{\Phi}\boldsymbol{\beta_0}$, and (1.1), i.e., $\boldsymbol{y}=\boldsymbol{\Phi}\boldsymbol{\beta_0}+\boldsymbol{e}$, are called *Block-sparse Exact Recovery Condition (Block-ERC)* and *block-sparse stable recovery condition.*

Sparsity level in the block-wise world is called *Block-Sparsity Level (Block-SL)* or *block-sparsity bound*, which is represented by $Block{-}SL(\boldsymbol{\Phi})$. As it can be easily derived, supposing equally-sized blocks of length $d$, i.e., $d_1{=}\cdots{=}d_K{=}d$, to improve the conventional ERC and stable recovery conditions, the $d$ times block-sparsity level have to be greater than the sparsity level, i.e., $d{\times}Block{-}SL(\boldsymbol{\Phi}){>}SL(\boldsymbol{\Phi})$.

Such sparse representations whose non-zero entries appear in a few blocks are referred to as *block-sparse* or *block $k$-sparse representation* [EB09]; [SPH09]; [BHE11]; [EV12] , which $k$ is the maximum number of active blocks.

Similarly, for a block k-sparse representation, for all $p{\geq}0$ we have $\|\boldsymbol{\beta_0}\|_{p,0}{\leq}k{<}Block{-}SL(\boldsymbol{\Phi})$. From conventional element-wise sparsity point of view, for equally-sized blocks, block $k$-sparse representation is equivalent to *$kd$-sparse* representation, i.e., $\|\boldsymbol{\beta_0}\|_0{\leq}kd$.

In this section, we review the following main block-sparse exact recovery conditions:

- Block-ERC based on Spark,
- Block-ERC based on null space property,
- Block-ERC based on mutual coherence constant, and
- Block-ERC based on cumulative mutual coherence constant.

**1) Block-ERC based on Spark** From algorithmic point of view, Ganesh and his colleagues proved that supposing a $k$-subspace sparse measurement vector $\boldsymbol{y}$, if $k<Spark_S(\boldsymbol{\Phi})/2$ then the sparse decomposition is necessarily unique in their proposed algorithms of subspace matching pursuit and subspace base pursuit [GZM09].

**2) Block-ERC based on null space property** In literature, this type of Block-ERC can be divided into the following groups:

**Based on block null space property:** [SPH09] demonstrated that the conventional NSP can be generalised to block-sparse representation. By proposing the following condition, they showed the equivalence of the optimisation problems $P_{2,1}$ and $P_{2,0}$:

$$Q_{2,1}\left(S_b\left(\boldsymbol{\beta}\right),\boldsymbol{\Phi}\right)\overset{\text{def}}{=}\max_{\boldsymbol{x}\in Ker(\boldsymbol{\Phi})\setminus\{\boldsymbol{0}\}}\frac{\sum\limits_{k\in S_b(\boldsymbol{\beta})}\left|\sum\limits_{j=1}^{d}\left|x_j[k]\right|^2\right|^{\frac{1}{2}}}{\sum\limits_{k=1}^{K}\left|\sum\limits_{j=1}^{d}\left|x_j[k]\right|^2\right|^{\frac{1}{2}}}<\frac{1}{2},$$

where, $S_b(\boldsymbol{\beta})$ is subset of a set with all subset of size $k$ of $\{1,\cdots,K\}$, i.e., block $k$-sparse representation, and $d$ is the length of equally-sized blocks. Also, he mentioned that his result can be generalised to $P_{2,p}$, $0<p\leq 1$.

**Based on fusion null space property:** In addition, the fusion NSP defined in the fusion frame framework proposed by Boufounos et al., in a special case in which all the subspaces share the same dimension, becomes similar to the NSP proposed by Stojnic et al. [BKR11].

Moreover, the notions of exact, stable and robust NSP for fusion frames are presented in [Aya14]; [ADR16].

**3) Block-ERC based on mutual coherence constant** In literature, this type of Block-ERC can be divided into the following groups:

**Based on block-coherence:** Eldar and her colleagues extended the basic uncertainty principle to the block-sparse representations [EB09]; [EKB10b]; [EKB10a]. In other words, there is a limit on the block-sparsity level of the representations $\boldsymbol{\beta_1}$ and $\boldsymbol{\beta_2}$:

$$\|\boldsymbol{\beta_1}\|_{2,0}+\|\boldsymbol{\beta_2}\|_{2,0}\geq\frac{2}{d\,\overline{M}_{Inter}^{Eldar}\left(\boldsymbol{\Phi_1},\boldsymbol{\Phi_2}\right)},\tag{1.19}$$

where, the basic block-coherence of Eldar et al., i.e., $\overline{M}_{Inter}^{Eldar}(\mathbf{\Phi_1}, \mathbf{\Phi_2})$, is the maximal coherence between the blocks of the orthonormal bases $\mathbf{\Phi_1}$ and $\mathbf{\Phi_2}$, i.e.,:

$$\overline{M}_{Inter}^{Eldar}(\mathbf{\Phi_1}, \mathbf{\Phi_2}) = \max_{k,k'} \frac{1}{d} \left\| \mathbf{\Phi_1}^T[k]\, \mathbf{\Phi_2}\left[k'\right] \right\|_{2\to 2},$$

where, $d$ is the length of the equally-sized blocks.

Notice that for orthonormal bases $\mathbf{\Phi_1}$, $\mathbf{\Phi_2}$ and their concatenation $[\mathbf{\Phi_1}, \mathbf{\Phi_2}]$, we have $\overline{M}_{Inter}^{Eldar}(\mathbf{\Phi_1}, \mathbf{\Phi_2}) = M_{Inter}^{Eldar}([\mathbf{\Phi_1}, \mathbf{\Phi_2}])$.

Again, supposing the dimension of two orthonormal bases $\mathbf{\Phi_1}$ and $\mathbf{\Phi_2}$ is $m$ by $m$, and $m = R\,d$ ($R$ is an integer), it is proved that $\overline{M}_{Inter}^{Eldar}(\mathbf{\Phi_1}, \mathbf{\Phi_2}) \geq 1/\sqrt{d\,m}$ [EB09]; [EKB10b]; [EKB10a].

Notice that for $d$=1, all the formulation reduces to its conventional one described in Section 1.3.3.

Then, they proposed Block-ERC for ensuring the uniqueness of the solution of their proposed recovery algorithms of $\ell_2/\ell_1$-optimisation program [EM09], block matching pursuit and block orthogonal matching pursuit [EB09]; [EKB10b]; [EKB10a] based on their proposed block version of the conventional MCC.

Their Block-ERC for recovering a block $k$-sparse representation through their proposed recovery algorithms and for equally-sized blocks of length $d$ is [EB09]; [EKB10b]; [EKB10a]:

$$\|\boldsymbol{\beta_0}\|_{2,0} < \frac{1 + \left(d\, M_{Inter}^{Eldar}(\mathbf{\Phi})\right)^{-1} \left(1 - (d-1)\, M_{Intra}^{Eldar}(\mathbf{\Phi})\right)}{2}. \tag{1.20}$$

In a similar work, Ziaei et al. proved nearly the same Block-ERC of (1.20), for the different block size setting, where, all the $d$-s in (1.20) and $M_{Inter}^{Eldar}(\mathbf{\Phi})$ have been replaced by $\max_{i,j\neq i} \sqrt{d_i d_j}$ [Zia+10].

Ziaei et al. proved that their proposed modification of the $\ell_2/\ell_1$-optimisation program for different-size blocks outperforms the $\ell_2/\ell_1$-optimisation program [EM09] and basis pursuit [CDS01], in terms of the error rate [Zia+10].

In the presence of noise, Ben-Haim and Eldar proposed block-sparse stable recovery conditions for block orthogonal matching pursuit and block-thresholding algorithms [BHE11].

**Based on fusion coherence:** In addition, in the context of fusion frame theory, Boufounos et al. have proved that if $k < (1 + M_F^{-1}(\mathbf{\Phi}))/2$, then the unique recovery of his $\ell_0$ and $\ell_1$ norm optimisation problems is guaranteed [BKR11].

Later, Ayaz et al. improves some of the results of Boufounos et al. in the fusion frame setup [Aya14]; [ADR16].

The recovery condition based on the fusion coherence in the special setting, where, all the subspaces share the same dimension, reduces to (1.20) [BKR11].

**Based on mutual subspace coherence:** There is a relationship between subspace Spark and mutual subspace coherence, i.e., $Spark_S(\mathbf{\Phi}) \geq 1+M_S^{-1}(\mathbf{\Phi})$, which is similar to its conventional relationship [GZM09].

From algorithmic point of view, Ganesh and his colleagues proved similar conditions in recovering the representation. Supposing that the measurement vector $\boldsymbol{y}$ is *k-subspace sparse*, they demonstrated that if $k<(1+M_S^{-1}(\mathbf{\Phi}))/2$, then subspace matching pursuit and subspace base pursuit algorithms are guaranteed to find the $k$ subspaces exactly [GZM09].

In a related work based on $M_S(\mathbf{\Phi})$, Elhamifar and Vidal proved that if the measurement vector $\boldsymbol{y}$ has a unique block $k$-sparse representation vector $\boldsymbol{\beta_0}$, and the subspaces spanned by the columns of the blocks are disjoint, then under the condition

$$k < \frac{1 + \left(\frac{1+\epsilon_p}{1-\epsilon_p} M_S\left(\mathbf{\Phi}\right)\right)^{-1}}{1 + \sqrt{\frac{1+\sigma_p}{1+\epsilon_p}}},$$

the optimisation problems $P_{p,1}$ and $P_{p,0}$ are equivalent, and also under the condition

$$k < \frac{1 + \left(\frac{1+\epsilon'_p}{1-\epsilon'_p} M_S\left(\mathbf{\Phi}\right)\right)^{-1}}{2},$$

the optimisation problems $P_{(p,1),q,0}(\boldsymbol{\beta_\Phi}, \boldsymbol{r})$ and $P_{(p,0),q,0}(\boldsymbol{\beta_\Phi}, \boldsymbol{r})$ are equivalent, where, $\sigma_p$, $\epsilon_p$, and $\epsilon'_p$ are some intra-block $p$-restricted isometry constants which are defined based on the restricted isometry properties introduced in the paper [EV11]; [EV12].

**4) Block-ERC based on cumulative mutual coherence** The mutual subspace coherence of Ganesh et al. [GZM09] represents only the most extreme correlation between the subspaces of the dictionary and does not offer a comprehensive description of the dictionary.

In order to extract more information and to better characterise the dictionary, Elhamifar and Vidal focused on the cumulative version of mutual subspace coherence.

In the block-wise world, Elhamifar and Vidal proved that if the measurement vector $\boldsymbol{y}$ has a unique block $k$-sparse representation vector $\boldsymbol{\beta_0}$, and the subspaces spanned by the columns of the blocks are disjoint, then under

$$\sqrt{\frac{1+\sigma_p}{1+\epsilon_p}} M_S\left(\mathbf{\Phi}, k\right) + M_S\left(\mathbf{\Phi}, k-1\right) < \frac{1-\epsilon_p}{1+\epsilon_p},$$

the optimisation problems $P_{p,1}$ and $P_{p,0}$ are equivalent, and also under

$$M_S\left(\mathbf{\Phi}, k\right) + M_S\left(\mathbf{\Phi}, k-1\right) < \frac{1-\epsilon'_p}{1+\epsilon'_p},$$

the optimisation problems $P_{(p,1),q,0}(\boldsymbol{\beta_\Phi}, \boldsymbol{r})$ and $P_{(p,0),q,0}(\boldsymbol{\beta_\Phi}, \boldsymbol{r})$ are equivalent, where, $\sigma_p$, $\epsilon_p$, and $\epsilon'_p$ are some intra-block $p$-restricted isometry constants which are defined based on the restricted isometry properties introduced in the paper [EV12].

## 1.5 Block-sparsity and multi-modality

### 1.5.1 Introduction

In Section 1.2, first the concept of multi-modality is discussed in general. Then, we showed that due to the fact that a brain neuronal activity have electric and magnetic properties, the brain imaging modality of EEG or MEG individually, cannot completely describe the electromagnetic behaviour of the brain neurons. On the other hand, EEG and MEG measure the electrical and magnetic activities of the same cerebral currents, respectively. Since they have complementary information, they can be an appropriate choice as brain imaging modalities in a multi-modality framework. Therefore, we took the distributed EEG and MEG source reconstruction problem as a real-world example of multi-modality.

However, the distributed EEG or/and MEG source reconstruction problems are vastly underdetermined, i.e., the number of EEG or/and MEG sensors is significantly less than the number of brain sources, e.g., in a typical problem there could be 30 sensors against 3000 sources.

Hence, in order to recover a unique solution from the USLE of distributed EEG or/and MEG source reconstruction problems, appropriate constraint(s) need(s) to be applied on the corresponding optimisation problem .

In Section 1.4, we pointed out that, in certain USLE a block of coefficients need to be penalized and not necessarily a single coefficient. Hence, the notion of block-sparsity was extended from the conventional notion of sparsity, which was reviewed in Section 1.3.

As a real-world example, we again took the distributed EEG or/and MEG source reconstruction problem. Since each brain source, which is a dipole, can be represented as a block of size $d$=3, and the source reconstruction problem is consistent with the constraint of block-sparsity.

Although the two mentioned notions of multi-modality and block-sparse representation theory are fundamentally independent, they meet each other in our main real-world problem, i.e., distributed EEG and MEG source reconstruction problem.

In order to recover a unique solution, in addition to imposing the block-sparsity constraint on the corresponding optimisation problem, the number of active blocks should be less than a threshold. The mentioned restricted number of active blocks is explained through recovery conditions, which are explained in Chapter 2.

Next, in Chapter 3 by following a framework to merge the lead-field matrices of each source position, we show that at level of grouping corresponding to maximum inter-group distance and minimum number of groups, the system satisfies the existing conditions.

Finally, in Chapter 4, we demonstrate the impact of EEG and MEG multi-modality and merging framework on brain source space segmentation.

### 1.5.2 Distributed EEG and MEG source reconstruction problem

Regardless of the chosen strategy to combine the EEG and MEG measurements and lead-field matrices, consider the combined measurement vector as $\boldsymbol{y}_{\mathbf{EMEG}}$, the lead-field matrix as $\boldsymbol{\Phi}_{\mathbf{EMEG}}$, and the true source vector as $\boldsymbol{\beta}_{\mathbf{0}}$. Then, the distributed EEG and MEG source reconstruction problem can be translated in mathematical form as an USLE model in an ideally noiseless case, i.e., $\boldsymbol{y}_{\mathbf{EMEG}}=\boldsymbol{\Phi}_{\mathbf{EMEG}}\boldsymbol{\beta}_{\mathbf{0}}$.

On the other hand, each current dipole in the source space, which can be represented as a vector, has three values in three dimensions of $x$, $y$ and $z$ in the Cartesian coordination. So, each current dipole can be represented as a block of length three, i.e., $\boldsymbol{\beta}_{\mathbf{0}}[k]\in\mathbb{R}^3$, $\forall k$. Then, each successive three elements in the source vector $\boldsymbol{\beta}_{\mathbf{0}}$ represents the activity of a current dipole.

From physiological a priori knowledge, it is known that for a given brain task only a few regions of brain will be activated, i.e., sparse regions. Furthermore, for a given active current dipole, the activity of each of values in three dimensions of $x$, $y$ and $z$ is not constrained, then sparsity inside each active block $\boldsymbol{\beta}_{\mathbf{0}}[k]$ is not important.

Therefore, among the infinitely many solutions of the above-mentioned USLE or source activity vectors $\tilde{\boldsymbol{\beta}}$, solutions with the fewest *active blocks* of dimension three would be of interest and not necessarily solutions with the fewest non-zero *scalar* entries.

Ultimately, our real-world distributed EEG and MEG source reconstruction problem perfectly complies with the notions of multi-modality and block-sparse representation theory.

As shown in figure 1.11, without imposing appropriate constraints on the optimisation problem corresponding to the EEG and MEG multi-modal USLE in the source reconstruction problem, there would be infinitely many solutions $\boldsymbol{\beta}_{\mathbf{1}},\cdots,\boldsymbol{\beta}_{\boldsymbol{\infty}}$ (left branch), but by applying block-sparsity constraints, the desired block-sparse solution $\boldsymbol{\beta}_{\mathbf{0}}$ (right branch), can be recovered.

Notice, the left branch in figure 1.11, in contrary to the right branch, is obtained without exploiting any anatomical information from MRI[15], i.e., the volume conduction head model is spherical and is computed analytically. By utilising anatomical constraint, the position of sources is restricted to be placed on a specific points, hence, the resulted lead-field matrix is more precise.

In addition, in order to *uniquely* recover the desired block-sparse solution $\boldsymbol{\beta}_{\mathbf{0}}$, we need to consider block-sparse exact recovery conditions, which are going to be discussed in the next chapter.

[15] *Magnetic Resonance Imaging*

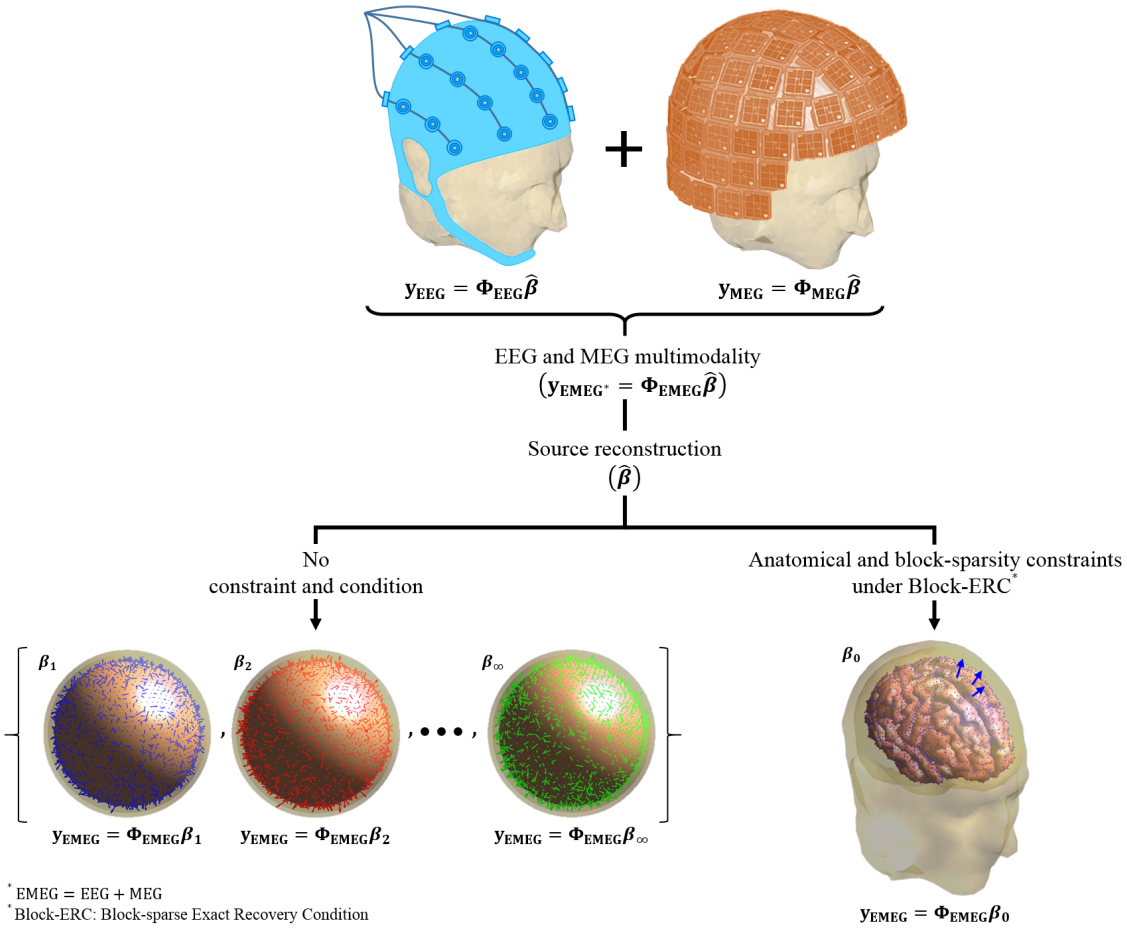



Figure 1.11: EEG and MEG source reconstruction problem using the measurements stored in $\boldsymbol{y_{EEG}}$ and $\boldsymbol{y_{MEG}}$ vectors, and their corresponding lead-field matrices, leads to infinitely many solutions of $\boldsymbol{\beta_1}, \cdots, \boldsymbol{\beta_\infty}$, while by imposing anatomical and block-sparsity constraints under recovery conditions, a unique solution $\boldsymbol{\beta_0}$ can be recovered.

Chapter 2

# Block-sparse exact recovery conditions in general dictionaries

Contents



## 2.1 Introduction

This chapter contains the main contribution of the Ph.D. thesis. In this chapter, a general framework for block-sparse exact signal recovery in an USLE[1] problem, i.e., Block-ERC[2], is proposed. The framework consists of four theoretical and one algorithmic-dependent Block-ERC.

[1] *Underdetermined System(s) of Linear Equations*
[2] *Block-sparse Exact Recovery Condition(s)*

The mentioned generality is in terms of

- structure of the dictionary, i.e., blocks could be orthonormal bases or there may exist intra-block orthonormality, however, in one of the proposed Block-ERC the computation of the Moore-Penrose pseudo-inverse of each block is involved, which requires the blocks to be full column rank,
- norm in the corresponding optimisation problem,
- length of blocks,
- parameters of the proposed block-sparsity measures, e.g., we are not constrained to only Euclidean norm criterion for determining the active blocks, and
- there is no constraint on the relationship between the dimensions of size of the dictionary $\mathbf{\Phi}\in\mathbb{R}^{m\times n}$, except $m<n$.

As mentioned, the recovery conditions are proposed for the exact solution of the optimisation problem, therefore the *uniqueness* of the representation is the center of our focus. In addition, the exact recovery conditions ensure the uniqueness of the Support of the solution, i.e., the problem is Support recovery, and not coefficient recovery.

Through the aforementioned generalisations, we can relax some constraints as mentioned before, and improve the results of Eldar et al. [EB09]; [EKB10b]; [EKB10a], and extend some findings of Donoho et al. [DH01]; [DE03b], Elad and and Bruckstein [EB01]; [EB02], and Gribonval and Nielsen [GN03b]; [GN03a] from element-wise (scalar-wise) to block-wise (vector-wise) framework [APJ16].

This chapter is a major extended version of [APJ16]. Section 2.2 presents the required terminology and tools to establish the main block-sparse recovery conditions which are introduced in Section 2.3. At last, to demonstrate the supremacy of the proposed theoretical recovery conditions in comparison to the state-of-the-art results, some numerical experiments are implemented in Section 2.4, and finally conclusions in Section 2.5 terminate this chapter.

## 2.2 Block-sparsity

Firstly, we need to redefine the structure of the dictionary and the representation vector and consequently the model. Then, an appropriate optimisation problem yielding block-sparse solutions should be investigated. Finally, suitable block-wise characterisations of the dictionary, which would be the base material for proposing the solution's uniqueness will be required. All the above mentioned requirements for establishing Block-ERC are investigated in this section through the following parts of *structure and model*, *optimisation problem*, and *characterisations*.

### 2.2.1 Structure and model

Consider the representation vector $\hat{\boldsymbol{\beta}}$ can be viewed as a concatenation of $K$ individual blocks:

$$\hat{\boldsymbol{\beta}} = \left[\hat{\boldsymbol{\beta}}^T[1], \cdots, \hat{\boldsymbol{\beta}}^T[k], \cdots, \hat{\boldsymbol{\beta}}^T[K]\right]^T,$$

where, the $k^{th}$ block of $\hat{\boldsymbol{\beta}}$ can be represented as:

$$\hat{\boldsymbol{\beta}}[k] = \left[\hat{\beta}_1[k], \cdots, \hat{\beta}_{d_k}[k]\right]^T,$$

where, $\hat{\beta}_{d_k}[k]=[\hat{\beta}_i \mid i=\sum_{j=1}^{k-1} d_j+d_k]$, and the length of $k^{th}$ block is $d_k$, whereas the vector of the blocks' length is:

$$\boldsymbol{d} = [d_1, \cdots, d_K],$$

where, $\sum_{k=1}^{K} d_k=n$, i.e., the block structure is non-overlapped, and there is no constraining relationship between any $d_k$ and $m$. Then, the block structure of the representation vector $\hat{\boldsymbol{\beta}}$, which is consecutive $d_k$-length vectors, is assumed to be known a priori. Similarly, the following block-wise structure is assumed for the dictionary $\boldsymbol{\Phi}$, which can be viewed as a concatenation of all $K$ individual blocks:

$$\boldsymbol{\Phi} = [\boldsymbol{\Phi}[1], \cdots, \boldsymbol{\Phi}[k], \cdots, \boldsymbol{\Phi}[K]],$$

where, $\boldsymbol{\Phi}[k]\in\mathbb{R}^{m\times d_k}$, and as it is mentioned there is not any imposed relationship between $m$ and $n$, other than $m<n$. The $k^{th}$ block is defined as the $d_k$ columns of matrix $\boldsymbol{\Phi}$:

$$\boldsymbol{\Phi}[k] = [\boldsymbol{\varphi}_1[k], \cdots, \boldsymbol{\varphi}_{d_k}[k]],$$

with $\boldsymbol{\varphi}_j[k]\in\mathbb{R}^m$, and without loss of generality, it is assumed that $\boldsymbol{\varphi}_j[k]$ has unit Euclidean norm, i.e., $\forall j,k : \|\boldsymbol{\varphi}_j[k]\|_2=1$. The block structure of $\boldsymbol{\Phi}$ and $\hat{\boldsymbol{\beta}}$ is shown schematically in figure 2.1. The matrix multiplication can be decomposed as sum of multiplication of blocks to reconstruct the measurement vector $\boldsymbol{y}\in span\{\boldsymbol{\Phi}\}\in\mathbb{R}^m$. Therefore, as the model we use the noiseless USLE explained in (1.2), which can be viewed as the block-wise structure, i.e., $\boldsymbol{y}=\boldsymbol{\Phi}\hat{\boldsymbol{\beta}}=\sum_{k=1}^{K}\boldsymbol{\Phi}[k]\hat{\boldsymbol{\beta}}[k]$.

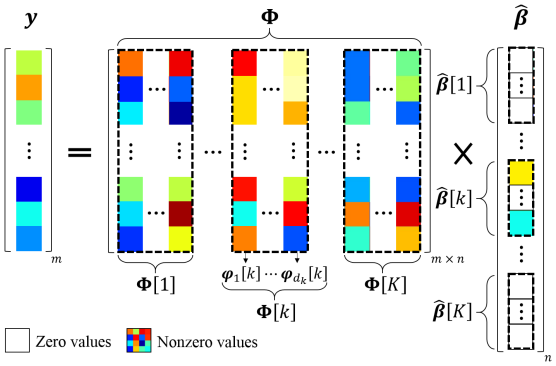



Figure 2.1: Block structure of the dictionary $\boldsymbol{\Phi}$ and the representation vector $\hat{\boldsymbol{\beta}}$ in the noiseless linear model, i.e., $\boldsymbol{y}=\boldsymbol{\Phi}\hat{\boldsymbol{\beta}}=\sum_{k=1}^{K}\boldsymbol{\Phi}[k]\hat{\boldsymbol{\beta}}[k]$.

### 2.2.2 Optimisation problem

As mentioned before, in the conventional element-wise sparse recovery problem, the goal is to extract a representation vector among the set of eligible solutions which is the sparsest one, i.e., a representation with the fewest non-zero elements. But in the framework of block-sparsity, the fewest *active blocks* are of interest, and not fewest non-zero elements. Any active block $k$ has at least one non-zero element in $\boldsymbol{\beta}[k]$ which results in its non-zero $\ell_p$ norm (for any $p$), which can be used for determining the active blocks. Usually, $p$ is assigned to two, because the $\ell_2$ norm is a rotational invariant measure. Here we use the $\ell_p$ norm in more general cases of $0{\leq}p$ and $1{\leq}p$.

As we said, a representation $\boldsymbol{\beta}$ is called block k-sparse if it has at most $k$ active blocks, i.e., $\|\boldsymbol{\beta}\|_{p,0}{\leq}k$. The $\ell_{p,0}$ pseudo-mixed-norm of $\boldsymbol{\beta}$, i.e., $\sum_k I(\|\boldsymbol{\beta}[k]\|_p)$, which can be used for counting the number of active blocks, measures the activity of each block in $\ell_p$ norm sense and the sparsity of the active blocks in $\ell_0$ pseudo-norm sense. Therefore, block's $\ell_p$ norm for any $p$ could be a suitable criterion for determining the activity of blocks and then using another sparsity inducing norm applied on the resulted vector containing the $\ell_p$ norm of each block could implement the concept of block-sparsity.

The previously-introduced (equation (1.4), page 17) general constrained $\ell_p$ norm optimisation problem $P_{p,q,\varepsilon}(\boldsymbol{a},\boldsymbol{b}) : \min_{\boldsymbol{a}} \|\boldsymbol{a}\|_p \quad s.t. \quad \|\boldsymbol{b}\|_q{\leq}\varepsilon$, is defined to recover sparse solution, i.e., solution with the fewest non-zero elements, with the widely used problem $P_p : \min_{\boldsymbol{\beta}} \|\boldsymbol{\beta}\|_p \quad s.t. \quad \boldsymbol{y}{=}\boldsymbol{\Phi\beta}$. Then, the general constrained $\ell_{p_1,p_2}$ (pseudo-)mixed-norm optimisation problem $P_{(p_1,p_2),q,\varepsilon}(\boldsymbol{a},\boldsymbol{b}) : \min_{\boldsymbol{a}} \|\boldsymbol{a}\|_{p_1,p_2} \quad s.t. \quad \|\boldsymbol{b}\|_q{\leq}\varepsilon$, introduced (equation (1.16), page 34) to recover equally-sized block-sparse solution, i.e., solution with the fewest active blocks, where, $\forall k,\ d_k{=}d$, with the widely used problem $P_{p_1,p_2} : \min_{\boldsymbol{\beta}} \|\boldsymbol{\beta}\|_{p_1,p_2} \quad s.t. \quad \boldsymbol{y}{=}\boldsymbol{\Phi\beta}$. In this chapter, we introduce the following *constrained* $\ell^{\boldsymbol{w}}_{p_1,p_2}$ *weighted (pseudo-)mixed-norm optimisation problem* $P_{(\boldsymbol{w};p_1,p_2),q,\varepsilon}(\boldsymbol{a},\boldsymbol{b})$ in a more general case to recover differently-sized block-sparse solution, and cover all previously introduced optimisation problems:

**Definition 2.1 (Constrained $\ell^{\boldsymbol{w}}_{p_1,p_2}$ weighted (pseudo-)mixed-norm optimisation problem).**

$$P_{(\boldsymbol{w};p_1,p_2),q,\varepsilon}\left(\boldsymbol{a},\boldsymbol{b}\right) : \min_{\boldsymbol{a}} \|\boldsymbol{a}\|_{\boldsymbol{w};p_1,p_2} \quad s.t. \quad \|\boldsymbol{b}\|_q \leq \varepsilon,$$

$\ell^{\boldsymbol{w}}_{p_1,p_2}$ *weighted (pseudo-)mixed-norm* of a vector $\boldsymbol{a}$, i.e., $\|\boldsymbol{a}\|_{\boldsymbol{w};p_1,p_2}$, is defined by following definition:

**Definition 2.2 (Weighted (pseudo-)mixed-norm).** The $\ell^{\boldsymbol{w}}_{p_1,p_2}$ weighted (pseudo-)mixed-norm of a block-structured vector $\boldsymbol{a}{=}[\boldsymbol{a}^T[1],\cdots,\boldsymbol{a}^T[k],\cdots,\boldsymbol{a}^T[K]]^T$, where, $\boldsymbol{a}[k]{=}[a_1[k],\cdots,a_{d_k}[k]]^T$, is defined as:

$$\|\boldsymbol{a}\|_{\boldsymbol{w};p_1,p_2} \stackrel{\text{def}}{=} \begin{cases} \sum_k I\left(w_k \left\|\boldsymbol{a}\left[k\right]\right\|_{p_1}\right), & \text{for } \ p_2 = 0 \\ \left(\sum_k w_k^{p_2} \left\|\boldsymbol{a}\left[k\right]\right\|_{p_1}^{p_2}\right)^{\frac{1}{p_2}}, & \text{for } \ 0 < p_2 < +\infty \\ \max_k \left\{w_k \left\|\boldsymbol{a}\left[k\right]\right\|_{p_1}\right\}, & \text{for } \ p_2 = \infty, \end{cases}$$

where, $I(a)$ is the indicator function, i.e., $I(a) \stackrel{\text{def}}{=} \begin{cases} 1, & \text{if } |a|>0 \\ 0, & \text{if } a=0 \end{cases}$, and $\boldsymbol{w}=[w_1, \cdots, w_K]$ is the weight vector.

The $\ell^{\boldsymbol{w}}_{p_1,p_2}$ for either $0\leq p_1<1$ or $0\leq p_2<1$ is called weighted *pseudo*-mixed-norm, consequently, the corresponding problem $P_{(\boldsymbol{w};p_1,p_2),q,\varepsilon}(\boldsymbol{a},\boldsymbol{b})$ is called constrained $\ell^{\boldsymbol{w}}_{p_1,p_2}$ weighted *pseudo*-mixed-norm optimisation problem.

$\ell^{\boldsymbol{w}}_{p_1,p_2}$ weighted (pseudo-)mixed-norm when all the weights are 1, i.e., $\boldsymbol{w}=\mathbf{1}_{1\times K}$, is equal to the $\ell_{p_1,p_2}$ (pseudo-)mixed-norm, i.e., $\ell^{\mathbf{1}}_{p_1,p_2}\equiv\ell_{p_1,p_2}$. In addition, for $p_1>0$ and $p_2=0$, the $\ell^{\boldsymbol{w}}_{p_1,0}$ weighted pseudo-mixed-norm is independent of weight vector $\boldsymbol{w}$, so it equals to ordinary pseudo-mixed-norm, i.e., $\ell^{\boldsymbol{w}}_{p,0}\equiv\ell_{p,0}$ for $p>0$.

Although $P_{(\boldsymbol{w};p_1,p_2),q,\varepsilon}(\boldsymbol{a},\boldsymbol{b})$ with carefully selected values for $p_1$ and $p_2$, is defined generally and can cover lots of scenarios, but the proposed conditions in this study concentrate on the following special cases:

- $\boldsymbol{a}$: $\boldsymbol{a}$ is chosen to be the representation vector, i.e., $\boldsymbol{a}=\boldsymbol{\beta}$.
- $\boldsymbol{b}$: $\boldsymbol{b}$ is selected as $\boldsymbol{r}$, i.e., $\boldsymbol{b}=\boldsymbol{r}=\boldsymbol{y}-\boldsymbol{\Phi\beta}$.
- $\varepsilon$ (and $q$): The representation vector $\boldsymbol{\beta}$ is minimised in a block-sparse exact signal recovery problem, hence, $\varepsilon$ should be zero, i.e., $P_{(\boldsymbol{w};p_1,p_2),q,0}(\boldsymbol{\beta},\boldsymbol{r})$. For the sake of simplicity we refer to $P_{(\boldsymbol{w};p_1,p_2),q,0}(\boldsymbol{\beta},\boldsymbol{r})$ as $P_{\boldsymbol{w};p_1,p_2}$, because regardless of choice of $q$, the $\|\boldsymbol{r}\|_q=0$ directly indicates to the noiseless model, so the only parameters to be determined are $\boldsymbol{w}$, $p_1$ and $p_2$. Therefore, the utilised optimisation problem is $P_{\boldsymbol{w};p_1,p_2}: \min_{\boldsymbol{\beta}} \|\boldsymbol{\beta}\|_{\boldsymbol{w};p_1,p_2} \quad s.t. \quad \boldsymbol{y}=\boldsymbol{\Phi\beta}$.
- $\boldsymbol{w}$: Throughout this thesis, only two cases for $\boldsymbol{w}$ is considered. In a block-structured vector $\boldsymbol{a}=[\boldsymbol{a}^T[1], \cdots, \boldsymbol{a}^T[K]]^T$ with length vector $\boldsymbol{d}=[d_1, \cdots, d_K]$, $w_k$ in $w_k\|\boldsymbol{a}[k]\|_{p_1}$ is either $d_k^{-1/p_1}$ or 1. Therefore, although in general the weight vector $\boldsymbol{w}$ can consist of any arbitrary values, in this work it is considered as $\boldsymbol{d}^{-1/p_1}$ and the corresponding problem is represented as $P_{\boldsymbol{w};p_1,p_2}$, except when mentioned explicitly that it is a vector of ones, i.e., $\mathbf{1}_{1\times K}$, and it can be represented as $P_{\mathbf{1};p_1,p_2}$. But $P_{\mathbf{1};p_1,p_2}$ is equal to $P_{p_1,p_2}$, so we refer this case to $P_{p_1,p_2}$. Therefore, we have two cases of $P_{p_1,p_2}$ and $P_{\boldsymbol{w};p_1,p_2}$, where the weight vector $\boldsymbol{w}$ is:
$$\forall k \in \{1, \cdots, K\}, \qquad w_k = d_k^{-\frac{1}{p_1}}.$$
- $p_1$ and $p_2$: The main proposed Block-ERC include the cases $0=p_2\leq p_1$, and the more general one of $0\leq p_2\leq 1\leq p_1$.

It is obvious that, if the size of all the blocks is chosen to be 1, i.e., $\forall k, d_k=1$, then the block-sparse exact signal recovery reduces to the conventional exact signal recovery, i.e., $P_{\boldsymbol{w};p_1,p_2}\equiv P_{p_1,p_2}\equiv P_{p_2}$. In addition, the problem $P_{\boldsymbol{w};p_1,p_2}$ for $0=p_2<p_1$ is independent of the

weight vector $\boldsymbol{w}$ and equals to ordinary pseudo-mixed-norm problem, i.e., $P_{\boldsymbol{w};p,0}{\equiv}P_{p,0}$ for $p{>}0$.

As it can be seen in Section 1.4.2, most of the previous works are based on Euclidean norm as a measure of block activity, i.e., $P_{2,p}$. However, it has been shown that by using norms other than Euclidean, the performance of the selection of group of variables increases [ZRY09]. So, we study the impact of other norms.

In this chapter, we first introduce Block-ERC based on $P_{p,0}$ in two different cases of $p{\geq}0$ and $p{\geq}1$. Then, we generalise the results to $P_{p_1,p_2}$ and $P_{\boldsymbol{w};p_1,p_2}$, where, the activity of blocks is measured by $\ell_{p_1}$ norm, $p_1{\geq}1$, and the sparsity of the blocks by $\ell_{p_2}$ (pseudo-)norm, $0{\leq}p_2{\leq}1$.

**Remark 2.1 ($P_{\boldsymbol{w};p_1,p_2}$ v.s. $P_{p_1,p_2}$).** Notice that in general:

- For equally-sized blocks, i.e., $\forall k, d_k{=}d$, we propose Block-ERC based on $P_{p_1,p_2}$, where, $p_1$ can have zero value, while for a more general case of differently-sized blocks we propose Block-ERC based on $P_{\boldsymbol{w};p_1,p_2}$ for $p_1{\neq}0$, to have finite values for the elements of the weight vector $\boldsymbol{w}$, because $w_k{=}d_k^{-1/p_1}$.
- As mentioned before, if $0{=}p_2{<}p_1$ then the $\ell^{\boldsymbol{w}}_{p_1,p_2}$ weighted pseudo-mixed-norm optimisation problem is equivalent to the $\ell_{p_1,p_2}$ pseudo-mixed-norm optimisation problem, i.e., $\forall p{>}0$, $P_{\boldsymbol{w};p,0}{\equiv}P_{p,0}$.

The selection of the values of the (pseudo-)mixed-norm coefficients $p_1$ and $p_2$ in the corresponding optimisation problems $P_{p_1,p_2}$ and $P_{\boldsymbol{w};p_1,p_2}$ is very important. Since the optimisation problems $P_{p_1,p_2}$ and $P_{\boldsymbol{w};p_1,p_2}$ are presented in a general case, by selecting a certain values of the (pseudo-)mixed-norm coefficients $p_1$ and $p_2$, one can switch the problem between the conventional sparsity and block-sparsity problem. Map of the values of the (pseudo-)mixed-norm coefficients $p_1$ and $p_2$, which leads to sparsity and block-sparsity problems is shown in figure 2.2.

Figure 2.2 represents schematically our regions of interest in orange colour for the (pseudo-)mixed-norm coefficients $p_1$ and $p_2$ in a block sparsity problems $P_{p_1,p_2}$ or $P_{\boldsymbol{w};p_1,p_2}$, compared to the traditional sparsity problem, shown in blue colour region. As it can be seen in figure 2.2, considering additional constraint of proposed block sparsity (orange-coloured regions) on the representation vector, gives rise to a wider region of the (pseudo-)mixed-norm coefficients of the corresponding optimisation problem compared to the existing block sparsity (green-coloured regions), and conventional sparsity (blue-coloured regions). In figure 2.2, the most frequently studied cases are represented by solid circles. For instance, $P_{0,1}$ ($\equiv P_0$) and $P_{1,1}$ ($\equiv P_1$), which are represented by blue circles (left and right one, respectively) are the widely used cases in sparse exact recovery problems, whereas $P_{2,0}$ and $P_{2,1}$, which are represented by green circles (down and top one, respectively) are the widely used cases in block-sparse exact recovery problems.

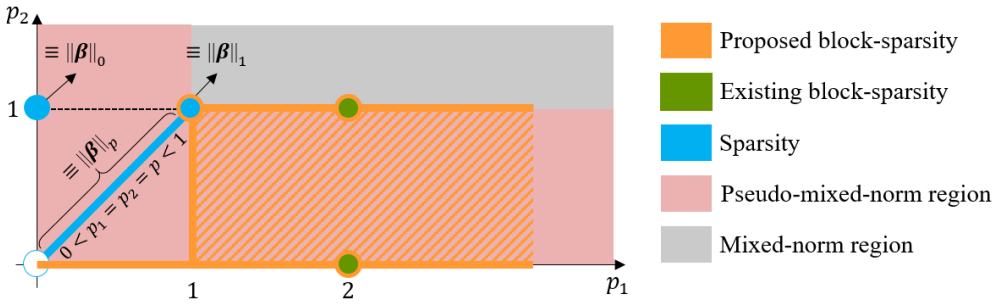



Figure 2.2: Sparsity and block sparsity regions in terms of $\|\boldsymbol{\beta}\|_{p_1,p_2}$, and $\|\boldsymbol{\beta}\|_{\boldsymbol{w};p_1,p_2}$, assuming $\forall k, d_k{=}d$. The most commonly studied cases are indicated with solid circles.

**Remark 2.2 (Sparsity regions in figure 2.2).** For $0{<}p_1{=}p_2{=}p{\leq}1$, the block-sparse exact signal recovery optimisation problems $P_{p,p}$ and $P_{\boldsymbol{w};p,p}$ reduce to the traditional exact signal recovery problem, i.e., $P_{p,p}{\equiv}P_p$ and $P_{\boldsymbol{w};p,p}{\equiv}d^{-1/p}\,P_p$ $(\forall k, d_k{=}d)$. In other words, in the common region of sparsity and block-sparsity, i.e., $p_1{=}p_2{=}1$ (the right blue circle in figure 2.2), and under certain conditions, the conventional exact signal recovery problem can recover the block-sparse representation. In addition, for $p_1{=}0$ and $p_2{=}1$ (the left blue circle in figure 2.2), the $P_{0,1}$ problem reduces to the classic exact signal recovery problem, i.e., $P_{0,1}{\equiv}P_0$, but it is still non-convex.

**Remark 2.3 (Block-sparsity regions in figure 2.2).** As shown in figure 2.2 the proposed block-sparsity region (orange region) covers the existing most frequently studied cases in block sparsity problem (green circles) and the widely used tractable case in the traditional sparsity problem (the right blue circle) in literature.

### 2.2.3 Characterisations

To define the Block-ERC, it is necessary to introduce some notations. Except when mentioned explicitly, (1) the block length $d_k \in \mathbb{N}$, where, $\sum_{k=1}^{K} d_k = n$, and $d_{max}$ and $d_{min}$ are the maximum and minimum block length, respectively, and (2) the characterisations coincide with the results known in the conventional element-wise case if all blocks are of size 1, i.e., $d_1 = \cdots = d_K = 1$. We first give some general definitions, which then will be used to prove uniqueness results.

**Definition 2.3 (Block-Support and Block-Cardinality).** The *Block-Support* and *Block-Cardinality* of a representation vector $\boldsymbol{\beta} \in \mathbb{R}^n$ are defined as:

$$\begin{aligned} &\text{Block-Support:} && S_b(\boldsymbol{\beta}) \stackrel{\text{def}}{=} \left\{ k : \|\boldsymbol{\beta}[k]\|_p \neq 0, 1 \leq k \leq K \right\} && \forall p \geq 0, \\ &\text{Block-Cardinality:} && |S_b(\boldsymbol{\beta})| \stackrel{\text{def}}{=} \|\boldsymbol{\beta}\|_{p,0} && \forall p \geq 0. \end{aligned}$$

$\|\boldsymbol{\beta}\|_{p,0}$ simply counts the number of active blocks in the sense of their non-zero $\ell_p$ norm.

In fact, considering the definition of $\|\boldsymbol{\beta}\|_{p,0}$, each block is first represented by its $\ell_p$ norm. So, non-zero representatives indicate active blocks. Then, by applying the $\ell_0$ pseudo-norm on the blocks' representatives, the active blocks are counted. When $d_1 = \cdots = d_K = 1$, the Block-Support and Block-Cardinality will be equivalent to the conventional Support and Cardinality (see Section 1.3.2), respectively.

**Definition 2.4 (Dictionary Kernel).** The *Block-Kernel* or *block null space* of a dictionary $\boldsymbol{\Phi} \in \mathbb{R}^{m \times n}$, is equivalent to its usual Kernel:

$$\begin{aligned} Block\text{-}Ker(\boldsymbol{\Phi}) &\stackrel{\text{def}}{=} \left\{ \boldsymbol{x} \in \mathbb{R}^n, \sum_{k=1}^{K} \boldsymbol{\Phi}[k]\, \boldsymbol{x}[k] = \boldsymbol{\Phi}\boldsymbol{x} = \boldsymbol{0} \right\} \\ &= Ker(\boldsymbol{\Phi}). \end{aligned}$$

Actually, the Block-Kernel is the block-wise definition of the conventional Kernel of a dictionary. Because of their equivalency and for the sake of simplicity, we use the notation of conventional Kernel.

Regarding the uniqueness of the solution of $P_{p,0}$ problem, an interesting question is "What is the minimal value of $\|\boldsymbol{x}\|_{p,0}$ for $\boldsymbol{x} \in Ker(\boldsymbol{\Phi}) \backslash \{\boldsymbol{0}\}$, and $p > 0$?". To elucidate the significance of the mentioned question, assume that $\boldsymbol{\beta_0}$ and $\boldsymbol{\beta_1}$ are both solutions to problem $P_{p,0}$, i.e., $\boldsymbol{y} = \boldsymbol{\Phi\beta_0} = \boldsymbol{\Phi\beta_1}$. Then $\boldsymbol{\Phi}(\boldsymbol{\beta_0} - \boldsymbol{\beta_1}) = \boldsymbol{0}$, so $\boldsymbol{x} = \boldsymbol{\beta_0} - \boldsymbol{\beta_1} \in Ker(\boldsymbol{\Phi})$. Clearly, $\|\boldsymbol{x}\|_{p,0} = \|\boldsymbol{\beta_0} - \boldsymbol{\beta_1}\|_{p,0} \leq \|\boldsymbol{\beta_0}\|_{p,0} + \|\boldsymbol{\beta_1}\|_{p,0} = 2\|\boldsymbol{\beta_0}\|_{p,0}$ since both $\|\boldsymbol{\beta_0}\|_{p,0}$ and $\|\boldsymbol{\beta_1}\|_{p,0}$ attain the minimum $\ell_{p,0}$ pseudo-mixed-norm. Consequently, assuming two solutions we reach to $\|\boldsymbol{x}\|_{p,0}/2 \leq \|\boldsymbol{\beta_0}\|_{p,0}$.

Hence, if $\|\boldsymbol{\beta_0}\|_{p,0} < \min_{\boldsymbol{x} \in Ker(\boldsymbol{\Phi}) \backslash \{\boldsymbol{0}\}} \|\boldsymbol{x}\|_{p,0}/2$, then no other solution $\boldsymbol{\beta_1}$ exists for which $\|\boldsymbol{\beta_0}\|_{p,0} = \|\boldsymbol{\beta_1}\|_{p,0}$. We stress the importance of the latter through the following definition which is a straightforward generalisation of the conventional Spark [GN03b]; [DE03b]:

**Definition 2.5 (Block-Spark).** The *Block-Spark* of a dictionary can be defined based on $\ell_{p,0}$ pseudo-mixed-norm:

$$Block{-}Spark(\boldsymbol{\Phi}) \stackrel{\text{def}}{=} \min_{\boldsymbol{x}\in Ker(\boldsymbol{\Phi})\setminus\{\mathbf{0}\}} \|\boldsymbol{x}\|_{p,0} \qquad \forall p \geq 0.$$

Block-Spark characterises the block null space of a dictionary using the $\ell_{p,0}$ pseudo-mixed-norm.

By definition, any vector $\boldsymbol{x}$ in the block null space of the dictionary, i.e., $\sum_{k=1}^{K} \boldsymbol{\Phi}[k]\boldsymbol{x}[k]{=}\mathbf{0}$, must satisfy $\forall p{\geq}0, \|\boldsymbol{x}\|_{p,0}{\geq}Block{-}Spark(\boldsymbol{\Phi})$. This important block-wise characterisation of the dictionary is the extension of the conventional Spark defined in equation (1.5) page 21.

Because of the fact that the proposed characterisation is a block-wise generalisation of the conventional element-wise case, in a unit block size setting, i.e., $\forall k, d_k{=}1$, our proposed Block-Spark is equivalent to the conventional Spark, i.e., $Block{-}Spark(\boldsymbol{\Phi}){\equiv}Spark(\boldsymbol{\Phi})$.

The following property shows the relationship between the proposed Block-Spark and the conventional Spark, which will be used in Section 2.3.1 to demonstrate the improvement of the recovery conditions based on Block-Spark compared to the ones based on the conventional Spark:

**Property 2.1 (Block-Spark v.s. Spark).** Let $d_k$ be the block length of the $k^{th}$ block in $\boldsymbol{\Phi}$, and denoting $\bar{d}{=}\sum_{k\in S_b(\boldsymbol{x}_b^\star)} d_k / |S_b(\boldsymbol{x}_b^\star)|$ the average block length, with $\boldsymbol{x}_b^\star{=}\arg\min_{\boldsymbol{x}\in Ker(\boldsymbol{\Phi})\setminus\{\mathbf{0}\}} \|\boldsymbol{x}\|_{p,0}$, $\forall p{\geq}0$, then we have:

$$\bar{d}\, Block{-}Spark(\boldsymbol{\Phi}) \geq Spark(\boldsymbol{\Phi}).$$

*Proof.* Let $\boldsymbol{x}^\star{\in}\arg\min_{\boldsymbol{x}\in Ker(\boldsymbol{\Phi})\setminus\{\mathbf{0}\}} \|\boldsymbol{x}\|_0$ and $\boldsymbol{x}_b^\star{\in}\arg\min_{\boldsymbol{x}\in Ker(\boldsymbol{\Phi})\setminus\{\mathbf{0}\}} \|\boldsymbol{x}\|_{p,0}$. Obviously, $\|\boldsymbol{x}^\star\|_0{\leq}\|\boldsymbol{x}_b^\star\|_0$, and $\|\boldsymbol{x}_b^\star\|_0{\leq}\sum_{k\in S_b(\boldsymbol{x}_b^\star)} d_k$, indeed the whole block is activated even if only a single element is needed. Now letting $\bar{d}{=}|S_b(\boldsymbol{x}_b^\star)|^{-1}\sum_{k\in S_b(\boldsymbol{x}_b^\star)} d_k{=}\|\boldsymbol{x}_b^\star\|_{p,0}^{-1}\sum_{k\in S_b(\boldsymbol{x}_b^\star)} d_k$, the transitivity of the inequalities yields $\|\boldsymbol{x}^\star\|_0{\leq}\bar{d}\,\|\boldsymbol{x}_b^\star\|_{p,0}$, which is exactly $\bar{d}\, Block{-}Spark(\boldsymbol{\Phi}){\geq}Spark(\boldsymbol{\Phi})$. □

Due to non-convexity of the pseudo-mixed-norm $\|\cdot\|_{p,0}$, Block-Spark is not computationally tractable. Therefore, the following block coherence measure, which is the block-wise extension of the conventional MCC[3] is proposed:

[3] *Mutual Coherence Constant*

**Definition 2.6 (Block coherence).** The $(q,p)$-*Block Mutual Coherence Constant (Block-MCC$_{q,p}$)* of a dictionary is defined $\forall(q,p)\in\mathbb{R}^2_{>0}$ as:

$$M_{q,p}\left(\boldsymbol{\Phi}\right) \stackrel{\text{def}}{=} \max_{k,k'\neq k} \frac{d_k^{-\frac{1}{p}}\, d_{k'}^{\frac{1}{q}}}{d_{max}} \left\| \boldsymbol{\Phi}^{\dagger}\left[k\right] \boldsymbol{\Phi}\left[k'\right] \right\|_{q\to p},$$

where, $\boldsymbol{\Phi}^{\dagger}[k]$ is Moore-Penrose pseudo-inverse of the full column rank block $\boldsymbol{\Phi}[k]\in\mathbb{R}^{m\times d_k}$, and $\|\boldsymbol{\Phi}\|_{q\to p}$ is the $\ell_{q\to p}$ operator-norm of dictionary $\boldsymbol{\Phi}$, which can be computed through any of the following expressions [Tro04b]; [GL13]:

$$\|\boldsymbol{A}\|_{q\to p} \stackrel{\text{def}}{=} \max_{\boldsymbol{a}\neq\boldsymbol{0}} \frac{\|\boldsymbol{A}\boldsymbol{a}\|_p}{\|\boldsymbol{a}\|_q} = \max_{\|\boldsymbol{a}\|_q=1} \|\boldsymbol{A}\boldsymbol{a}\|_p = \max_{\|\boldsymbol{a}\|_q\leq 1} \|\boldsymbol{A}\boldsymbol{a}\|_p .$$

For a unit block size setting, $d_1=\cdots=d_K=1$, and $\ell_2$-normalized columns of $\boldsymbol{\Phi}$, as expected, Block-MCC$_{q,p}$ is equivalent to conventional MCC (page 22), i.e., $M_{q,p}(\boldsymbol{\Phi})\equiv M(\boldsymbol{\Phi})$, because each block of $d_k$ columns $\boldsymbol{\Phi}[k]$ will be a single column, i.e., $\boldsymbol{\Phi}[k]=\boldsymbol{\varphi}_k$, and for the vector $\boldsymbol{\varphi}_k$, we have $\boldsymbol{\varphi}_k^{\dagger}=\boldsymbol{\varphi}_k^T$ and for the scalar $\boldsymbol{\varphi}_k^T\boldsymbol{\varphi}_{k'}$, we have $\|\boldsymbol{\varphi}_k^T\boldsymbol{\varphi}_{k'}\|_{q\to p}=|\boldsymbol{\varphi}_k^T\boldsymbol{\varphi}_{k'}|$ for $\forall(q,p)\in\mathbb{R}^2_{\geq 0}$, so $M_{q,p}(\boldsymbol{\Phi})=\max_{k,k'\neq k}|\boldsymbol{\varphi}_k^T\boldsymbol{\varphi}_{k'}|=M(\boldsymbol{\Phi})$.

Although Block-MCC$_{q,p}$ is valid for $\forall\,(q,p)\in\mathbb{R}^2_{>0}$, in practice, because of the computational complexity, only some basic operator-norms can be calculated [Tro04b]. Table 2.1 explains these basic operator-norms.

In order to investigate the bounds and relationship of the proposed Block-MCC$_{q,p}$ to the existing dictionary characterisations, we need to study the properties of the $\ell_{q\to p}$ operator-norm of a matrix, which in turn requires to establish the bounds of division of norm of two vectors:

**Property 2.2 (Bounds of two vector norms division).** $\forall(q,p)\in\mathbb{R}^2_{>0}$, and $\forall\boldsymbol{a}\in\mathbb{R}^d$, we have:

$$\min\left\{1, d^{\frac{1}{p}-\frac{1}{q}}\right\} \leq \frac{\|\boldsymbol{a}\|_p}{\|\boldsymbol{a}\|_q} \leq \max\left\{1, d^{\frac{1}{p}-\frac{1}{q}}\right\}.$$

The proof of Property 2.2 is provided in Section A.1 (page 119). The general bounds of

| | $p=1$ | $p=2$ | $p=\infty$ |
|---|---|---|---|
| $q=1$ | Maximum $\ell_1$ norm of a column | Maximum $\ell_2$ norm of a column | Maximum absolute entry of matrix |
| $q=2$ | NP-hard | Maximum singular value | Maximum $\ell_2$ norm of a row |
| $q=\infty$ | NP-hard | NP-hard | Maximum $\ell_1$ norm of a row |

Table 2.1: Computational complexity of $\ell_{q\to p}$ operator-norm for different basic $(q,p)$ pairs [Tro04b].

Property 2.2 include the following special cases $\forall \boldsymbol{a}\in\mathbb{R}^d$ [GL13]:

$$1 \leq \frac{\|\boldsymbol{a}\|_1}{\|\boldsymbol{a}\|_2} \leq \sqrt{d},$$
$$1 \leq \frac{\|\boldsymbol{a}\|_2}{\|\boldsymbol{a}\|_\infty} \leq \sqrt{d},$$
$$1 \leq \frac{\|\boldsymbol{a}\|_1}{\|\boldsymbol{a}\|_\infty} \leq d.$$

**Property 2.3 (Bounds of two (pseudo-)mixed-norms division).** $\forall (q,p)\in\mathbb{R}^2_{>0}$, for $\boldsymbol{a}=[\boldsymbol{a}[1],\ldots,\boldsymbol{a}[K]]$ and $\forall k$, $\boldsymbol{a}[k]\in\mathbb{R}^{d_k}$, and for weighted (pseudo-)mixed-norms we have:

$$\frac{\|\boldsymbol{a}\|_{\boldsymbol{w};p,1}}{\|\boldsymbol{a}\|_{\boldsymbol{w};q,1}} \geq \min_k \min\left\{1, d_k^{\frac{1}{q}-\frac{1}{p}}\right\} = \begin{cases} 1, & \text{if } 0<q\leq p \\ d_{max}^{\frac{1}{q}-\frac{1}{p}}, & \text{if } 0<p<q \end{cases},$$
$$\frac{\|\boldsymbol{a}\|_{\boldsymbol{w};p,1}}{\|\boldsymbol{a}\|_{\boldsymbol{w};q,1}} \leq \max_k \max\left\{1, d_k^{\frac{1}{q}-\frac{1}{p}}\right\} = \begin{cases} d_{max}^{\frac{1}{q}-\frac{1}{p}}, & \text{if } 0<q\leq p \\ 1, & \text{if } 0<p<q \end{cases}.$$

Similarly, for non-weighted (pseudo-)mixed-norms we have:

$$\frac{\|\boldsymbol{a}\|_{p,1}}{\|\boldsymbol{a}\|_{q,1}} \geq \min_k \min\left\{1, d_k^{\frac{1}{p}-\frac{1}{q}}\right\} = \begin{cases} d_{max}^{\frac{1}{p}-\frac{1}{q}}, & \text{if } 0<q\leq p \\ 1, & \text{if } 0<p<q \end{cases},$$
$$\frac{\|\boldsymbol{a}\|_{p,1}}{\|\boldsymbol{a}\|_{q,1}} \leq \max_k \max\left\{1, d_k^{\frac{1}{p}-\frac{1}{q}}\right\} = \begin{cases} 1, & \text{if } 0<q\leq p \\ d_{max}^{\frac{1}{p}-\frac{1}{q}}, & \text{if } 0<p<q \end{cases}.$$

*Proof.* For any real fractions $x_1/y_1, \ldots, x_K/y_K$ with positive denominators, we have [pah05]:

$$\min\left\{\frac{x_1}{y_1},\ldots,\frac{x_K}{y_K}\right\} \leq \frac{x_1+\ldots+x_K}{y_1+\ldots+y_K} \leq \max\left\{\frac{x_1}{y_1},\ldots,\frac{x_K}{y_K}\right\},$$

where, equality happens if and only if all fractions $x_1/y_1, \ldots, x_K/y_K$ are equal. Then, for any $k$, considering $x_k=\|\boldsymbol{a}[k]\|_p/d_k^{1/p}$ and $y_k=\|\boldsymbol{a}[k]\|_q/d_k^{1/q}$, we have $x_k/y_k=d_k^{1/q-1/p}\|\boldsymbol{a}[k]\|_p/\|\boldsymbol{a}[k]\|_q$. On the other hand, from Property 2.2 (bounds of two vector norms division), we have $\forall(q,p)\in\mathbb{R}^2_{>0}, \forall\boldsymbol{b}\in\mathbb{R}^d$ : $\min\{1,d^{1/p-1/q}\}\leq\|\boldsymbol{b}\|_p/\|\boldsymbol{b}\|_q\leq\max\{1,d^{1/p-1/q}\}$. Consequently, $\min\{1,d_k^{1/q-1/p}\}\leq x_k/y_k\leq\max\{1,d_k^{1/q-1/p}\}$, which by considering the above-mentioned inequality, we have:

$$\min_k \min\left\{1, d_k^{\frac{1}{q}-\frac{1}{p}}\right\} \leq \frac{\sum_k x_k}{\sum_k y_k} \leq \max_k \max\left\{1, d_k^{\frac{1}{q}-\frac{1}{p}}\right\}.$$

On the other hand, from Definition 2.2 (weighted (pseudo-)mixed-norm) we have $\sum_k x_k=\|\boldsymbol{a}\|_{\boldsymbol{w};p,1}$ and $\sum_k y_k=\|\boldsymbol{a}\|_{\boldsymbol{w};q,1}$, which proves the bounds. Similarly, for any $k$, considering $x_k=\|\boldsymbol{a}[k]\|_p$ and $y_k=\|\boldsymbol{a}[k]\|_q$, the bounds for division of two non-weighted (pseudo-)mixed-norms can be obtained. □

**Property 2.4 ($\ell_{q\to p}$ operator-norm properties).** Assuming $\boldsymbol{A}\in\mathbb{R}^{m\times n}$, $\boldsymbol{B}\in\mathbb{R}^{m\times n}$, and $\boldsymbol{C}\in\mathbb{R}^{n\times l}$, the $\ell_{q\to p}$ operator-norm of a matrix satisfies the following properties:

- Nonnegativity: $\forall(q,p)\in\mathbb{R}^2_{\geq 0}: \|\boldsymbol{A}\|_{q\to p}\geq 0$.
- Positivity: $\forall(q,p)\in\mathbb{R}^2_{\geq 0}: \|\boldsymbol{A}\|_{q\to p}=0$ if and only if $\boldsymbol{A}=\boldsymbol{0}$.
- Homogeneity: $\forall q\in\mathbb{R}_{\geq 0}, \forall p\in\mathbb{R}_{>0}, \forall\alpha\in\mathbb{R}: \|\alpha\boldsymbol{A}\|_{q\to p}=|\alpha|\,\|\boldsymbol{A}\|_{q\to p}$.
- Triangle inequality: $\forall q\in\mathbb{R}_{\geq 0}, \forall p\in\mathbb{R}_{\geq 1}$, for $p=0: \|\boldsymbol{A}+\boldsymbol{B}\|_{q\to p}\leq\|\boldsymbol{A}\|_{q\to p}+\|\boldsymbol{B}\|_{q\to p}$.

**Remark 2.4 (Generalised matrix norm).** Any matrix norm definition that satisfies the above four properties is called *generalised matrix norm* [HR12]. All the above-mentioned properties hold true $\forall q\in\mathbb{R}_{\geq 0}$, in contrast to $p$. Table 2.2 summarises the ranges of $p$ in which different properties hold true. As it can be seen in table 2.2, for $\forall q\in\mathbb{R}_{\geq 0}$ and $\forall p\in\mathbb{R}_{\geq 1}$, the $\ell_{q\to p}$ operator-norm satisfies all the above four properties, hence it is a generalised matrix norm in the mentioned range for $q$ and $p$.

- Submultiplicativity: In general, we have:

$$\forall(q,p)\in\mathbb{R}^2_{>0}, \qquad \|\boldsymbol{A}\,\boldsymbol{C}\|_{q\to p}\leq\|\boldsymbol{A}\|_{q\to p}\,\|\boldsymbol{C}\|_{q\to p}\max\left\{1, n^{\frac{1}{q}-\frac{1}{p}}\right\}.$$

  For $q\geq p$ we have $\max\{1, n^{1/q-1/p}\}=1$, and then there exists submultiplicativity property, i.e., for $q\geq p>0$, we have $\|\boldsymbol{A}\,\boldsymbol{C}\|_{q\to p}\leq\|\boldsymbol{A}\|_{q\to p}\,\|\boldsymbol{C}\|_{q\to p}$.

**Remark 2.5 (Matrix norm).** Any norm definition that satisfies the above five properties is called *matrix norm* [HR12]. The $\ell_{q\to p}$ operator-norm $\forall q\in\mathbb{R}_{\geq 0}$ and $\forall p\in\mathbb{R}_{\geq 1}$ satisfies the first four properties (generalised matrix norm), whereas for $q\geq p>0$ satisfies the fifth property. Then, $\ell_{q\to p}$ operator-norm for $q\geq p\geq 1$ satisfies all the above five properties, hence it is a matrix norm.

- Bounds: We define the following four types of bounds and inequalities for the $\ell_{q\to p}$ operator-norm:
  1) $\forall\,(q,p,q',p')\in\mathbb{R}^4_{>0}$, we have:

$$\|\boldsymbol{A}\|_{q\to p}\geq\max\left\{\min\left\{1, m^{\frac{1}{p}-\frac{1}{p'}}\right\}\|\boldsymbol{A}\|_{q\to p'}, \min\left\{1, n^{\frac{1}{q'}-\frac{1}{q}}\right\}\|\boldsymbol{A}\|_{q'\to p}\right\},$$
$$\|\boldsymbol{A}\|_{q\to p}\leq\min\left\{\max\left\{1, m^{\frac{1}{p}-\frac{1}{p'}}\right\}\|\boldsymbol{A}\|_{q\to p'}, \max\left\{1, n^{\frac{1}{q'}-\frac{1}{q}}\right\}\|\boldsymbol{A}\|_{q'\to p}\right\},$$

| $\forall q\in\mathbb{R}_{\geq 0}$ | $p=0$ | $0<p<1$ | $p\geq 1$ |
|---|---|---|---|
| $\ell_{q\to p}$ properties | N, P, T | N, P, H | N, P, H, T |

Table 2.2: Properties of $\ell_{q\to p}$ operator-norm for different ranges of $p$, while $\forall q\in\mathbb{R}_{\geq 0}$, where N, P, H, and T stand for the existence of nonnegativity, positivity, homogeneity, and triangle inequality properties, respectively.

| | | | |
|---|---|---|---|
| $1 \le \frac{\|\boldsymbol{A}\|_{1\to1}}{\|\boldsymbol{A}\|_{1\to2}} \le m^{\frac{1}{2}},$ | $1 \le \frac{\|\boldsymbol{A}\|_{1\to1}}{\|\boldsymbol{A}\|_{1\to\infty}} \le m,$ | $n^{-\frac{1}{2}} \le \frac{\|\boldsymbol{A}\|_{1\to1}}{\|\boldsymbol{A}\|_{2\to2}} \le m^{\frac{1}{2}},$ | $n^{-\frac{1}{2}} \le \frac{\|\boldsymbol{A}\|_{1\to1}}{\|\boldsymbol{A}\|_{2\to\infty}} \le m,$ |
| $n^{-1} \le \frac{\|\boldsymbol{A}\|_{1\to1}}{\|\boldsymbol{A}\|_{\infty\to\infty}} \le m,$ | $1 \le \frac{\|\boldsymbol{A}\|_{2\to2}}{\|\boldsymbol{A}\|_{1\to2}} \le n^{\frac{1}{2}},$ | $1 \le \frac{\|\boldsymbol{A}\|_{2\to2}}{\|\boldsymbol{A}\|_{1\to\infty}} \le (m\,n)^{\frac{1}{2}},$ | $1 \le \frac{\|\boldsymbol{A}\|_{2\to2}}{\|\boldsymbol{A}\|_{2\to\infty}} \le m^{\frac{1}{2}},$ |
| $n^{-\frac{1}{2}} \le \frac{\|\boldsymbol{A}\|_{2\to2}}{\|\boldsymbol{A}\|_{\infty\to\infty}} \le m^{\frac{1}{2}},$ | $m^{-\frac{1}{2}} \le \frac{\|\boldsymbol{A}\|_{\infty\to\infty}}{\|\boldsymbol{A}\|_{1\to2}} \le n,$ | $1 \le \frac{\|\boldsymbol{A}\|_{\infty\to\infty}}{\|\boldsymbol{A}\|_{1\to\infty}} \le n,$ | $1 \le \frac{\|\boldsymbol{A}\|_{\infty\to\infty}}{\|\boldsymbol{A}\|_{2\to\infty}} \le n^{\frac{1}{2}},$ |
| $1 \le \frac{\|\boldsymbol{A}\|_{1\to2}}{\|\boldsymbol{A}\|_{1\to\infty}} \le m^{\frac{1}{2}},$ | $n^{-\frac{1}{2}} \le \frac{\|\boldsymbol{A}\|_{1\to2}}{\|\boldsymbol{A}\|_{2\to\infty}} \le m^{\frac{1}{2}},$ | $n^{-\frac{1}{2}} \le \frac{\|\boldsymbol{A}\|_{1\to\infty}}{\|\boldsymbol{A}\|_{2\to\infty}} \le 1.$ | |

Table 2.3: Inequalities of basic tractable $\ell_{q\to p}$ operator-norms based on table 2.1 (page 54) for a matrix $\boldsymbol{A}\in\mathbb{R}^{m\times n}$.

which the bounds are based on the operator norm having one of the original domains, either $q$, or $p$.

2) In addition, the bounds can be based on the operator norm having totally new domains, e.g., $q'$ and $p'$, i.e., $\forall\,(q,p,q',p')\in\mathbb{R}^4_{>0}$ we have:

$$\|\boldsymbol{A}\|_{q\to p} \ge \min\left\{1, m^{\frac{1}{p}-\frac{1}{p'}}\right\} \min\left\{1, n^{\frac{1}{q'}-\frac{1}{q}}\right\} \|\boldsymbol{A}\|_{q'\to p'}\,,$$
$$\|\boldsymbol{A}\|_{q\to p} \le \max\left\{1, m^{\frac{1}{p}-\frac{1}{p'}}\right\} \max\left\{1, n^{\frac{1}{q'}-\frac{1}{q}}\right\} \|\boldsymbol{A}\|_{q'\to p'}\,.$$

**Remark 2.6 ($\ell_{q\to p}$ operator-norm inequalities).** The above-mentioned general inequalities for basic tractable $\ell_{q\to p}$ operator-norms based on table 2.1 (page 54) is shown in table 2.3, which includes the following standard inequalities [GL13]:

$$\|\boldsymbol{A}\|_{1\to\infty} \le \|\boldsymbol{A}\|_{2\to2} \le \sqrt{mn}\,\|\boldsymbol{A}\|_{1\to\infty}\,,$$
$$\frac{1}{\sqrt{n}}\,\|\boldsymbol{A}\|_{\infty\to\infty} \le \|\boldsymbol{A}\|_{2\to2} \le \sqrt{m}\,\|\boldsymbol{A}\|_{\infty\to\infty}\,,$$
$$\frac{1}{\sqrt{m}}\,\|\boldsymbol{A}\|_{1\to1} \le \|\boldsymbol{A}\|_{2\to2} \le \sqrt{n}\,\|\boldsymbol{A}\|_{1\to1}\,.$$

In addition, in general (tractable and intractable), the $\ell_{q\to p}$ operator-norm inequalities for a fixed $p$ or $q$ is shown schematically in figure 2.3.

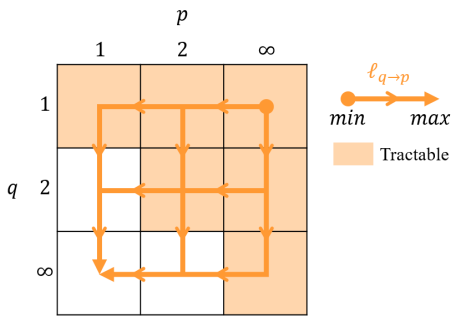



Figure 2.3: $\ell_{q\to p}$ operator-norm inequalities for common $\ell_{q\to p}$ operator-norms according to table 2.1 (page 54).

3) Another useful lower- and upper-bound $\forall\,(q,p,q',p')\in\mathbb{R}^4_{>0}$ are:

$$\|\boldsymbol{A}\|_{q\to p}\geq\frac{\min\left\{1,m^{\frac{1}{p}-\frac{1}{p'}}\right\}\min\left\{1,n^{\frac{1}{q'}-\frac{1}{q}}\right\}\min\left\{1,m^{\frac{1}{p'}-\frac{1}{2}}\right\}\min\left\{1,n^{\frac{1}{2}-\frac{1}{q'}}\right\}}{\sqrt{\min\{m,n\}}}\|\boldsymbol{A}\|_F\,,$$

$$\|\boldsymbol{A}\|_{q\to p}\leq\max\left\{1,m^{\frac{1}{p}-\frac{1}{p'}}\right\}\max\left\{1,n^{\frac{1}{q'}-\frac{1}{q}}\right\}\max\left\{1,m^{\frac{1}{p'}-\frac{1}{2}}\right\}\max\left\{1,n^{\frac{1}{2}-\frac{1}{q'}}\right\}\|\boldsymbol{A}\|_F\,,$$

where, the Frobenius norm is defined as $\|\boldsymbol{A}\|_F=\sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n}|a_{i,j}|^2}$.

4) In addition, the $\ell_{q\to p}$ operator-norm of any matrix $\boldsymbol{A}$ is less than or equal to the $\ell_{q\to p}$ operator-norm of another matrix $\boldsymbol{B}$, in which all the elements are the maximum absolute value of the elements of first matrix. It also holds true, when all the on-diagonal entries of $\boldsymbol{A}$ and $\boldsymbol{B}$ are set to zero:

$$\forall i,j,\forall\,(q,p)\in\mathbb{R}^2_{>0},\qquad\text{if }\ |a_{i,j}|\leq b_{i,j}=\max_{i,j}|a_{i,j}|\Rightarrow\|\boldsymbol{A}\|_{q\to p}\leq\|\boldsymbol{B}\|_{q\to p}\,,$$

and

$$\forall i,j,\forall\,(q,p)\in\mathbb{R}^2_{>0},\qquad\text{if }\begin{cases}|a_{i,j}|\leq b_{i,j}=\max\limits_{i,j}|a_{i,j}|\,, & i\neq j\\ a_{i,j}=b_{i,j}=0, & i=j\end{cases}\Rightarrow\|\boldsymbol{A}\|_{q\to p}\leq\|\boldsymbol{B}\|_{q\to p}\,.$$

See Section A.2 (page 121) for the proof of Property 2.4.

Now, based on the above-mentioned properties of the $\ell_{q\to p}$ operator-norm introduced in Property 2.4 ($\ell_{q\to p}$ operator-norm properties), in the following property we investigate the possible relationship between different Block-MCC$_{q,p}$ characterisations (Definition 2.6, page 54) with basic $(q,p)$ pairs according to table 2.1 (page 54):

**Property 2.5 (Block-MCC$_{q,p}$ inequalities).** The different Block-MCC$_{q,p}$ characterisations calculated for basic tractable $\ell_{q\to p}$ operator-norms of table 2.1 (page 54) have the following relationships:

$$\begin{aligned}M_{1,1}(\boldsymbol{\Phi})\leq M_{1,2}(\boldsymbol{\Phi})\leq M_{1,\infty}(\boldsymbol{\Phi})\,,\\ M_{2,2}(\boldsymbol{\Phi})\leq M_{1,2}(\boldsymbol{\Phi})\leq M_{1,\infty}(\boldsymbol{\Phi})\,,\\ M_{2,2}(\boldsymbol{\Phi})\leq M_{2,\infty}(\boldsymbol{\Phi})\leq M_{1,\infty}(\boldsymbol{\Phi})\,,\\ M_{\infty,\infty}(\boldsymbol{\Phi})\leq M_{2,\infty}(\boldsymbol{\Phi})\leq M_{1,\infty}(\boldsymbol{\Phi})\,,\end{aligned}$$

while their general relationship is represented schematically in figure 2.4.

The proof of Property 2.5 is provided in Section A.3 (page 126).

Since the recovery conditions started with orthonormal bases, the following property shows Definition 2.6 (block coherence) in special case of intra-block orthonormality:

**Property 2.6 (Block-MCC$_{q,p}$ for intra-block orthonormality).** If the dictionary $\boldsymbol{\Phi}$ has intra-block orthonormality, i.e., for $1\leq k\leq K$, $\boldsymbol{\Phi}^T[k]\boldsymbol{\Phi}[k]=\boldsymbol{I}_{d_k}$, then $\forall(q,p)\in\mathbb{R}^2_{>0}$ we have:

$$M_{q,p}(\boldsymbol{\Phi})=\max_{k,k'\neq k}\frac{d_k^{-\frac{1}{p}}\,d_{k'}^{\frac{1}{q}}}{d_{max}}\left\|\boldsymbol{\Phi}^T[k]\,\boldsymbol{\Phi}\left[k'\right]\right\|_{q\to p}\,.$$

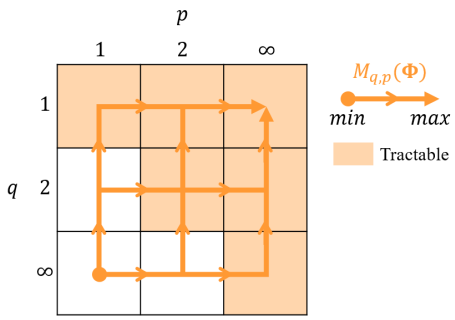



Figure 2.4: Block-MCC$_{q,p}$ inequalities for common $\ell_{q\to p}$ operator-norms according to table 2.1 (page 54).

*Proof.* Taking into account that, for an orthonormal matrix we have $\boldsymbol{A}^{\dagger}=\boldsymbol{A}^{T}$, the result is immediately obtained from the Definition 2.6 (block coherence). □

Block-MCC$_{q,p}$ is a measure of similarity between different blocks of the columns of the dictionary, which is the block-wise extension of the conventional MCC (page 22). Since the dictionary coherence has an inverse effect on the theoretical recovery conditions, the lower bounds for coherence is desired. When a dictionary has intra-block orthonormality, Block-MCC$_{q,p}$ can be bounded, as shown in the following property:

**Property 2.7 (Block-MCC$_{q,p}$ upper-bound with intra-block orthonormality I).** For a dictionary $\boldsymbol{\Phi}$ with intra-block orthonormality, i.e., for $1\leq k\leq K$, $\boldsymbol{\Phi}^{T}[k]\boldsymbol{\Phi}[k]=\boldsymbol{I}_{d_k}$, the upper-bound of Block-MCC$_{q,p}$ is shown in table 2.4.

The proof of Property 2.7 is provided in Section A.4 (page 129). Later, in Property 2.9, we will show another upper-bounds for Block-MCC$_{q,p}$ based on the relationship with conventional MCC. Property 2.8 investigates the bounds of Block-MCC$_{q,p}$ in terms of the conventional MCC (page 22).

**Property 2.8 (Block-MCC$_{q,p}$ bounds).** Block-MCC$_{q,p}$ is bounded based on the conventional MCC, i.e., $M(\boldsymbol{\Phi})\overset{\text{def}}{=}\max_{k,k'\neq k}|\langle\boldsymbol{\varphi}_k,\boldsymbol{\varphi}_{k'}\rangle|$.
1) Suppose for a dictionary with full column rank blocks, we have $M(\boldsymbol{\Phi})<d_{max}^{1/q-2}(d_{max}-1)^{-1/2}$

| $(q,p)$ | $(1,1)$ | $(1,2)$ | $(1,\infty)$ | $(2,2)$ | $(2,\infty)$ | $(\infty,\infty)$ |
|---|---|---|---|---|---|---|
| $\max M_{q,p}(\boldsymbol{\Phi})$ | $d_{min}^{-\frac{1}{2}}\,m^{\frac{1}{2}}$ | $d_{min}^{-\frac{1}{2}}\,m^{\frac{1}{2}}$ | $m$ | $d_{min}^{-\frac{1}{2}}\,d_{max}^{-\frac{1}{2}}$ | $d_{max}^{-\frac{1}{2}}\,m^{\frac{1}{2}}$ | $d_{max}^{-\frac{1}{2}}\,m^{\frac{1}{2}}$ |

Table 2.4: Upper-bound of Block-MCC$_{q,p}$ obtained based on the relationship with unit $\ell_{2\to 2}$ operator-norm, for different basic $(q,p)$ pairs and for a dictionary $\boldsymbol{\Phi}\in\mathbb{R}^{m\times n}$ with intra-block orthonormality.

| $(q=q', p=p')$ | $(1,1)$ | $(1,2)$ | $(1,\infty)$ | $(2,2)$ | $(2,\infty)$ | $(\infty,\infty)$ |
|---|---|---|---|---|---|---|
| $\max M_{q,p}(\mathbf{\Phi})$ | $d_{max}^{\frac{1}{2}}$ | $d_{max}^{\frac{1}{2}}$ | $d_{max}$ | $1$ | $d_{max}^{\frac{1}{2}}$ | $d_{max}^{\frac{1}{2}}$ |

Table 2.5: Upper-bound of Block-MCC$_{q,p}$ obtained based on the relationship with MCC, for different basic values of $(q,p)$ pairs and for a dictionary with intra-block orthonormality.

and $q{\geq}p{\geq}1$, then:

$$0 \leq M_{q,p}\left(\mathbf{\Phi}\right) \leq \frac{d_{max}^{\frac{3}{2}-\frac{1}{p}}\, M\left(\mathbf{\Phi}\right)}{1 - d_{max}^{2-\frac{1}{q}}\left(d_{max}-1\right)^{\frac{1}{2}}\, M\left(\mathbf{\Phi}\right)}.$$

2) For a dictionary with intra-block orthonormality, $\forall(q,p,q',p'){\in}\mathbb{R}^4_{>0}$ we have:

$$0 \leq M_{q,p}\left(\mathbf{\Phi}\right) \leq \frac{M\left(\mathbf{\Phi}\right)}{d_{max}} \max_{k,k'\neq k} d_k^{\frac{1}{2}-\frac{1}{p}}\, d_{k'}^{\frac{1}{q}+\frac{1}{2}}\, \max\left\{1, d_k^{\frac{1}{p}-\frac{1}{p'}}\right\} \max\left\{1, d_{k'}^{\frac{1}{q'}-\frac{1}{q}}\right\} \times \max\left\{1, d_k^{\frac{1}{p'}-\frac{1}{2}}\right\} \max\left\{1, d_{k'}^{\frac{1}{2}-\frac{1}{q'}}\right\}.$$

The proof of Property 2.8 is provided in Section A.5 (page 131). In addition, we have:

**Property 2.9 (Block-MCC$_{q,p}$ upper-bound with intra-block orthonormality II).** For a dictionary with orthonormal blocks, Block-MCC$_{q,p}$ is upper-bounded based on its relationship with MCC, i.e., $\forall(q,p,q',p'){\in}\mathbb{R}^4_{>0}$, we have:

$$0 \leq M_{q,p}\left(\mathbf{\Phi}\right) \leq d_{max}^{-1} \max_{k,k'\neq k} d_k^{\frac{1}{2}-\frac{1}{p}}\, d_{k'}^{\frac{1}{q}+\frac{1}{2}}\, \max\left\{1, d_k^{\frac{1}{p}-\frac{1}{p'}}\right\} \max\left\{1, d_{k'}^{\frac{1}{q'}-\frac{1}{q}}\right\} \times \max\left\{1, d_k^{\frac{1}{p'}-\frac{1}{2}}\right\} \max\left\{1, d_{k'}^{\frac{1}{2}-\frac{1}{q'}}\right\}.$$

The mentioned upper-bound of Block-MCC$_{q,p}$ for basic tractable $(q,p)$ pairs is shown in table 2.5.

*Proof.* Since MCC is upper-bounded by one, from Property 2.8 the proof is done. □

Notice that in Property 2.7 we introduced the bounds of Block-MCC$_{q,p}$ based on the bounds of the operator-norms in terms of the unit $\ell_{2\to2}$ operator-norm, whereas now in Property 2.9 we introduced another upper-bound based on its relationship with MCC and considering the upper-bound of MCC. By minimising upper-bounds of Block-MCC$_{q,p}$ in properties 2.7 (table 2.4) and 2.9 (table 2.5), the final resulted upper-bounds are shown in table 2.6.

**Remark 2.7.** By comparing the bounds of conventional MCC, i.e., $1/\sqrt{m}{\leq}M(\mathbf{\Phi}){\leq}1$ (equation (1.7), page 22), with the upper-bounds of the proposed dictionary coherence in table 2.6, where, there is intra-block orthonormality and considering $d_{min}{=}d_{max}{=}m$ (orthonormal complete bases), it can be seen that, although the obtained upper-bound for Block-MCC$_{1,\infty}$ is

| $(q,p)$ | $(1,1)$ and $(1,2)$ | $(1,\infty)$ | $(2,2)$ | $(2,\infty)$ and $(\infty,\infty)$ |
|---|---|---|---|---|
| $\max M_{q,p}(\boldsymbol{\Phi})$ | $\min\left\{d_{min}^{-\frac{1}{2}}\, m^{\frac{1}{2}}, d_{max}^{\frac{1}{2}}\right\}$ | $\min\{m, d_{max}\}$ | $d_{min}^{-\frac{1}{2}}\, d_{max}^{-\frac{1}{2}}$ | $\min\left\{d_{max}^{-\frac{1}{2}}\, m^{\frac{1}{2}}, d_{max}^{\frac{1}{2}}\right\}$ |

Table 2.6: Upper-bound of Block-MCC$_{q,p}$ obtained based on properties 2.7 (table 2.4) and 2.9 (table 2.5), for different basic values of $(q,p)$ pairs and for a dictionary $\boldsymbol{\Phi}\in\mathbb{R}^{m\times n}$ with intra-block orthonormality.

greater than that of conventional MCC ($m$ v.s. 1), but both characterisations have the same upper-bound for $(q,p)$ pairs of $(1,1)$, $(1,2)$, $(2,\infty)$, and $(\infty,\infty)$, while the upper-bound for $(2,2)$ is $1/m$, which is much less than the conventional unit upper-bound, and even less than the lower-bound of conventional MCC ($1/m$ v.s. $1/\sqrt{m}$).

If MCC is small enough, the dictionary is said to be *incoherent*, similarly if Block-MCC$_{q,p}$ is small enough, the dictionary will be *block-incoherent*. The term block-incoherent is previously used in some studies, of course with a different definition for the characterisation of the coherence of the dictionary which is called *block-coherence* [PV07]; [EB09]; [EKB10b].

Next, we demonstrate the relationship between the proposed general Block-MCC$_{q,p}$ and the conventional characterisation of Donoho and Huo, i.e., $\tilde{M}(\boldsymbol{\Phi_1},\boldsymbol{\Phi_2})\overset{\text{def}}{=}\max\{\max_{i,j}|\boldsymbol{\Phi_1}^{-1}\boldsymbol{\Phi_2}|_{i,j}, \max_{i,j}|\boldsymbol{\Phi_2}^{-1}\boldsymbol{\Phi_1}|_{i,j}\}$ explained on page 28 [DH01], in sparsity, and block-coherence of Eldar et al., i.e., $M_{Inter}^{Eldar}(\boldsymbol{\Phi})\overset{\text{def}}{=}\max_{k,k'\neq k}\|\boldsymbol{\Phi}^T[k]\boldsymbol{\Phi}[k']\|_{2\to 2}/d$ explained on page 36 equation (1.17) [EKB10a], in block-sparsity domain.

**Property 2.10 (Block-MCC$_{q,p}$ v.s. coherence of Donoho and Huo).** For a dictionary $\boldsymbol{\Phi}=[\boldsymbol{\Phi_1},\boldsymbol{\Phi_2}]$ with two invertible square blocks of same size, $d_1=d_2=m$, we have:

$$M_{1,\infty}(\boldsymbol{\Phi}) = \tilde{M}(\boldsymbol{\Phi_1},\boldsymbol{\Phi_2}),$$

and

$$\{M_{1,1}(\boldsymbol{\Phi}), M_{1,2}(\boldsymbol{\Phi}), M_{2,2}(\boldsymbol{\Phi}), M_{2,\infty}(\boldsymbol{\Phi}), M_{\infty,\infty}(\boldsymbol{\Phi})\} \leq \tilde{M}(\boldsymbol{\Phi_1},\boldsymbol{\Phi_2}),$$

where, $\tilde{M}(\boldsymbol{\Phi_1},\boldsymbol{\Phi_2})$ is the dictionary characterisation of Donoho and Huo, i.e., $\tilde{M}(\boldsymbol{\Phi_1},\boldsymbol{\Phi_2})\overset{\text{def}}{=}\max\{\max_{i,j}|\boldsymbol{\Phi_1}^{-1}\boldsymbol{\Phi_2}|_{i,j}, \max_{i,j}|\boldsymbol{\Phi_2}^{-1}\boldsymbol{\Phi_1}|_{i,j}\}$, and $M_{q,p}(\boldsymbol{\Phi})$ is the proposed Block-MCC$_{q,p}$, i.e., $M_{q,p}(\boldsymbol{\Phi})\overset{\text{def}}{=}d_{max}^{-1}\max_{k,k'\neq k} d_k^{-1/p} d_{k'}^{1/q}\|\boldsymbol{\Phi}^\dagger[k]\boldsymbol{\Phi}[k']\|_{q\to p}$.

*Proof.* Considering that $\ell_{1\to\infty}$ operator-norm represents the maximum absolute value of matrix, i.e., $\|\boldsymbol{A}\|_{1\to\infty}=\max_{i,j}|a_{i,j}|$ (table 2.1, page 54), and the fact that here $\boldsymbol{\Phi}[k]=\boldsymbol{\Phi_k}$, the definition $\tilde{M}(\boldsymbol{\Phi_1},\boldsymbol{\Phi_2})\overset{\text{def}}{=}\max\{\max_{i,j}|\boldsymbol{\Phi_1}^{-1}\boldsymbol{\Phi_2}|_{i,j}, \max_{i,j}|\boldsymbol{\Phi_2}^{-1}\boldsymbol{\Phi_1}|_{i,j}\}$ (page 28) can be rewritten as $\tilde{M}(\boldsymbol{\Phi_1},\boldsymbol{\Phi_2})=\max_{k,k'\neq k}\|\boldsymbol{\Phi}^{-1}[k]\boldsymbol{\Phi}[k']\|_{1\to\infty}$. On the other hand for the invertible block $\boldsymbol{\Phi}[k]$, we have $\boldsymbol{\Phi}^{-1}[k]=\boldsymbol{\Phi}^\dagger[k]$, so $\tilde{M}(\boldsymbol{\Phi_1},\boldsymbol{\Phi_2})=\max_{k,k'\neq k}\|\boldsymbol{\Phi}^\dagger[k]\boldsymbol{\Phi}[k']\|_{1\to\infty}$. On the other hand, $M_{1,\infty}(\boldsymbol{\Phi})=d_{max}^{-1}\max_{k,k'\neq k} d_{k'}\|\boldsymbol{\Phi}^\dagger[k]\boldsymbol{\Phi}[k']\|_{1\to\infty}=\max_{k,k'\neq k}\|\boldsymbol{\Phi}^\dagger[k]\boldsymbol{\Phi}[k']\|_{1\to\infty}$, because $\forall k, d_k=d_{max}$ ($d_1=d_2=m$). Then, $M_{1,\infty}(\boldsymbol{\Phi})=\tilde{M}(\boldsymbol{\Phi_1},\boldsymbol{\Phi_2})$. In addition from Property 2.5 (Block-MCC$_{q,p}$ inequalities), we know that $M_{1,\infty}(\boldsymbol{\Phi})$ upper-bounds the other basic tractable $(q,p)$ pairs of table 2.1 (page 54), hence, the proof for the second part is done. □

**Property 2.11 (Block-MCC$_{q,p}$ v.s. block-coherence of Eldar et al.).** For a dictionary with equally-sized blocks of length $d$ and intra-block orthonormality, $\forall (q,p)\in\mathbb{R}^2_{>0}$ we have:

$$M_{q,p}(\boldsymbol{\Phi}) \geq d^{\frac{1}{q}-\frac{1}{p}} \min\left\{1, d^{\frac{1}{p}-\frac{1}{2}}\right\} \min\left\{1, d^{\frac{1}{2}-\frac{1}{q}}\right\} M_{Inter}^{Eldar}(\boldsymbol{\Phi}),$$
$$M_{q,p}(\boldsymbol{\Phi}) \leq d^{\frac{1}{q}-\frac{1}{p}} \max\left\{1, d^{\frac{1}{p}-\frac{1}{2}}\right\} \max\left\{1, d^{\frac{1}{2}-\frac{1}{q}}\right\} M_{Inter}^{Eldar}(\boldsymbol{\Phi}),$$

where, $M_{Inter}^{Eldar}(\boldsymbol{\Phi})$ is the block-coherence of Eldar et al., i.e., $M_{Inter}^{Eldar}(\boldsymbol{\Phi})\stackrel{\text{def}}{=}\max_{k,k'\neq k}\|\boldsymbol{\Phi}^T[k]\boldsymbol{\Phi}[k']\|_{2\to 2}/d$, and $M_{q,p}(\boldsymbol{\Phi})$ is the proposed Block-MCC$_{q,p}$, i.e., $M_{q,p}(\boldsymbol{\Phi})\stackrel{\text{def}}{=}d_{max}^{-1}\max_{k,k'\neq k} d_k^{-1/p} d_{k'}^{1/q}\|\boldsymbol{\Phi}^\dagger[k]\boldsymbol{\Phi}[k']\|_{q\to p}$. Then, for $q{=}p{=}2$, we have $M_{2,2}(\boldsymbol{\Phi}){=}M_{Inter}^{Eldar}(\boldsymbol{\Phi})$.

*Proof.* From Property 2.6 (Block-MCC$_{q,p}$ for intra-block orthonormality; page 58) in a special setting of $d_1{=}\cdots{=}d_K{=}d$, we have $\forall(q,p)\in\mathbb{R}^2_{>0}$ : $M_{q,p}(\boldsymbol{\Phi}){=}d^{1/q-1/p-1}\max_{k,k'\neq k}\|\boldsymbol{\Phi}^T[k]\boldsymbol{\Phi}[k']\|_{q\to p}$. Then, from Property 2.4 ($\ell_{q\to p}$ operator-norm properties; second set of bounds), we can determine the bounds of $\ell_{q\to p}$ operator-norm based on $\ell_{q'\to p'}$ operator-norm, i.e., $\forall \boldsymbol{A}\in\mathbb{R}^{m\times n}, \forall\,(q,p,q',p')\in\mathbb{R}^4_{>0}$ : $\|\boldsymbol{A}\|_{q\to p}\geq(\leq)\min(\max)\{1, m^{1/p-1/p'}\}\min(\max)\{1, n^{1/q'-1/q}\}\|\boldsymbol{A}\|_{q'\to p'}$, and by selecting $q'{=}p'{=}2$, we have:

$$M_{q,p}(\boldsymbol{\Phi}) \geq d^{\frac{1}{q}-\frac{1}{p}-1} \min\left\{1, d^{\frac{1}{p}-\frac{1}{2}}\right\} \min\left\{1, d^{\frac{1}{2}-\frac{1}{q}}\right\} \max_{k,k'\neq k} \left\|\boldsymbol{\Phi}^T[k]\,\boldsymbol{\Phi}\left[k'\right]\right\|_{2\to 2},$$
$$M_{q,p}(\boldsymbol{\Phi}) \leq d^{\frac{1}{q}-\frac{1}{p}-1} \max\left\{1, d^{\frac{1}{p}-\frac{1}{2}}\right\} \max\left\{1, d^{\frac{1}{2}-\frac{1}{q}}\right\} \max_{k,k'\neq k} \left\|\boldsymbol{\Phi}^T[k]\,\boldsymbol{\Phi}\left[k'\right]\right\|_{2\to 2}.$$

Next, considering the definition $M_{Inter}^{Eldar}(\boldsymbol{\Phi})\stackrel{\text{def}}{=}\max_{k,k'\neq k}\|\boldsymbol{\Phi}^T[k]\boldsymbol{\Phi}[k']\|_{2\to 2}/d$ presented on page 36 equation (1.17), the proof is done. □

The above introduced coherence constant Block-MCC$_{q,p}$ characterises the dictionary by means of the maximum coherence, but with the expense of heavier computations it is possible to better characterise it using a cumulative coherence.

**Definition 2.7 (Cumulative Block-MCC$_{q,p}$).** The *cumulative Block-MCC$_{q,p}$* of a dictionary with full column rank blocks, is defined for all integers $1{\leq}k{\leq}K{-}1$ and $\forall(q,p)\in\mathbb{R}^2_{>0}$ as:

$$M_{q,p}(\boldsymbol{\Phi},k) \stackrel{\text{def}}{=} \max_{|\Lambda|=k}\max_{j\notin\Lambda}\sum_{i\in\Lambda}\frac{d_i^{-\frac{1}{p}} d_j^{\frac{1}{q}}}{d_{max}}\left\|\boldsymbol{\Phi}^\dagger[i]\,\boldsymbol{\Phi}[j]\right\|_{q\to p},$$

where, $\Lambda{\subset}\{1,\cdots,K\}$ of Cardinality $k$.

Cumulative Block-MCC$_{q,p}$ measures the maximum total Block-MCC$_{q,p}$ between a fixed block and a collection of other blocks, which is the block-wise generalisation of the conventional cumulative MCC defined in (1.8), i.e., $M(\boldsymbol{\Phi},k)\stackrel{\text{def}}{=}\max_{|\Lambda|=k}\max_{j\notin\Lambda}\sum_{i\in\Lambda}|\langle\boldsymbol{\varphi}_i,\boldsymbol{\varphi}_j\rangle|$.

As expected, for a unit block size scenario, i.e., $d_1{=}\cdots{=}d_K{=}1$, cumulative Block-MCC$_{q,p}$ in Definition 2.7 is equivalent to the conventional cumulative MCC in (1.8) presented on page 23. The following property shows the cumulative Block-MCC$_{q,p}$ properties, which are similar to the cumulative MCC's one presented on page 24:

**Property 2.12 (Cumulative Block-MCC$_{q,p}$ properties).** For any dictionary $\mathbf{\Phi}$ with full column rank blocks, and Block-MCC$_{q,p}$ $M_{q,p}(\mathbf{\Phi})$, $\forall(q,p){\in}\mathbb{R}^2_{>0}$ we have:

$$M_{q,p}\left(\mathbf{\Phi},1\right)=M_{q,p}\left(\mathbf{\Phi}\right),$$

and

$$M_{q,p}\left(\mathbf{\Phi},k\right)\leq k\,M_{q,p}\left(\mathbf{\Phi}\right).$$

*Proof.* An investigation of the formula of cumulative Block-MCC$_{q,p}$, shows that $\forall(q,p){\in}\mathbb{R}^2_{>0}$ and for $k{=}1$, it reduces to Block-MCC$_{q,p}$, i.e.,:

$$M_{q,p}\left(\mathbf{\Phi},1\right)=\max_{|\Lambda|=1}\max_{j\notin\Lambda}\sum_{i\in\Lambda}\frac{d_i^{-\frac{1}{p}}d_j^{\frac{1}{q}}}{d_{max}}\left\|\mathbf{\Phi}^{\dagger}\left[i\right]\mathbf{\Phi}\left[j\right]\right\|_{q\to p}=\max_{i}\max_{j\neq i}\frac{d_i^{-\frac{1}{p}}d_j^{\frac{1}{q}}}{d_{max}}\left\|\mathbf{\Phi}^{\dagger}\left[i\right]\mathbf{\Phi}\left[j\right]\right\|_{q\to p}=M_{q,p}\left(\mathbf{\Phi}\right).$$

To prove the second part, similar to the proof of Proposition 2.1 in [Tro04a], $\forall(q,p){\in}\mathbb{R}^2_{>0}$ we have:

$$M_{q,p}\left(\mathbf{\Phi},k\right)=\max_{|\Lambda|=k}\max_{j\notin\Lambda}\sum_{i\in\Lambda}\frac{d_i^{-\frac{1}{p}}d_j^{\frac{1}{q}}}{d_{max}}\left\|\mathbf{\Phi}^{\dagger}\left[i\right]\mathbf{\Phi}\left[j\right]\right\|_{q\to p}\leq\max_{|\Lambda|=k}\sum_{i\in\Lambda}M_{q,p}\left(\mathbf{\Phi}\right)=k\,M_{q,p}\left(\mathbf{\Phi}\right).$$

□

According to the inter-block coherence defined by Eldar et al. in (1.17) page 36, i.e., $M_{Inter}^{Eldar}(\mathbf{\Phi})\overset{\text{def}}{=}\max_{k,k'\neq k}\|\mathbf{\Phi}^T[k]\mathbf{\Phi}[k']\|_{2\to2}/d$, [EB09]; [EKB10b]; [EKB10a], and the concept of cumulative coherence introduced by Tropp [Tro04a] (page 24), we extend the inter-block coherence and introduce the following definition:

**Definition 2.8 (Eldar's cumulative coherence).** The cumulative inter-block coherence constant of a general dictionary $\mathbf{\Phi}$ is defined for all integers $1{\leq}k{\leq}K{-}1$ as:

$$M_{Inter}^{Eldar}\left(\mathbf{\Phi},k\right)\overset{\text{def}}{=}\max_{|\Lambda|=k}\max_{j\notin\Lambda}\sum_{i\in\Lambda}\frac{1}{d}\left\|\mathbf{\Phi}^{T}\left[i\right]\mathbf{\Phi}\left[j\right]\right\|_{2\to2},$$

where, $\Lambda{\subset}\{1,\cdots,K\}$ is of Cardinality $k$ and $d_1{=}\cdots{=}d_K{=}d$.

As expected, for $d{=}1$, cumulative inter-block coherence constant in Definition 2.8 is equivalent to the conventional cumulative MCC defined in (1.8) presented on page 23. The following property shows the properties of cumulative inter-block coherence constant, which are similar to the properties of cumulative MCC presented on page 24:

**Property 2.13 (Eldar's cumulative coherence properties).** For any general dictionary $\mathbf{\Phi}$ with equally-sized block structure, i.e., $d_1{=}\cdots{=}d_K{=}d$, we have:

$$M_{Inter}^{Eldar}\left(\mathbf{\Phi},1\right)=M_{Inter}^{Eldar}\left(\mathbf{\Phi}\right),$$

and

$$M_{Inter}^{Eldar}\left(\mathbf{\Phi},k\right)\leq k\,M_{Inter}^{Eldar}\left(\mathbf{\Phi}\right).$$

*Proof.* It is similar to the proof of Property 2.12 (cumulative Block-MCC$_{q,p}$ properties). □

In the following property, the relationship between our two proposed cumulative coherences, i.e., cumulative Block-MCC$_{q,p}$ (Definition 2.7) and Eldar's cumulative coherence (Definition 2.8), is investigated:

**Property 2.14 (Cumulative Block-MCC$_{q,p}$ v.s. Eldar's cumulative coherence).** For a dictionary $\mathbf{\Phi}$ with orthonormal equally-sized blocks, we have:

$$M_{2,2}\left(\mathbf{\Phi},k\right)=M_{Inter}^{Eldar}\left(\mathbf{\Phi},k\right)$$

*Proof.* From Definition 2.7 (cumulative Block-MCC$_{q,p}$), we have $\forall(q,p){\in}\mathbb{R}_{>0}^{2}$ : $M_{q,p}(\mathbf{\Phi},k){\overset{\text{def}}{=}}d_{max}^{-1}\max_{|\Lambda|=k}\max_{j\notin\Lambda}\sum_{i\in\Lambda}d_i^{-1/p}d_j^{1/q}\|\mathbf{\Phi}^{\dagger}[i]\mathbf{\Phi}[j]\|_{q\to p}$. On the other hand, for orthonormal blocks we have $\mathbf{\Phi}^{\dagger}[i]{=}\mathbf{\Phi}^{T}[i]$, and by selecting $q{=}p{=}2$ for an equally-sized block structure, i.e., $d_1{=}\cdots{=}d_K{=}d$, we get $M_{q,p}(\mathbf{\Phi},k){=}d^{-1}\max_{|\Lambda|=k}\max_{j\notin\Lambda}\sum_{i\in\Lambda}\|\mathbf{\Phi}^{T}[i]\mathbf{\Phi}[j]\|_{2\to2}{=}M_{Inter}^{Eldar}(\mathbf{\Phi},k)$. □

## 2.3 Block-sparse exact recovery condition

In this section Block-ERC based on above-mentioned different characterisations and properties of a dictionary will be studied, which can be considered as the main theoretical contribution of this thesis.

### 2.3.1 Block-ERC based on Block-Spark

To introduce the Block-ERC we need a block-wise version of the conventional uncertainty principle defined in (1.9) page 26. Therefore, generalising the ideas from [GR97]; [DS89]; [DH01]; [EB01]; [EB02]; [DE03b]; [DE03a]; [GN03b]; [GN03a]; [BDE09], we propose the following lemma which is called *block-sparse uncertainty principle based on Block-Spark*:

**Lemma 2.1 (Block-UP[4] based on Block-Spark).** For any arbitrary non-zero signal $\boldsymbol{y}$ with two distinct representations $\boldsymbol{\beta_0}$ and $\boldsymbol{\beta_1}$ in any general dictionary $\boldsymbol{\Phi}$, i.e., $\boldsymbol{y}=\boldsymbol{\Phi\beta_0}=\boldsymbol{\Phi\beta_1}$, we have:

$$\forall p \geq 0, \quad \|\boldsymbol{\beta_0}\|_{p,0} + \|\boldsymbol{\beta_1}\|_{p,0} \geq Block{-}Spark(\boldsymbol{\Phi}).$$

The proof of Lemma 2.1 is provided in Section A.6 (page 136).

As expected, for conventional element-wise sparse recovery case, i.e., $d_1{=}\cdots{=}d_K{=}1$, block-sparse uncertainty principle based on Block-Spark is equivalent to the conventional Spark-based uncertainty principle of (1.9) on page 26, i.e., $\|\boldsymbol{\beta_0}\|_0{+}\|\boldsymbol{\beta_1}\|_0{\geq}Spark(\boldsymbol{\Phi})$.

Using the proposed Lemma 2.1 (Block-UP based on Block-Spark), we introduce the following *Block-ERC based on Block-Spark*:

**Theorem 2.1 (Block-ERC based on Block-Spark).** For any general dictionary $\boldsymbol{\Phi}$, such that $\boldsymbol{y}{=}\boldsymbol{\Phi\beta_0}$, $\forall p{\geq}0$, if

$$\|\boldsymbol{\beta_0}\|_{p,0} < \frac{Block{-}Spark(\boldsymbol{\Phi})}{2},$$

then $\boldsymbol{\beta_0}$ is the unique solution to the $P_{p,0}$ optimisation problem.

*Proof.* Considering Lemma 2.1 (Block-UP based on Block-Spark), suppose that in addition to $\boldsymbol{\beta_0}$ there is another solution $\boldsymbol{\beta_1}$, that satisfies the same linear model, i.e., $\boldsymbol{y}{=}\boldsymbol{\Phi\beta_0}{=}\boldsymbol{\Phi\beta_1}$.

Since it is assumed that the number of active blocks of the candidate solution $\boldsymbol{\beta_0}$ is less than $Block{-}Spark(\boldsymbol{\Phi})/2$, from Lemma 2.1 it can be deduced that any alternative solution such as $\boldsymbol{\beta_1}$ necessarily is denser and has more than $Block{-}Spark(\boldsymbol{\Phi})/2$ active blocks. □

[4] *Block-sparse Uncertainty Principle*

As expected, for $d_1=\cdots=d_K=1$, Theorem 2.1 (Block-ERC based on Block-Spark) is equivalent to the conventional Spark-based ERC of (1.10) on page 26, i.e., $\|\boldsymbol{\beta_0}\|_0<Spark(\boldsymbol{\Phi})/2$.

As mentioned earlier, from a mathematical point of view, exploiting the block structure information of the representation leads to improved recovery conditions, i.e., conditions with higher sparsity levels, i.e., more non-zero coefficients.

For investigating this claim, consider a block-sparse representation $\boldsymbol{\beta_0}$, which satisfies the condition of the Theorem 2.1 (Block-ERC based on Block-Spark), i.e., $\|\boldsymbol{\beta_0}\|_{p,0}<Block{-}Spark(\boldsymbol{\Phi})/2$. The number of non-zero elements is at most $d_{max}$ times greater than the number of active blocks, i.e., $\forall p\geq 0$, $\|\boldsymbol{\beta_0}\|_0\leq d_{max}\,\|\boldsymbol{\beta_0}\|_{p,0}$, where, equality occurs when all of the elements in each of the blocks are non-zero in an equally-sized blocks case. Therefore, using Theorem 2.1 (Block-ERC based on Block-Spark) we have:

$$\|\boldsymbol{\beta_0}\|_0 < d_{max}\,\frac{Block{-}Spark(\boldsymbol{\Phi})}{2}. \tag{2.1}$$

Now, if we treat the block-sparse representation $\boldsymbol{\beta_0}$ as a conventional sparse representation, i.e., without exploiting its block structure, the sufficient recovery condition would be $\|\boldsymbol{\beta_0}\|_0<Spark(\boldsymbol{\Phi})/2$ as explained in (1.10) on page 26, and by comparing with (2.1), it is clear that for showing the benefit of the assumption of the block structure on the recovery condition, the relation between $d_{max}\,Block{-}Spark(\boldsymbol{\Phi})$ and $Spark(\boldsymbol{\Phi})$ should be investigated.

According to the relationship in Property 2.1 (Block-Spark v.s. Spark, page 53), i.e., $\bar{d}\,Block{-}Spark(\boldsymbol{\Phi})\geq Spark(\boldsymbol{\Phi})$, it is clear that the right-hand side of the inequality (2.1) is greater than or equal to the right-hand side of the inequality (1.10), i.e., $Spark(\boldsymbol{\Phi})/2$, because we have $d_{max}\geq\bar{d}$.

Therefore, exploiting the block structure information of the representation and using the proposed characterisation of the dictionary, named Block-Spark (Definition 2.5, page 52), improves the conventional Spark-based ERC presented in (1.10) by increasing the sparsity level, hence, weakening the corresponding conditions.

**Remark 2.8.** Although the Block-ERC based on Block-Spark in Theorem 2.1 (Block-ERC based on Block-Spark) improves the conventional condition of Gorodnitsky and Rao [GR97], Donoho et al. [DS89]; [DH01]; [DE03b]; [DE03a], Elad and Bruckstein [EB01]; [EB02], Gribonval and Nielsen [GN03b]; [GN03a], and Bruckstein et al. [BDE09], when it comes to practical computations both conditions would become unrealistic, because both the characterisation Block-Spark (and corresponding Spark) and the related optimisation problem $P_{p,0}$ (and corresponding $P_0$) are non-convex.

Therefore, the next step would focus on at least the related optimisation problems which are tractable. In the next section we propose conditions for more general optimisation problems of $P_{\boldsymbol{w};p_1,p_2}$ and $P_{p_1,p_2}$, where, $0\leq p_2\leq 1\leq p_1$.

### 2.3.2 Block-ERC based on block null space property

By refining and generalising the ideas from Donoho et al. [DH01]; [DE03b], Elad and Bruckstein [EB01]; [EB02], Gribonval and Nielsen [GN03a]; [GN03b]; [GN04]; [GN07], Feuer and Nemirovski [FN03a], Zhang [Zha05], Cohen et al. [CDD09], Stojnic et al. [SXH08]; [SPH09] and Boufounos et al. [BKR11], we propose the following theorem, called *Block Null Space Property (Block-NSP)*.

Block-NSP includes more general case of differently-sized blocks, i.e., $d_1 \neq \cdots \neq d_K$, arbitrary dictionary and integrating the block-structure information into the proposed characterisation, e.g., $Q_{\boldsymbol{w};p_1,p_2}(S_b(\boldsymbol{\beta}),\boldsymbol{\Phi})$, and problems, e.g., $P_{\boldsymbol{w};p_1,p_2}$ and $P_{p_1,p_2}$, where, $0 \leq p_2 \leq 1 \leq p_1$.

**Theorem 2.2 (Block-NSP).** Let $\boldsymbol{\Phi}$ be a general dictionary in $\boldsymbol{y}=\boldsymbol{\Phi\beta_0}$ and $S_b(\boldsymbol{\beta}) \subset \{1,\cdots,K\}$ a set of block indices. For $0 \leq p_2 \leq 1 \leq p_1$ and Block-Support $S_b(\boldsymbol{\beta_0})$, define:

$$Q_{\boldsymbol{w};p_1,p_2}\left(S_b\left(\boldsymbol{\beta}\right),\boldsymbol{\Phi}\right) \stackrel{\text{def}}{=} \max_{\boldsymbol{x}\in Ker(\boldsymbol{\Phi})\setminus\{\boldsymbol{0}\}} \frac{\sum\limits_{k\in S_b(\boldsymbol{\beta})} {w_k}^{p_2} \left| \sum\limits_{j=1}^{d_k} |x_j[k]|^{p_1} \right|^{\frac{p_2}{p_1}}}{\sum\limits_{k=1}^{K} w_k^{p_2} \left| \sum\limits_{j=1}^{d_k} |x_j[k]|^{p_1} \right|^{\frac{p_2}{p_1}}},$$

with the convention of $x^0 = \lim\limits_{i\to 0^+} x^i = \begin{cases} 1 & x\neq 0 \\ 0 & x=0 \end{cases}$ and $w_k = d_k^{-1/p_1}, \forall k$ are the elements of the weight vector $\boldsymbol{w}$.
If $Q_{\boldsymbol{w};p_1,p_2}(S_b(\boldsymbol{\beta}),\boldsymbol{\Phi}) < 1/2$, then for all $\boldsymbol{\beta_0}$ whose Block-Support lies within $S_b(\boldsymbol{\beta})$, i.e., $S_b(\boldsymbol{\beta_0}) \subset S_b(\boldsymbol{\beta})$, $\boldsymbol{\beta_0}$ is the unique solution to the $P_{\boldsymbol{w};p_1,p_2}$.

Notice that for the case of equally-sized blocks, i.e., $d_1 = \cdots = d_K = d$, if $Q_{p_1,p_2}(S_b(\boldsymbol{\beta}),\boldsymbol{\Phi}) < 1/2$ and $S_b(\boldsymbol{\beta_0}) \subset S_b(\boldsymbol{\beta})$, then $\boldsymbol{\beta_0}$ is the unique solution to the $P_{p_1,p_2}$, where, $Q_{p_1,p_2}(S_b(\boldsymbol{\beta}),\boldsymbol{\Phi})$ is the unweighted variant of $Q_{\boldsymbol{w};p_1,p_2}(S_b(\boldsymbol{\beta}),\boldsymbol{\Phi})$:

$$Q_{p_1,p_2}\left(S_b\left(\boldsymbol{\beta}\right),\boldsymbol{\Phi}\right) \stackrel{\text{def}}{=} \max_{\boldsymbol{x}\in Ker(\boldsymbol{\Phi})\setminus\{\boldsymbol{0}\}} \frac{\sum\limits_{k\in S_b(\boldsymbol{\beta})} \left| \sum\limits_{j=1}^{d} |x_j\left[k\right]|^{p_1} \right|^{\frac{p_2}{p_1}}}{\sum\limits_{k=1}^{K} \left| \sum\limits_{j=1}^{d} |x_j\left[k\right]|^{p_1} \right|^{\frac{p_2}{p_1}}}.$$

The proof of Theorem 2.2 is provided in Section A.7 (page 137).

As expected, for $d_1=\cdots=d_K$ equal to 1, $p_2$ equal to $p$, and for all $p_1\geq 1$, Block-NSP (Theorem 2.2) is equivalent to the conventional NSP (page 27).

Theorem 2.2 (Block-NSP) determines sufficient conditions on $S_b(\boldsymbol{\beta})$ by determining a 50% upper threshold on the concentration of the $\ell_{p_1}$ norm of blocks of Kernel $\boldsymbol{x}$ in Block-Support $S_b(\boldsymbol{\beta})$, such that guarantee the uniqueness of the solution to the $P_{\boldsymbol{w};p_1,p_2}$ minimisation problem.

**Corollary 2.1.** $\boldsymbol{\beta}$ is the unique solution of $P_{p,0}$ (or $P_{\boldsymbol{w};p,0}$) minimisation problem, if $\forall p\geq 1$

$$|S_b(\boldsymbol{\beta})| < \frac{Block{-}Spark(\boldsymbol{\Phi})}{2},$$

where, $|S_b(\boldsymbol{\beta})|$ is the Block-Cardinality of $\boldsymbol{\beta}$ defined in Definition 2.3 (page 52).

*Proof.* In Theorem 2.2 (Block-NSP) for a special case of $p_2=0$, $\forall p_1=p\geq 1$ we have

$$\forall S_b(\boldsymbol{\beta}), \qquad Q_{\boldsymbol{w};p,0}(S_b(\boldsymbol{\beta}),\boldsymbol{\Phi}) \leq \max_{\boldsymbol{x}\in Ker(\boldsymbol{\Phi})\setminus\{\mathbf{0}\}} \frac{|S_b(\boldsymbol{\beta})|}{\|\boldsymbol{x}\|_{\boldsymbol{w};p,0}}.$$

Maximum value of the right-hand side is obtained by minimising $\|\boldsymbol{x}\|_{\boldsymbol{w};p,0}$, which the minimum value is defined according to Definition 2.5 as Block-Spark, so

$$\forall S_b(\boldsymbol{\beta}), \qquad Q_{\boldsymbol{w};p,0}(S_b(\boldsymbol{\beta}),\boldsymbol{\Phi}) \leq \frac{|S_b(\boldsymbol{\beta})|}{Block{-}Spark(\boldsymbol{\Phi})}.$$

But from Theorem 2.2 (Block-NSP), it is known that for unique solution of the problem $P_{\boldsymbol{w};p,0}$, it is required that $Q_{\boldsymbol{w};p,0}(S_b(\boldsymbol{\beta}),\boldsymbol{\Phi})<1/2$. On the other hand, $\forall p\geq 1$, $P_{\boldsymbol{w};p,0}$ is equal to $P_{p,0}$, which proves the corollary. □

Following corollary is a double proof for Theorem 2.1 (Block-ERC based on Block-Spark, page 65) for $p\geq 1$ instead of $p\geq 0$.

**Corollary 2.2.** For any general dictionary, $\forall p\geq 1$ if

$$\|\boldsymbol{\beta_0}\|_{p,0} < \frac{Block{-}Spark(\boldsymbol{\Phi})}{2},$$

then $\boldsymbol{\beta_0}$ is the unique solution to the $P_{p,0}$ (or $P_{\boldsymbol{w};p,0}$) optimisation problem.

*Proof.* From Theorem 2.2 (Block-NSP), we have $S_b(\boldsymbol{\beta_0})\subset S_b(\boldsymbol{\beta})$, so $\|\boldsymbol{\beta_0}\|_{p,0}=|S_b(\boldsymbol{\beta_0})|\leq|S_b(\boldsymbol{\beta})|$ and using Corollary 2.1 the proof is done. □

**Remark 2.9.** Although the proposed Block-NSP (Theorem 2.2) is a generalisation of all the previous null space properties, in general the characterisation $Q_{\boldsymbol{w};p_1,p_2}(S_b(\boldsymbol{\beta}),\boldsymbol{\Phi})$ is not easy to calculate. Therefore, in the next section the tractable characterisation of coherence will be taken into consider.

### 2.3.3 Block-ERC based on block mutual coherence constant

Eldar et al. extended the basic uncertainty principle defined in (1.11) on page 28 to block-sparse domain defined in (1.19) on page 39 [EB09]; [EKB10b]; [EKB10a]. Now we generalise the results of Eldar et al., which will be also considered as a generalisation of the basic uncertainty principle [DS89]; [DH01]; [EB01]; [EB02] and name it *basic block-sparse uncertainty principle*.

**Lemma 2.2 (Basic Block-UP[5]).** Supposing $\boldsymbol{\beta_1}$ and $\boldsymbol{\beta_2}$ are two distinct block-structured representations of the non-zero signal $\boldsymbol{y}$ in two orthonormal $m\times m$ matrices $\boldsymbol{\Phi_1}$ and $\boldsymbol{\Phi_2}$ with similar block structure (same number of blocks and blocks' length vector $\boldsymbol{d}$), respectively, i.e.,:

$$\boldsymbol{y}=\sum_{k=1}^{K}\boldsymbol{\Phi_1}\left[k\right]\boldsymbol{\beta_1}\left[k\right]=\sum_{k=1}^{K}\boldsymbol{\Phi_2}\left[k\right]\boldsymbol{\beta_2}\left[k\right],$$

then $\forall(q,p)\in\mathbb{R}^2_{\geq 1}$, and $\forall r\in\mathbb{R}_{\geq 0}$:

$$\left\|\boldsymbol{\beta_1}\right\|_{r,0}+\left\|\boldsymbol{\beta_2}\right\|_{r,0}\geq\frac{2}{d_{min}^{-\frac{1}{2}}d_{max}^{\frac{3}{2}}\,\overline{M}_{q,\frac{p}{p-1}}\left(\boldsymbol{\Phi_1},\boldsymbol{\Phi_2}\right)},$$

where, $\overline{M}_{q,p/(p-1)}(\boldsymbol{\Phi_1},\boldsymbol{\Phi_2})$ is defined based on the parameter changing of $p\to p/(p-1)$ in $\overline{M}_{q,p}(\boldsymbol{\Phi_1},\boldsymbol{\Phi_2})$, which is called *basic Block-MCC*$_{q,p}$. Basic Block-MCC$_{q,p}$ is derived from the special case of Property 2.6 (Block-MCC$_{q,p}$ for intra-block orthonormality, page 58), when the characterisation is extracted from two separated matrices instead of one matrix, and is defined as follows:

$$\forall(q,p)\in\mathbb{R}^2_{>0},\qquad\overline{M}_{q,p}\left(\boldsymbol{\Phi_1},\boldsymbol{\Phi_2}\right)=\max_{k,k'}\frac{d_k^{-\frac{1}{p}}\,d_{k'}^{\frac{1}{q}}}{d_{max}}\left\|\boldsymbol{\Phi_1}^T\left[k\right]\boldsymbol{\Phi_2}\left[k'\right]\right\|_{q\to p}.$$

The proof of Lemma 2.2 is provided in Section A.8 (page 140).

From the notation point of view, notice that the basic Block-MCC$_{q,p}$ of two orthonormal matrices is equal to the Block-MCC$_{q,p}$ of the dictionary built from the concatenation of those previous orthonormal matrices, i.e., $\overline{M}_{q,p/(p-1)}(\boldsymbol{\Phi_1},\boldsymbol{\Phi_2})=M_{q,p/(p-1)}([\boldsymbol{\Phi_1},\boldsymbol{\Phi_2}])$.

As expected, for $d_1=\cdots=d_K$ equal to 1, Lemma 2.2 reduces to the basic uncertainty principle explained in (1.11), i.e., $(\|\boldsymbol{\beta_1}\|_0+\|\boldsymbol{\beta_2}\|_0)\geq 2/\overline{M}(\boldsymbol{\Phi_1},\boldsymbol{\Phi_2})$.

In comparison of different uncertainty principles, it should be taken into account that a typical uncertainty principle I is weaker than uncertainty principle II, when minimal sum of $\ell_0$ pseudo-norm or $\ell_{r,0}$ pseudo-mixed-norm of two candidate solutions $\boldsymbol{\beta_1}$ and $\boldsymbol{\beta_2}$ in I is higher than II. To see the effect of block-sparsity on the uncertainty principle, using $\|\boldsymbol{\beta_1}\|_0\leq d_{max}\|\boldsymbol{\beta_1}\|_{r,0}$ and $\|\boldsymbol{\beta_2}\|_0\leq d_{max}\|\boldsymbol{\beta_2}\|_{r,0}$, $\forall r\geq 0$, in (1.11), we get $(\|\boldsymbol{\beta_1}\|_{r,0}+\|\boldsymbol{\beta_2}\|_{r,0})\geq 2/(d_{max}\overline{M}(\boldsymbol{\Phi_1},\boldsymbol{\Phi_2}))$. Then, comparing with Lemma 2.2, we need to demonstrate the relationship between $d_{max}\overline{M}(\boldsymbol{\Phi_1},\boldsymbol{\Phi_2})$ and $d_{min}^{-1/2}d_{max}^{3/2}\overline{M}_{q,p/(p-1)}(\boldsymbol{\Phi_1},\boldsymbol{\Phi_2})$.

[5] *Block-sparse Uncertainty Principle*

**Property 2.15 (Basic Block-MCC$_{q,p}$ v.s. basic MCC).** Supposing $\mathbf{\Phi_1}$ and $\mathbf{\Phi_2}$ are two orthonormal matrices, $\forall(q,p)\in\mathbb{R}^2_{>0}$ we have:

$$0 \leq \frac{d_{min}^{-\frac{1}{2}} d_{max}^{\frac{3}{2}}\, \overline{M}_{q,\frac{p}{p-1}}(\mathbf{\Phi_1},\mathbf{\Phi_2})}{d_{max}\, \overline{M}(\mathbf{\Phi_1},\mathbf{\Phi_2})} \leq d_{min}^{-\frac{1}{2}} d_{max}^{-\frac{1}{2}} \max_{k,k'} d_k^{\frac{1}{p}-\frac{1}{2}} d_{k'}^{\frac{1}{q}+\frac{1}{2}} \max\left\{1, d_k^{\frac{1}{2}-\frac{1}{p}}\right\} \max\left\{1, d_{k'}^{\frac{1}{2}-\frac{1}{q}}\right\}.$$

The above upper-bound for different values of $q$ and $p$ are shown in table 2.7.

*Proof.* From second part of Property 2.8 (Block-MCC$_{q,p}$ bounds, page 59), which the dictionary has intra-block orthonormality, in a special case of two distinct orthonormal bases instead of one dictionary, $\forall(q,p,q',p')\in\mathbb{R}^4_{>0}$ we have:

$$0 \leq \overline{M}_{q,p}(\mathbf{\Phi_1},\mathbf{\Phi_2}) \leq \frac{\overline{M}(\mathbf{\Phi_1},\mathbf{\Phi_2})}{d_{max}} \max_{k,k'} d_k^{\frac{1}{2}-\frac{1}{p}} d_{k'}^{\frac{1}{q}+\frac{1}{2}} \max\left\{1, d_k^{\frac{1}{p}-\frac{1}{p'}}\right\} \max\left\{1, d_{k'}^{\frac{1}{q'}-\frac{1}{q}}\right\} \times \max\left\{1, d_k^{\frac{1}{p'}-\frac{1}{2}}\right\} \max\left\{1, d_{k'}^{\frac{1}{2}-\frac{1}{q'}}\right\}.$$

Then, by parameter changing of $p\to p/(p-1)$, and then selecting $q'=p'=2$, we get:

$$\begin{aligned} 0 \leq \overline{M}_{q,\frac{p}{p-1}}(\mathbf{\Phi_1},\mathbf{\Phi_2}) &\leq \frac{\overline{M}(\mathbf{\Phi_1},\mathbf{\Phi_2})}{d_{max}} \max_{k,k'} d_k^{\frac{1}{p}-\frac{1}{2}} d_{k'}^{\frac{1}{q}+\frac{1}{2}} \max\left\{1, d_k^{1-\frac{1}{p}-\frac{1}{p'}}\right\} \max\left\{1, d_{k'}^{\frac{1}{q'}-\frac{1}{q}}\right\} \times \max\left\{1, d_k^{\frac{1}{p'}-\frac{1}{2}}\right\} \max\left\{1, d_{k'}^{\frac{1}{2}-\frac{1}{q'}}\right\} \\ &= \frac{\overline{M}(\mathbf{\Phi_1},\mathbf{\Phi_2})}{d_{max}} \max_{k,k'} d_k^{\frac{1}{p}-\frac{1}{2}} d_{k'}^{\frac{1}{q}+\frac{1}{2}} \max\left\{1, d_k^{\frac{1}{2}-\frac{1}{p}}\right\} \max\left\{1, d_{k'}^{\frac{1}{2}-\frac{1}{q}}\right\}. \end{aligned}$$

Next, by multiplying by $d_{min}^{-1/2} d_{max}^{1/2} > 0$, the proof is done. □

**Remark 2.10.** As it is mentioned before, in order to compare the proposed basic block-sparse uncertainty principle in Lemma 2.2 (Basic Block-UP, page 69), i.e., $(\|\boldsymbol{\beta_1}\|_{r,0}+\|\boldsymbol{\beta_2}\|_{r,0}) \geq 2/(d_{min}^{-1/2} d_{max}^{3/2} \overline{M}_{q,p/(p-1)}(\mathbf{\Phi_1},\mathbf{\Phi_2}))$, with the conventional basic uncertainty principle explained in (1.11) on page 28, i.e., $(\|\boldsymbol{\beta_1}\|_0+\|\boldsymbol{\beta_2}\|_0) \geq 2/\overline{M}(\mathbf{\Phi_1},\mathbf{\Phi_2})$, the relationship between $d_{min}^{-1/2} d_{max}^{3/2} \overline{M}_{q,p/(p-1)}(\mathbf{\Phi_1},\mathbf{\Phi_2})$ and $d_{max}\overline{M}(\mathbf{\Phi_1},\mathbf{\Phi_2})$ should be investigated. The mentioned comparison is shown in table 2.7. Since the two terms are in the denominator of the right-hand side of the uncertainty principle, the smaller values lead to more weakened uncertainty principles, hence, improve the conditions. Therefore, the values in table 2.7, which are less than or equal to one, correspond to weakened uncertainty principles due to

| | $0<q\,\&\,p\leq 2$ | $q\,\&\,p\geq 2$ | $0<q\leq 2\,\&\,p\geq 2$ | $q\geq 2\,\&\,0<p\leq 2$ |
|---|---|---|---|---|
| $\frac{d_{min}^{-\frac{1}{2}} d_{max}^{\frac{3}{2}}\, \overline{M}_{q,\frac{p}{p-1}}(\mathbf{\Phi_1},\mathbf{\Phi_2})}{d_{max}\, \overline{M}(\mathbf{\Phi_1},\mathbf{\Phi_2})} \leq$ | $d_{min}^{-\frac{1}{2}} d_{max}^{\frac{1}{p}+\frac{1}{q}-\frac{1}{2}}$ | $d_{min}^{-\frac{1}{2}} d_{max}^{\frac{1}{2}}$ | $d_{min}^{-\frac{1}{2}} d_{max}^{\frac{1}{q}}$ | $d_{min}^{-\frac{1}{2}} d_{max}^{\frac{1}{p}}$ |

Table 2.7: Upper-bound of $d_{min}^{-1/2} d_{max}^{3/2} \overline{M}_{q,p/(p-1)}(\mathbf{\Phi_1},\mathbf{\Phi_2}) / d_{max}\overline{M}(\mathbf{\Phi_1},\mathbf{\Phi_2})$ for different values of $q$ and $p$.

using block-sparsity. From table 2.7, it can be seen that for $\forall(q,p)\in\mathbb{R}^2_{\geq 2}$ and $d_{min}{=}d_{max}$, the proposed basic block-sparse uncertainty principle in Lemma 2.2 (Basic Block-UP) improves the conventional basic uncertainty principle explained in (1.11).

To compare the proposed basic block-sparse uncertainty principle in Lemma 2.2 (Basic Block-UP, page 69, with $d_{min}{=}d_{max}{=}d$) with the block-sparse uncertainty principle proposed by Eldar et al. in (1.19) on page 39, i.e., $\|\boldsymbol{\beta_1}\|_{2,0}{+}\|\boldsymbol{\beta_2}\|_{2,0}{\geq}2/(d\,\overline{M}^{Eldar}_{Inter}(\boldsymbol{\Phi_1},\boldsymbol{\Phi_2}))$, where $\overline{M}^{Eldar}_{Inter}(\boldsymbol{\Phi_1},\boldsymbol{\Phi_2}){=}\max_{k,k'}\|\boldsymbol{\Phi}_1^T[k]\boldsymbol{\Phi_2}[k']\|_{2\to 2}/d$, we need to investigate the relationship between $\overline{M}_{q,p/(p-1)}(\boldsymbol{\Phi_1},\boldsymbol{\Phi_2})$ and $\overline{M}^{Eldar}_{Inter}(\boldsymbol{\Phi_1},\boldsymbol{\Phi_2})$.

**Property 2.16 (Basic Block-MCC$_{q,p}$ v.s. basic block-coherence of Eldar et al.).** Supposing $\boldsymbol{\Phi_1}$ and $\boldsymbol{\Phi_2}$ are two orthonormal matrices with equally-sized blocks of length $d$, $\forall(q,p)\in\mathbb{R}^2_{>0}$ we have:

$$\frac{\overline{M}_{q,\frac{p}{p-1}}(\boldsymbol{\Phi_1},\boldsymbol{\Phi_2})}{\overline{M}^{Eldar}_{Inter}(\boldsymbol{\Phi_1},\boldsymbol{\Phi_2})}\geq d^{\frac{1}{q}+\frac{1}{p}-1}\min\left\{1,d^{\frac{1}{2}-\frac{1}{p}}\right\}\min\left\{1,d^{\frac{1}{2}-\frac{1}{q}}\right\},$$
$$\frac{\overline{M}_{q,\frac{p}{p-1}}(\boldsymbol{\Phi_1},\boldsymbol{\Phi_2})}{\overline{M}^{Eldar}_{Inter}(\boldsymbol{\Phi_1},\boldsymbol{\Phi_2})}\leq d^{\frac{1}{q}+\frac{1}{p}-1}\max\left\{1,d^{\frac{1}{2}-\frac{1}{p}}\right\}\max\left\{1,d^{\frac{1}{2}-\frac{1}{q}}\right\}.$$

Then, for $q{=}p{=}2$ the lower- and upper-bound are both equal to one, then $\overline{M}_{2,2}(\boldsymbol{\Phi_1},\boldsymbol{\Phi_2}){=}\overline{M}^{Eldar}_{Inter}(\boldsymbol{\Phi_1},\boldsymbol{\Phi_2})$. The above bounds for different values of $q$ and $p$ are shown in table 2.8.

*Proof.* It follows from Property 2.11 (Block-MCC$_{q,p}$ v.s. block-coherence of Eldar et al., page 62) in a special case of two orthonormal matrices $\forall(q,p)\in\mathbb{R}^2_{>0}$:

$$\overline{M}_{q,p}(\boldsymbol{\Phi_1},\boldsymbol{\Phi_2})\geq d^{\frac{1}{q}-\frac{1}{p}}\min\left\{1,d^{\frac{1}{p}-\frac{1}{2}}\right\}\min\left\{1,d^{\frac{1}{2}-\frac{1}{q}}\right\}\overline{M}^{Eldar}_{Inter}(\boldsymbol{\Phi_1},\boldsymbol{\Phi_2}),$$
$$\overline{M}_{q,p}(\boldsymbol{\Phi_1},\boldsymbol{\Phi_2})\leq d^{\frac{1}{q}-\frac{1}{p}}\max\left\{1,d^{\frac{1}{p}-\frac{1}{2}}\right\}\max\left\{1,d^{\frac{1}{2}-\frac{1}{q}}\right\}\overline{M}^{Eldar}_{Inter}(\boldsymbol{\Phi_1},\boldsymbol{\Phi_2}).$$

Then, by parameter changing of $p{\to}p/(p{-}1)$, the proof is done. □

**Remark 2.11.** To compare the proposed basic block-sparse uncertainty principle in Lemma 2.2 (Basic Block-UP, page 69, with $d_{min}{=}d_{max}{=}d$), i.e., $(\|\boldsymbol{\beta_1}\|_{r,0}{+}\|\boldsymbol{\beta_2}\|_{r,0}){\geq}2/(d\overline{M}_{q,p/(p-1)}(\boldsymbol{\Phi_1},\boldsymbol{\Phi_2}))$, with the block-sparse uncertainty principle proposed by Eldar et al. in (1.19), i.e., $\|\boldsymbol{\beta_1}\|_{2,0}{+}\|\boldsymbol{\beta_2}\|_{2,0}{\geq}2/(d\,\overline{M}^{Eldar}_{Inter}(\boldsymbol{\Phi_1},\boldsymbol{\Phi_2}))$, the

| | $0<q\,\&\,p\leq 2$ | $q\,\&\,p\geq 2$ | $0<q\leq 2\,\&\,p\geq 2$ | $q\geq 2\,\&\,0<p\leq 2$ |
|---|---|---|---|---|
| $\frac{\overline{M}_{q,\frac{p}{p-1}}(\boldsymbol{\Phi_1},\boldsymbol{\Phi_2})}{\overline{M}^{Eldar}_{Inter}(\boldsymbol{\Phi_1},\boldsymbol{\Phi_2})}\geq$ | 1 | $d^{\frac{1}{q}+\frac{1}{p}-1}$ | $d^{\frac{1}{p}-\frac{1}{2}}$ | $d^{\frac{1}{q}-\frac{1}{2}}$ |
| $\frac{\overline{M}_{q,\frac{p}{p-1}}(\boldsymbol{\Phi_1},\boldsymbol{\Phi_2})}{\overline{M}^{Eldar}_{Inter}(\boldsymbol{\Phi_1},\boldsymbol{\Phi_2})}\leq$ | $d^{\frac{1}{q}+\frac{1}{p}-1}$ | 1 | $d^{\frac{1}{q}-\frac{1}{2}}$ | $d^{\frac{1}{p}-\frac{1}{2}}$ |

Table 2.8: Bounds of $\overline{M}_{q,p/(p-1)}(\boldsymbol{\Phi_1},\boldsymbol{\Phi_2})/\overline{M}^{Eldar}_{Inter}(\boldsymbol{\Phi_1},\boldsymbol{\Phi_2})$ for different values of $q$ and $p$.

relationship between $\overline{M}_{q,p/(p-1)}(\boldsymbol{\Phi_1},\boldsymbol{\Phi_2})$ and $\overline{M}^{Eldar}_{Inter}(\boldsymbol{\Phi_1},\boldsymbol{\Phi_2})$ should be investigated. With the same reasoning as the one used in Remark 2.10 (page 70), the upper-bound values in table 2.8, which are less than or equal to one, correspond to weakened uncertainty principles. From table 2.8, it can be seen that for $\forall(q,p)\in\mathbb{R}^2_{\geq 2}$, the proposed basic block-sparse uncertainty principle in Lemma 2.2 improves the basic block-sparse uncertainty principle proposed by Eldar et al. in (1.19) on page 39.

As mentioned before, we had lower-bound for the basic MCC, i.e., $\overline{M}(\boldsymbol{\Phi_1},\boldsymbol{\Phi_2})\geq 1/\sqrt{m}$ (page 22), and for basic block-coherence of Eldar et al., i.e., $\overline{M}^{Eldar}_{Inter}(\boldsymbol{\Phi_1},\boldsymbol{\Phi_2})\geq 1/\sqrt{d\,m}$ (page 40). Next, we show the corresponding lower-bound for the proposed basic Block-MCC$_{q,p}$ in a special case of $q{=}p{=}2$, i.e., $\overline{M}_{2,2}(\boldsymbol{\Phi_1},\boldsymbol{\Phi_2})$.

**Property 2.17 (Basic Block-MCC$_{2,2}$ lower-bound).** Supposing $\boldsymbol{\Phi_1}$ and $\boldsymbol{\Phi_2}$ are two $m{\times}m$ orthonormal matrices, we have:

$$\overline{M}_{2,2}\left(\boldsymbol{\Phi_1},\boldsymbol{\Phi_2}\right)\geq\frac{d_{min}\,d_{max}^{-\frac{3}{2}}}{\sqrt{m}}. \tag{2.2}$$

The proof of Property 2.17 is provided in Section A.9 (page 144).

**Remark 2.12 (Block-incoherency).** It can be seen that for special settings of $d_1{=}\cdots{=}d_K{=}1$ and equally-sized blocks, i.e., $d_1{=}\cdots{=}d_K{=}d$, Property 2.17 is equivalent to the conventional bound, i.e., $\overline{M}(\boldsymbol{\Phi_1},\boldsymbol{\Phi_2})\geq 1/\sqrt{m}$, and the bound of Eldar et al., i.e., $\overline{M}^{Eldar}_{Inter}(\boldsymbol{\Phi_1},\boldsymbol{\Phi_2})\geq 1/\sqrt{d\,m}$, respectively. Generally, for settings of $d_{min}\,d_{max}^{-3/2}{\leq}1$, the lower-bound in Property 2.17 is less than or equal to the lower-bound in conventional case, which means that in the proposed block-structured scenario, dictionaries can be more *block-incoherent* compared to the conventional case. Since there is a direct relationship between the sparsity level of recovery condition and block-incoherency, hence improved recovery conditions are obtained in block-structured scenario. In addition, notice that in the proof of all above basic two orthonormal bases we used the relationships of a dictionary with intra-block orthonormality, because for an orthonormal base $\boldsymbol{\Phi}{\in}\mathbb{R}^{m\times m}$, with $\boldsymbol{\Phi}^T\boldsymbol{\Phi}{=}\boldsymbol{I}_m$, we have $\boldsymbol{\Phi}^T[k]\boldsymbol{\Phi}[k]{=}\boldsymbol{I}_{d_k}$.

Returning back from basic two orthonormal bases $\boldsymbol{\Phi_1}$ and $\boldsymbol{\Phi_2}$ to the dictionaries $\boldsymbol{\Phi}$, the corresponding relations based on Block-MCC$_{q,p}$ (Definition 2.6, page 54) should be established. Thanks to Block-MCC$_{q,p}$, we can overcome the issue of intractability of the Block-Spark (Definition 2.5, page 52). Now, we introduce the following *Block-ERC based on Block-MCC$_{q,p}$*:

**Theorem 2.3 (Block-ERC based on Block-MCC$_{q,p}$).** For any dictionary $\boldsymbol{\Phi}$ with full column rank blocks, let $\boldsymbol{y}=\boldsymbol{\Phi\beta_0}$, then supposing Block-MCC$_{q,p}$ $M_{q,p}(\boldsymbol{\Phi})$, Block-Support relation $S_b(\boldsymbol{\beta_0})\subset S_b(\boldsymbol{\beta})$, $\forall(q,p)\in\mathbb{R}^2_{\geq 1}$, and $\forall r\in\mathbb{R}_{\geq 0}$, if

$$\|\boldsymbol{\beta_0}\|_{r,0} < \frac{1+(d_{max}M_{q,p}(\boldsymbol{\Phi}))^{-1}\min\limits_{k}\min\left\{1,d_k^{\frac{1}{q}-\frac{1}{p}}\right\}}{2},$$

then $\boldsymbol{\beta_0}$ is the unique solution to the $P_{\boldsymbol{w};q,1}$ problem.

The proof of Theorem 2.3 is provided in Section A.10 (page 145).

For $d_1=\cdots=d_K=1$, Theorem 2.3 converges to its conventional element-wise counterpart explained in (1.13) on page 29, i.e., $\|\boldsymbol{\beta_0}\|_0<(1+M^{-1}(\boldsymbol{\Phi}))/2$.

As mentioned before (page 38), the right-hand side of the equation in Theorem 2.3 is called Block-SL and is represented as $Block\text{-}SL(\boldsymbol{\Phi})$. In order to make a comparison of Block-SL of Theorem 2.3, for different basic tractable $(q,p)$ pairs of table 2.1 (page 54), let us represent it as $Block\text{-}SL_{q,p}(\boldsymbol{\Phi})$ and call it $(q,p)$-Block-Sparsity Level (Block-SL$_{q,p}$). In the following property we investigate the possible relationship between different Block-SL$_{q,p}$:

**Property 2.18 (Block-SL$_{q,p}$ inequalities).** The different Block-SL$_{q,p}$ calculated for basic tractable $\ell_{q\to p}$ operator-norms of table 2.1 (page 54) have the following relationships:

$$\begin{aligned}
&Block\text{-}SL_{1,\infty}(\boldsymbol{\Phi}) \leq Block\text{-}SL_{1,2}(\boldsymbol{\Phi}) \leq Block\text{-}SL_{1,1}(\boldsymbol{\Phi}),\\
&Block\text{-}SL_{1,\infty}(\boldsymbol{\Phi}) \leq Block\text{-}SL_{1,2}(\boldsymbol{\Phi}) \leq Block\text{-}SL_{2,2}(\boldsymbol{\Phi}),\\
&Block\text{-}SL_{1,\infty}(\boldsymbol{\Phi}) \leq Block\text{-}SL_{2,\infty}(\boldsymbol{\Phi}) \leq Block\text{-}SL_{2,2}(\boldsymbol{\Phi}),\\
&Block\text{-}SL_{1,\infty}(\boldsymbol{\Phi}) \leq Block\text{-}SL_{2,\infty}(\boldsymbol{\Phi}) \leq Block\text{-}SL_{\infty,\infty}(\boldsymbol{\Phi}),
\end{aligned}$$

which is shown in figure 2.5(a), while their general relationship for equally-sized case, i.e., $d_1=\cdots=d_K=d$, is represented schematically in figure 2.5(b).

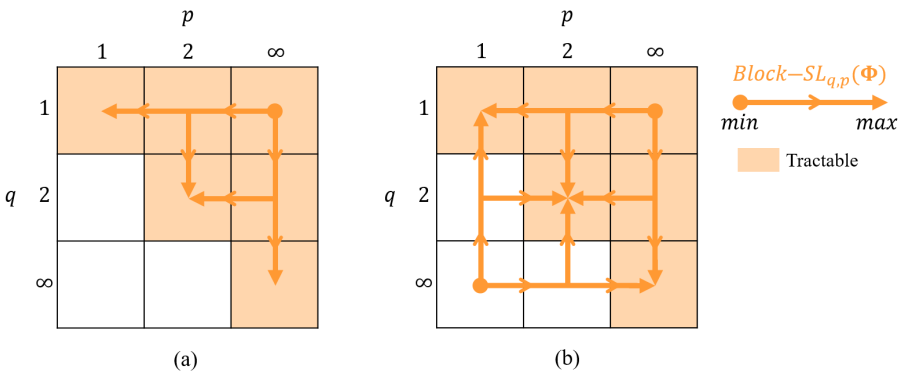



Figure 2.5: Block-SL$_{q,p}$ inequalities for (a) differently-sized, and (b) equally-sized block structure, for common $\ell_{q\to p}$ operator-norms according to table 2.1 (page 54).

The proof of Property 2.18 is provided in Section A.11 (page 148).

Now we need to prove the claim of improved recovery conditions in the block-wise domain through comparing sparsity level with block-sparsity level. To this end, supposing that $\boldsymbol{\beta_0}$ has a block-sparse structure and satisfies the condition of Theorem 2.3 (Block-ERC based on Block-MCC$_{q,p}$), and considering the fact that $\forall r \geq 0$, $\|\boldsymbol{\beta_0}\|_0 \leq d_{max}\, \|\boldsymbol{\beta_0}\|_{r,0}$, we have:

$$\forall (q,p) \in \mathbb{R}^2_{\geq 1}, \qquad \|\boldsymbol{\beta_0}\|_0 < \frac{d_{max} + M^{-1}_{q,p}\left(\boldsymbol{\Phi}\right) \min\limits_k \min \left\{1, d_k^{\frac{1}{q}-\frac{1}{p}}\right\}}{2}.$$

Comparing with $\|\boldsymbol{\beta_0}\|_0 < (1+M^{-1}(\boldsymbol{\Phi}))/2$ in (1.13) on page 29, we see that the relationship between $d_{max}+M^{-1}_{q,p}(\boldsymbol{\Phi}) \min_k \min\{1, d_k^{1/q-1/p}\}$ and $1+M^{-1}(\boldsymbol{\Phi})$ should be investigated. We show this relationship in the following property, which the proof is provided in Section A.12 (page 149):

**Property 2.19 (SL[6] v.s. Block-SL$_{q,p}$).** For a dictionary $\boldsymbol{\Phi}$ with intra-block orthonormality, and Block-MCC$_{q,p}$ $M_{q,p}(\boldsymbol{\Phi})$, $\forall (q,p,q',p') \in \mathbb{R}^4_{>0}$ if the MCC $M(\boldsymbol{\Phi})$ is small enough, i.e.,

$$M\left(\boldsymbol{\Phi}\right) \leq \frac{1 - \dfrac{d_{max} \min\limits_k \min \left\{1, d_k^{\frac{1}{q}-\frac{1}{p}}\right\}}{\max\limits_{k,k' \neq k} d_k^{\frac{1}{2}-\frac{1}{p}}\, d_{k'}^{\frac{1}{q}+\frac{1}{2}} \max \left\{1, d_k^{\frac{1}{p}-\frac{1}{p'}}\right\} \max \left\{1, d_{k'}^{\frac{1}{q'}-\frac{1}{q}}\right\} \max \left\{1, d_k^{\frac{1}{p'}-\frac{1}{2}}\right\} \max \left\{1, d_{k'}^{\frac{1}{2}-\frac{1}{q'}}\right\}}}{d_{max} - 1},$$

then the proposed sparsity level in Theorem 2.3 (block-ERC based on Block-MCC$_{q,p}$, page 73) is higher than the conventional sparsity level, i.e.,

$$d_{max} + M^{-1}_{q,p}\left(\boldsymbol{\Phi}\right) \min_k \min \left\{1, d_k^{\frac{1}{q}-\frac{1}{p}}\right\} \geq 1 + M^{-1}\left(\boldsymbol{\Phi}\right).$$

**Remark 2.13.** Therefore, exploiting the block structure information of the representation and using the proposed characterisation of the dictionary, named Block-MCC$_{q,p}$ (Definition 2.6, page 54), improves the conventional MCC-based ERC of Donoho et al. [DH01]; [DE03b]; [DET06], Gribonval and Nielsen [GN03a]; [GN07], Tropp [Tro04a], and Bruckstein et al. [BDE09] by increasing the sparsity level, hence weakening the corresponding conditions. The mentioned improvement is under condition on conventional MCC, which ensures that the dictionary is sufficiently incoherent, i.e., MCC is small enough.

In order to have a clear sense of the upper-bound on the conventional MCC in Property 2.19 (SL v.s. Block-SL$_{q,p}$), this value is calculated in table 2.9 for the basic tractable $(q,p)$ pairs of table 2.1 (page 54). Although, for the MCC values less than the upper-bounds in table 2.9, the supremacy of the proposed sparsity level is ensured, but these bounds are very pessimistic and in practice for even higher values of MCC the supremacy of the proposed sparsity level can be observed.

[6] Sparsity Level

| $(q=q', p=p')$ | $(1,1)$ | $(1,2)$ | $(1,\infty)$ | $(2,2)$ | $(2,\infty)$ | $(\infty,\infty)$ |
|---|---|---|---|---|---|---|
| $M(\boldsymbol{\Phi}) \leq$ | $\frac{1}{d_{max}^{\frac{1}{2}}\left(d_{max}^{\frac{1}{2}}+1\right)}$ | $\frac{1}{d_{max}^{\frac{1}{2}}\left(d_{max}^{\frac{1}{2}}+1\right)}$ | $\frac{1}{d_{max}}$ | $0$ | $\frac{1}{d_{max}^{\frac{1}{2}}\left(d_{max}^{\frac{1}{2}}+1\right)}$ | $\frac{1}{d_{max}^{\frac{1}{2}}\left(d_{max}^{\frac{1}{2}}+1\right)}$ |

Table 2.9: Upper-bound of MCC ensuring the supremacy of the proposed sparsity level, for different basic values of $(q,p)$ pairs and for a dictionary with intra-block orthonormality.

**Remark 2.14.** Table 2.9 demonstrates that our proposed sparsity level for the $\ell_{2\to 2}$ operator-norm, which is equal to the conventional $\ell_2$ matrix norm, requires the most strict upper-bound on $M(\boldsymbol{\Phi})$ in order to be higher than the conventional sparsity level, while the dictionaries with characterisations that use operator-norms other than $\ell_{2\to 2}$ can be less incoherent (more coherent, i.e., higher $M(\boldsymbol{\Phi})$ is allowed), and especially for $\ell_{1\to\infty}$ operator-norm the most relaxed upper-bound on $M(\boldsymbol{\Phi})$ is achieved. This is one of the advantages of utilising operator-norms, which enable us to have more relaxed conditions by exploiting norms other than the conventional $\ell_2$.

Now Block-ERC of Eldar et al., i.e., $\|\boldsymbol{\beta_0}\|_{2,0}<(1+(dM_{Inter}^{Eldar}(\boldsymbol{\Phi}))^{-1}(1-(d-1)M_{Intra}^{Eldar}(\boldsymbol{\Phi})))/2$, explained in (1.20) on page 40, can be compared to our Block-ERC proposed in Theorem 2.3 (Block-ERC based on Block-MCC$_{q,p}$, page 73). Supposing that all blocks sharing the same block length $d$, it is clear that the relationship between $(M_{Inter}^{Eldar}(\boldsymbol{\Phi}))^{-1}(1-(d-1)M_{Intra}^{Eldar}(\boldsymbol{\Phi}))$ and $M_{q,p}^{-1}(\boldsymbol{\Phi})\min\{1, d^{1/q-1/p}\}$ should be investigated.

**Lemma 2.3 (Eldar et al.'s v.s. proposed Block-SL$_{q,p}$).** For a dictionary with intra-block orthonormality, Block-ERC proposed in Theorem 2.3 (Block-ERC based on Block-MCC$_{q,p}$, page 73) for $q{=}p{=}2$ as the best case is equal to Block-ERC of Eldar et al. (equation (1.20), page 40).

The proof of Lemma 2.3 is provided in Section A.13 (page 151).

**Remark 2.15.** It is worth mentioning that in dictionaries with intra-block orthonormality, by definition of intra-block coherence of Eldar et al., $M_{Intra}^{Eldar}(\boldsymbol{\Phi})$ is equal to zero in Block-ERC of Eldar et al. (equation (1.20), page 40), i.e., $\|\boldsymbol{\beta_0}\|_{2,0}<(1+(dM_{Inter}^{Eldar}(\boldsymbol{\Phi}))^{-1}(1-(d-1)M_{Intra}^{Eldar}(\boldsymbol{\Phi})))/2=(1+(dM_{Inter}^{Eldar}(\boldsymbol{\Phi}))^{-1})/2$. On the other hand, based on lemma 2.3, the proposed Block-ERC for $q{=}p{=}2$ is equivalent to the condition of Eldar and her co-workers. Therefore, our proposed Block-ERC in Theorem 2.3 (page 73) in the following special setting, reduces to Block-ERC of Eldar et al.:

- $q{=}p{=}2$,
- equally-sized blocks, i.e., $d_1{=}\cdots{=}d_K{=}d$, and
- intra-block orthonormality of dictionary, i.e., for $1{\leq}k{\leq}K$, $\boldsymbol{\Phi}^T[k]\boldsymbol{\Phi}[k]{=}\boldsymbol{I}_d$.

In other words, in the mentioned special setting, *theoretically and independent of the recovery algorithm, the same Block-ERC can be achieved.*

### 2.3.4 Block-ERC based on cumulative coherence constant

All the coherence characterisations such as conventional MCC (page 22) [DH01], block-coherence of Eldar et al. (page 36) [EKB10a], mutual subspace coherence of Ganesh et al. (page 37) [GZM09], or our Block-MCC$_{q,p}$ (page 54) regardless of their definition in the block-wise or element-wise domain represent only the most extreme correlation between the atoms, blocks or subspaces of the dictionary and do not offer a comprehensive description of the dictionary.

Based on the proposed Eldar's cumulative coherence (Definition 2.8, page 63) for equally-sized blocks of length $d$, we introduce the following condition named *Block-ERC based on cumulative inter-block coherence constant*:

**Theorem 2.4 (Block-ERC based on Eldar's cumulative coherence).** For any general dictionary $\boldsymbol{\Phi}$ with equally-sized blocks of length $d$, and for $(d-1)M_{Intra}^{Eldar}(\boldsymbol{\Phi})+d\,M_{Inter}^{Eldar}(\boldsymbol{\Phi},k-1)<1$, if

$$M_{Inter}^{Eldar}(\boldsymbol{\Phi},k)+M_{Inter}^{Eldar}(\boldsymbol{\Phi},k-1)<\frac{1-(d-1)\,M_{Intra}^{Eldar}(\boldsymbol{\Phi})}{d},$$

where, $M_{Inter}^{Eldar}(\boldsymbol{\Phi},k)\overset{\text{def}}{=}d^{-1}\max_{|\Lambda|=k}\max_{j\notin\Lambda}\sum_{i\in\Lambda}\|\boldsymbol{\Phi}^T[i]\boldsymbol{\Phi}[j]\|_{2\to2}$ (Eldar's cumulative coherence, Definition 2.8, page 63), and $M_{Intra}^{Eldar}(\boldsymbol{\Phi})\overset{\text{def}}{=}\max_{\substack{i,j\neq i\\ k}}|\boldsymbol{\varphi}_i^T[k]\boldsymbol{\varphi}_j[k]|$ (page 36), then block $k$-sparse representation vector $\boldsymbol{\beta_0}$ can be recovered correctly from block orthogonal matching pursuit and $\ell_2/\ell_1$-optimisation program algorithms.

The proof of Theorem 2.4 is provided in Section A.14 (page 154).

For $d$ equal to 1, the Block-ERC based on cumulative inter-block coherence constant which is proposed in Theorem 2.4 is equivalent to its element-wise counterpart defined in (1.14) on page 31, i.e., the ERC based on cumulative MCC, i.e., $M(\boldsymbol{\Phi},k)+M(\boldsymbol{\Phi},k-1)<1$. The last proposed Block-ERC is the only algorithm-dependent condition, which the algorithms are explained in [EB09].

## 2.4 Numerical experiments

In this section, the block-sparsity levels resulted from the proposed Block-ERC based on Block-MCC$_{q,p}$ (Theorem 2.3, page 73), i.e., $Block{-}SL_{q,p}(\boldsymbol{\Phi}){=}(1{+}(d_{max}M_{q,p}(\boldsymbol{\Phi}))^{-1}\min_k \min\{1, d_k^{1/q-1/p}\})/2$, are going to be calculated. Block-MCC$_{q,p}$ is selected, because it is computationally tractable. In all of the numerical simulations performed below, we have:

1. Dictionaries are generated from independent and identically distributed (i.i.d.) random variables, with 40 rows and 400 columns.
2. All columns of the dictionaries are normalized to have unit $\ell_2$ norm.
3. The results are shown in terms of average and standard deviation over 100 repetitions of random dictionary generation.
4. All blocks share the same length $d$, i.e., $d_1{=}\cdots{=}d_K{=}d$.
5. Only the basic tractable operator-norms of table 2.1 (page 54) are considered for the calculation of the Block-MCC$_{q,p}$.

In Definition 2.6 (page 54), we proposed the characterisation Block-MCC$_{q,p}$ in a general format, which has three main free parameters of $d_k$, $\boldsymbol{\Phi}$, and $\ell_{q\to p}$ operator-norm, i.e., $M_{q,p}(\boldsymbol{\Phi}){=}\max_{k,k'\neq k}(d_k^{-1/p}d_{k'}^{1/q})/d_{max}\|\boldsymbol{\Phi}^\dagger[k]\boldsymbol{\Phi}[k']\|_{q\to p}$. Therefore, following parameters can be considered in the computation of sparsity level:

- Block length $d_k$: Block-MCC$_{q,p}$ is a function of the length of each block, i.e. $d_k$, $\forall k$.
- Type of dictionary $\boldsymbol{\Phi}$: Although Block-MCC$_{q,p}$ is defined for a dictionary in general, but in Property 2.6 (Block-MCC$_{q,p}$ for intra-block orthonormality, page 58), we investigated a special case, where, there is intra-block orthonormality, i.e., $\boldsymbol{\Phi}^T[k]\boldsymbol{\Phi}[k]{=}\boldsymbol{I}_d$, $\forall k$.
- $\ell_{q\to p}$ operator-norm in Block-MCC$_{q,p}$: Based on the table 2.1 (page 54), only six tractable operator-norms of $\ell_{1\to 1}$, $\ell_{1\to 2}$, $\ell_{1\to\infty}$, $\ell_{2\to 2}$, $\ell_{2\to\infty}$, and $\ell_{\infty\to\infty}$ are considered in the numerical experiments.

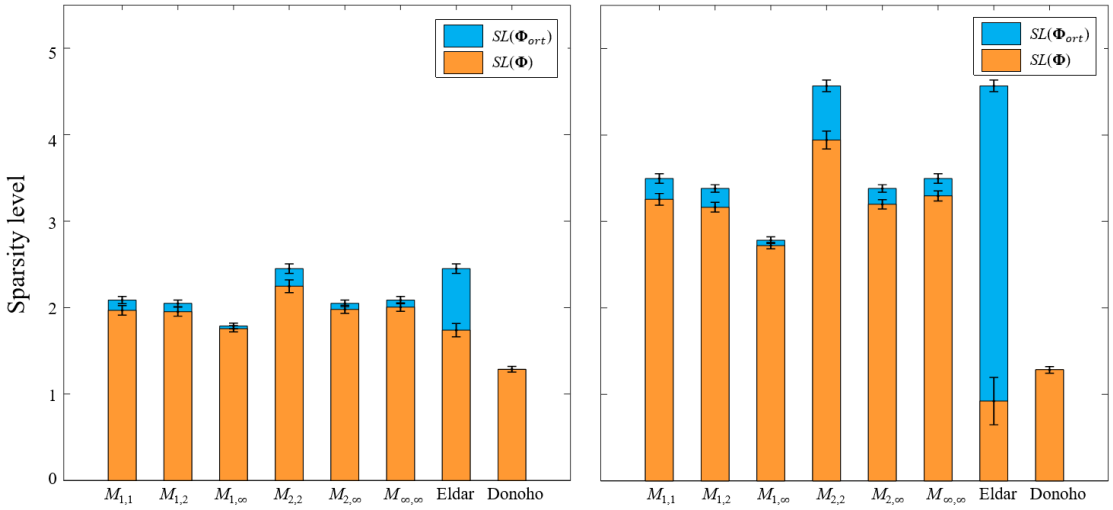
In order to investigate the behaviour of sparsity level in Block-ERC based on Block-MCC$_{q,p}$ as a function of the three above-mentioned parameters, in each of the following three experiments, one parameter out of three is kept fixed. Finally, for comparison with the existing results, the sparsity level in the sparsity domain introduced by Donoho et al. (equation (1.13), page 29), i.e., $SL^{Donoho}{=}(1{+}M^{-1}(\boldsymbol{\Phi}))/2$, and in the block-sparsity domain introduced by Eldar et al. (equation (1.20), page 40), i.e., $SL^{Eldar}{=}(d{+}(M_{Inter}^{Eldar}(\boldsymbol{\Phi}))^{-1}(1{-}(d{-}1)M_{Intra}^{Eldar}(\boldsymbol{\Phi}))/2$, are provided. Notice, in order to compare to the conventional sparsity level of Donoho et al., the block-sparsity level resulted in the Block-ERC, whether ours or the results of Eldar et al., are multiplied by $d$ to be transformed to the sparsity level, because the dictionary consists of equally-sized blocks. At last, the higher the sparsity level, the more improved the recovery condition.

### 2.4.1 Effect of dictionary and operator-norm type

In this experiment, assume that the block length $d$ is fixed. The goal is to investigate the sparsity level for two types of random dictionaries without and with intra-block orthonormality, and for the six pairs of $q$ and $p$ with tractable operator-norms.

Finally, the sparsity levels for two types of random dictionaries each with six different pairs of $q$ and $p$ are compared to the conventional sparsity levels of Eldar et al. and Donoho and his co-workers. The result of the experiment is shown in figure 2.6 for (a) $d$=2 and (b) $d$=4. The first six bars in figure 2.6(a) and (b) correspond to the proposed sparsity levels for random dictionaries without ($SL(\mathbf{\Phi})$, orange colour) and with ($SL(\mathbf{\Phi}_{ort})$, blue colour) intra-block orthonormality. Sparsity levels of Donoho et al. for the mentioned two types of dictionary are equal.

As it can be seen in figure 2.6(a) and (b), all the proposed sparsity levels based on Block-MCC$_{q,p}$ for random dictionaries $\mathbf{\Phi}$ without intra-block orthonormality are higher than the conventional sparsity levels introduced by Eldar et al. and Donoho et al. (orange colour), whereas for random dictionaries $\mathbf{\Phi}_{ort}$ with intra-block orthonormality (blue colour), the highest value corresponds to Block-MCC$_{2,2}$, which is equal to the sparsity level of Eldar et al., as proved theoretically in Lemma 2.3 (page 75), and explained in Remark 2.15. At last, all the proposed sparsity levels for two types of dictionaries are higher than the conventional sparsity level of Donoho et al. as proved theoretically in Property 2.19 (page 74), and explained in Remark 2.13.




(a) $d = 2$ (b) $d = 4$

Figure 2.6: Sparsity levels for six tractable Block-MCC$_{q,p}$, for dictionaries without ($SL(\mathbf{\Phi})$) and with ($SL(\mathbf{\Phi}_{ort})$) intra-block orthonormality compared to two conventional sparsity levels for (a) $d$=2, and (b) $d$=4.

### 2.4.2 Effect of block length and operator-norm type

In this experiment, assume that the dictionary is in general with full column rank blocks. The goal is to investigate the sparsity levels for the six pairs of $q$ and $p$ with tractable operator-norms and for different dictionaries with equally-sized blocks $d$=$\{1, 2, 4, 5, 8, 10\}$. Finally, the sparsity levels for six Block-MCC$_{q,p}$ computed for different values of $d$, are compared to the conventional sparsity levels of Donoho et al. and Eldar and her co-workers. As it can be seen in figure 2.7, for a general dictionary, for $d$>1 all the proposed sparsity levels based on Block-MCC$_{q,p}$ are higher than the conventional sparsity levels introduced by Donoho et al. and Eldar and her co-workers. By increasing the block length $d$, the difference between the proposed and conventional sparsity levels becomes more pronounced.

Assuming that the sparsity level corresponding to Block-MCC$_{q,p}$ is represented by $SL_{q,p}(\mathbf{\Phi})$, the following SL$_{q,p}$ inequalities was obtained from simulation experiments for all values of $d$: $SL_{1,\infty}(\mathbf{\Phi})$<$SL_{1,2}(\mathbf{\Phi})$<$SL_{2,\infty}(\mathbf{\Phi})$<$SL_{1,1}(\mathbf{\Phi})$<$SL_{\infty,\infty}(\mathbf{\Phi})$<$SL_{2,2}(\mathbf{\Phi})$, whereas $SL_{1,2}(\mathbf{\Phi})$, $SL_{2,\infty}(\mathbf{\Phi})$, $SL_{1,1}(\mathbf{\Phi})$, and $SL_{\infty,\infty}(\mathbf{\Phi})$ are closer to each other compared to the two other sparsity levels. Part of the sparsity level inequalities obtained from simulation experiments is proved theoretically in Property 2.18 (Block-SL$_{q,p}$ inequalities, page 73).

For a general dictionary with full column rank blocks, sparsity level of Eldar et al. is not computable for all values of $d$, because from Block-ERC of Eldar et al., i.e., $\|\boldsymbol{\beta_0}\|_{2,0}$<$(1+(dM_{Inter}^{Eldar}(\mathbf{\Phi}))^{-1}(1-(d-1)M_{Intra}^{Eldar}(\mathbf{\Phi})))/2$, (equation (1.20), page 40), the inequality $1-(d-1)M_{Intra}^{Eldar}(\mathbf{\Phi})+d\,M_{Inter}^{Eldar}(\mathbf{\Phi})$>0 should hold true in order to have a positive sparsity levels. At last, for $d$=1, the SL in block domain are equal to the scalar domain.

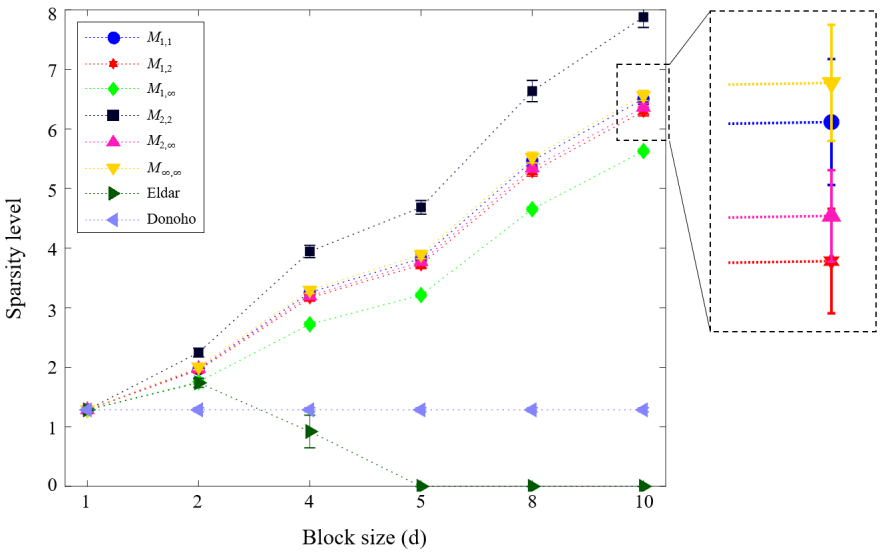



Figure 2.7: Sparsity levels for different values of block length $d$ using six Block-MCC$_{q,p}$, compared to two conventional sparsity levels proposed by Donoho et al. and Eldar and her co-workers.

### 2.4.3 Effect of block length and dictionary type

In this experiment, assume that the dictionary characterisation of Block-MCC$_{2,2}$ is used. The goal is to investigate the sparsity levels for two types of random dictionaries without and with intra-block orthonormality, and for different dictionaries with equally-sized blocks $d$=$\{1, 2, 4, 5, 8, 10\}$. Finally, the sparsity levels for two types of random dictionaries computed for different values of $d$, are compared to the conventional sparsity levels of Donoho et al. and Eldar and her co-workers.

As it can be seen in figure 2.8, for Block-MCC$_{2,2}$ as dictionary characterisation, the proposed sparsity levels for random dictionaries without intra-block orthonormality and with full column rank blocks ($\mathbf{\Phi}$, hollow downward orange triangles) are higher than the conventional sparsity levels introduced by Donoho et al. ($\mathbf{\Phi}$, hollow green circles) and Eldar et al. ($\mathbf{\Phi}$, hollow upward blue triangles) for all values of $d$.

Except for the results of Donoho et al., the sparsity levels for random dictionaries with intra-block orthonormality are higher than for random dictionaries without intra-block orthonormality (with full column rank blocks). Sparsity level of Donoho et al. is invariable to two types of dictionaries $\mathbf{\Phi}$ and $\mathbf{\Phi}_{ort}$, and also to changes in $d$ (the corresponding two set of markers are superimposed). At last, for $d$=1, the sparsity levels in block domain are equal to the scalar domain.

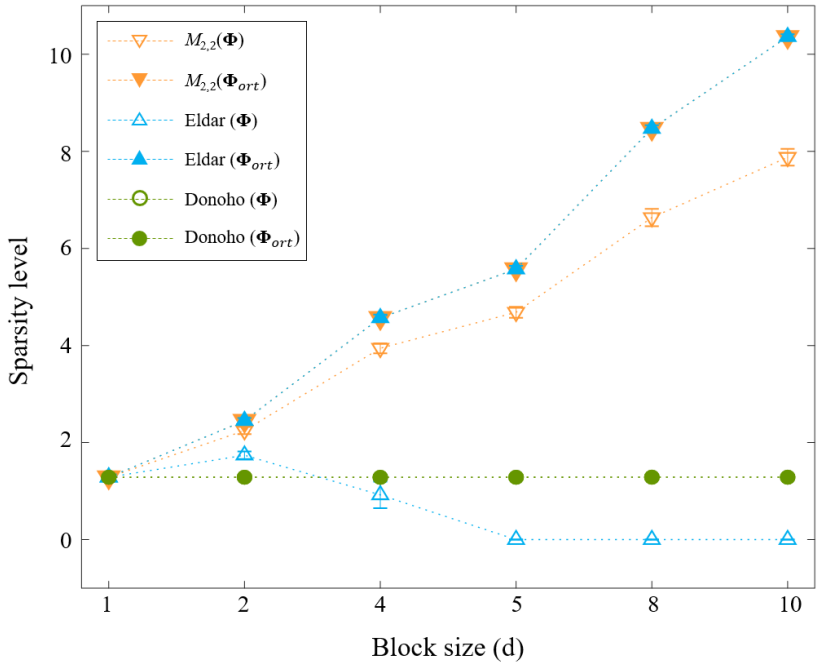



Figure 2.8: Sparsity levels for different values of block length $d$ for two types of random dictionaries without ($\mathbf{\Phi}$) and with ($\mathbf{\Phi}_{ort}$) intra-block orthonormality, compared to two conventional sparsity levels.

## 2.5 Conclusion

In this chapter, the sufficient conditions for unique representation of block-sparse recovery of an arbitrary signal $\boldsymbol{y}$ in a general arbitrary dictionary $\boldsymbol{\Phi}$ using general weighted (pseudo-)mixed-norm $P_{\boldsymbol{w};p_1,p_2}$ (and also $P_{p_1,p_2}$ as explained in Remark 2.1, $P_{\boldsymbol{w};p_1,p_2}$ v.s. $P_{p_1,p_2}$, page 50) optimisation problem, are proposed.

The dictionary is general (with full column rank blocks) and not restricted to be a union of two or more orthonormal bases, although we have some results in the special setting of the dictionary with intra-bock orthonormality. For theoretical conditions, weighted optimisation problems of $P_{\boldsymbol{w};p_1,p_2}$ (and also $P_{p_1,p_2}$ for $0\leq p_2\leq 1\leq p_1$) and $P_{\boldsymbol{w};p,0}$ (and also $P_{p,0}$ for $p>0$) are defined in their general form.

In addition, the proposed characterisations of Block-Spark, Block-NSP, Block-MCC$_{q,p}$, cumulative Block-MCC$_{q,p}$, and the properties are introduced in the general case.

Then, we demonstrated the relationship between the proposed Block-MCC$_{q,p}$, and conventional MCC and block-coherence proposed by Eldar et al. [EB09]; [EKB10b]; [EKB10a].

The proposed various properties and block-sparse uncertainty principles are defined in the general case and introduced to deduce recovery conditions for block-sparse representations.

We defined basic block-sparse uncertainty principle based on the proposed basic Block-MCC$_{q,p}$, upper-bounded it in terms of its conventional basic MCC, then lower-bounded in terms of $m$ and length of blocks, then showed its relationship with basic block-coherence of Eldar et al., and in the end we showed in which conditions the dictionary can be more block-incoherent relative to the conventional and existing block-wise cases.

We demonstrated that the proposed Block-ERC based on Block-Spark improves the conventional ERC based on Spark [DE03b]; [GN03b]. In addition, the proposed Block-NSP generalises the conventional NSP [DH01]; [EB01]; [GN03b]; [FN03a]; [Zha05]; [SXH08]; [CDD09] and another existing Block-wise NSP [SPH09], and Fusion NSP in special setting [BKR11].

Further, we showed that the proposed Block-ERC based on Block-MCC$_{q,p}$ improves the conventional ERC based on MCC [DH01].

Moreover, we improved the Block-ERC of Eldar et al. and demonstrated that our results algorithm-independently weaken the condition. In addition in special setting of $q=p=2$, intra-block orthonormality and equally-sized blocks our proposed Block-ERC and the results of Eldar et al. are equivalent [EB09].

Next, we proposed Block-MCC$_{q,p}$-based recovery conditions for convex problem $P_{\boldsymbol{w};p,1}$ (and $P_{p,1}$ for equally-sized blocks), where, $p\geq 1$.

Finally, we introduced the cumulative version of the block-coherence of Eldar et al. [EB09], named it cumulative inter-block coherence constant, demonstrated its relationship

with cumulative Block-MCC$_{q,p}$, and the inter-block coherence defined by Eldar et al., and proposed Block-ERC based on cumulative inter-block coherence constant.

All the contributions of this work are the natural generalisation of the existing conventional concepts, thus, they all reduce to the conventional ones for the unit block size, i.e., $\forall k$, $d_k$=1. Therefore, all the results are consistent with the previous findings.

In numerical experiments in Section 2.4, we showed that in computation of sparsity level, there are three types of parameters: block length $d_k$, existence of intra-block orthonormality, and $\ell_{q\to p}$ operator-norm in Block-MCC$_{q,p}$.

Then, by assuming that only two types of parameters are variable, we investigated the behaviour of sparsity level in three different experiments.

In this part, we are going to investigate the sparsity levels while all the three mentioned parameters are variable. In figure 2.9, the first twelve bars for each block size $d$, indicate $(d+M_{q,p}^{-1}(\mathbf{\Phi}))/2$.

The proposed sparsity levels are calculated for the six tractable basic operator-norms of $\ell_{1\to1}$, $\ell_{1\to2}$, $\ell_{1\to\infty}$, $\ell_{2\to2}$, $\ell_{2\to\infty}$, and $\ell_{\infty\to\infty}$, based on the table 2.1 (page 54). Then, repeated for dictionaries without ($\mathbf{\Phi}$) and with ($\mathbf{\Phi}_{ort}$) intra-block orthonormality.

The last four bars are the sparsity levels of the Block-ERC proposed by Eldar et al., i.e., $d$ times the right-hand side of the (1.20), and of the ERC proposed by Donoho et al., i.e., $(1+M^{-1}(\mathbf{\Phi}))/2$ defined in (1.13), for two types of dictionary, i.e., $\mathbf{\Phi}$ and $\mathbf{\Phi}_{ort}$.

As it can be seen in figure 2.9, all sparsity levels for $d$=1 converge to the conventional sparsity level of Donoho et al., whereas by increasing $d$, the sparsity level related to the Block-ERC, whether ours or the condition of Eldar et al., are significantly increased, which practically demonstrates the claim of weakened recovery conditions in the block-wise domain, which has been proved theoretically in Property 2.19 (SL v.s. Block-SL$_{q,p}$, page 74).

In addition, it can be seen in figure 2.9 that for $q$=$p$=2, our proposed sparsity level is equal to that of Eldar et al., which demonstrates the correctness of the claim in Lemma 2.3 and consequently Remark 2.15. As discussed in Remark 2.15 and can be seen in figure 2.9, for $q$=$p$=2 and intra-block orthonormality, the proposed sparsity level reduces to the sparsity level of Eldar and her co-workers.

Furthermore, for all variable parameters, we have $SL_{1,\infty}(\mathbf{\Phi})<SL_{1,2}(\mathbf{\Phi})<SL_{1,1}(\mathbf{\Phi})$, $SL_{1,\infty}(\mathbf{\Phi})<SL_{1,2}(\mathbf{\Phi})<SL_{2,2}(\mathbf{\Phi})$, $SL_{1,\infty}(\mathbf{\Phi})<SL_{2,\infty}(\mathbf{\Phi})<SL_{2,2}(\mathbf{\Phi})$, and $SL_{1,\infty}(\mathbf{\Phi}) < SL_{2,\infty}(\mathbf{\Phi}) < SL_{\infty,\infty}(\mathbf{\Phi})$, which is in agreement with Property 2.18 (Block-SL$_{q,p}$ inequalities).

In addition, it can be seen that for a given $q$ and $p$ pairs, the sparsity level of dictionaries $\mathbf{\Phi}_{ort}$ with intra-block orthonormality is higher than the sparsity level of dictionaries $\mathbf{\Phi}$ without intra-block orthonormality, i.e., conditions based on $\mathbf{\Phi}_{ort}$ are weaker.

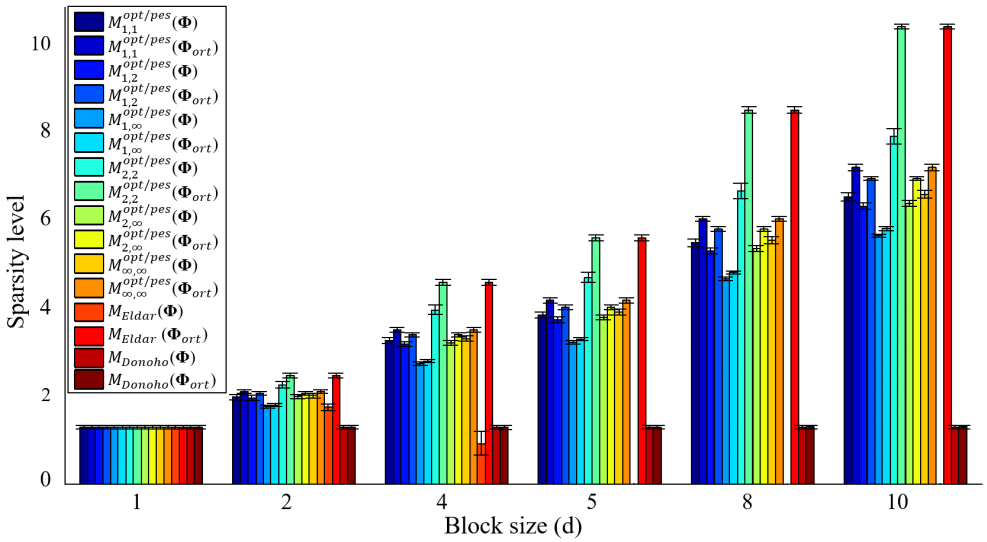



Figure 2.9: Sparsity level as a function of $d$ for different values of $q$ and $p$, averaged over 100 realisations of random dictionaries, with $m$=40 and $n$=400.

Future research should focus on:

- Introducing block-sparse recovery conditions based on the proposed cumulative Block-MCC$_{q,p}$ defined in Definition 2.7.
- Generalising the conventional dictionary characterisation of $\mu$ defined in [DE03b], to establish block-sparse recovery conditions.
- Transforming all the previously mentioned block-sparse *exact* recovery conditions to block-sparse *stable* recovery conditions. In stable or robust recovery conditions we have, $\|\boldsymbol{y}-\mathbf{\Phi}\hat{\boldsymbol{\beta}}\|_2<e$, where, $e$ is the noise level.
- Study on block-sparse optimisation algorithms, and the relationship between the *theoretical* and *algorithmic* block-sparse recovery conditions.

Chapter 3

# A general framework for block structure identification

Contents



## 3.1 Introduction

The structure of block-sparsity, which is defined in Section 2.2.1 had been assumed to be known. In other words, the block structure of the representation vector $\boldsymbol{\beta}$ and dictionary $\boldsymbol{\Phi}$ is assumed to be stored in $\boldsymbol{d}$=$[d_1, \cdots, d_K]$. However, in some problems the block structure prior knowledge might be not available. Therefore, the problem of block structure identification needs to be taken into account [Eks11].

In this chapter, the proposed block structure identification method is explained in Section 3.2. Next, the advantages of the proposed method including the improved recovery conditions and brain source space segmentation are investigated in sections 3.3 and 3.4, respectively. Then, some experiments on synthetic dictionary and real lead-field in Section 3.5 demonstrate the advantages of the proposed block structure identification method.

## 3.2 Block structure identification

As mentioned earlier, many real-world inverse problems are vastly underdetermined. In a fat coefficient matrix of a vastly USLE[1], the columns are more likely to be coherent. Therefore, computing any kind of dictionary characterisation based on inter-columns coherence leads to high values. As discussed in Chapter 2, because of the inverse relationship between the coherence characterisation of a general dictionary and the corresponding sparsity level in Block-ERC[2], the resulted Block-ERC would be tighter, which is not desired.

One possible solution to the issue of high coherence characterisation of dictionary or coefficient matrix in vastly USLE, is grouping the coherent columns of the dictionary to form some groups of columns, where, the columns inside each group or block are highly coherent. Then, computing the inter-blocks coherence characterisation leads to lower coherence characterisation of dictionary in comparison to the standard inter-columns coherence characterisation, because the resulted block of columns are more incoherent compared to the columns of dictionary. The proposed Block-MCC$_{q,p}$[3] is a suitable coherence characterisation of dictionary, due to its generality and flexibility in computing the coherence between differently-sized blocks of columns.

By clustering the coherent columns or blocks of columns of a given dictionary utilising general characterisation of Block-MCC$_{q,p}$ as the similarity measure, the block structure of the dictionary can be identified. Because each cluster of columns represents a block of the dictionary. In other words, clusters resulted from the clustering algorithm are equivalent to the blocks of the dictionary.

After applying clustering algorithm on the dictionary, each cluster, e.g., cluster $k$, contains some columns, where, the number of columns determines the size of corresponding block in the dictionary, i.e., $d_k$, therefore, the block structure $\boldsymbol{d}=[d_1, \cdots, d_K]$ is identified.

Each column in the dictionary $\boldsymbol{\Phi}\in\mathbb{R}^{m\times n}$ is multiplied to a corresponding element in the representation vector $\boldsymbol{\beta}\in\mathbb{R}^n$, i.e., $\boldsymbol{y}=\sum_{i=1}^{n}\beta_i\boldsymbol{\varphi}_i$. So, each column $\boldsymbol{\varphi}_i$ in dictionary is linked to a corresponding element in $\boldsymbol{\beta}$, i.e., $\beta_i$. Therefore, by determining the block structure of dictionary, the block structure of the representation vector is also identified.

By utilising the proposed clustering-based framework, we attain two main goals, which are represented graphically in figure 3.1 for a real-world problem of EEG[4]/MEG[5] source reconstruction USLE:

- improved Block-ERC, and
- support segmentation.

[1] *Underdetermined System(s) of Linear Equations*
[2] *Block-sparse Exact Recovery Condition(s)*
[3] $(q,p)$*-Block Mutual Coherence Constant*
[4] *ElectroEncephaloGraphy*
[5] *MagnetoEncephaloGraphy*

Assuming the EEG/MEG source reconstruction problem in figure 3.1, there are four brain sources of $s_1, \cdots, s_4$, as a simple example. The clustering tree is resulted from clustering the lead-fields of four sources. As shown graphically, the inter-cluster distance in clustering tree at the clustering level equivalent to three clusters is maximum. Then, the estimated number of clusters in the lead-field is three. As represented graphically in figure 3.1 (right part), the SL[%][6] for three clusters is higher than the SL[%] for four clusters, where, the clustering is not applied. So there is improvement in Block-ERC based on Block-MCC$_{q,p}$. The computation of different sparsity levels will be explained in Section 3.3.

The clustering structure corresponding to three clusters is $\mathbb{P}$=$\{s_1, s_2, [s_3, s_4]\}$, where, $s_3$ and $s_4$ are clustered to form a new cluster, while $s_1$ and $s_2$ are single-element clusters. Therefore, $\boldsymbol{d}$=$[3, 3, 6]$, because in EEG/MEG source reconstruction problem each source's lead-field corresponds to three columns in the whole lead-field matrix. The partitioning $\mathbb{P}$ of sources $s_1, \cdots, s_4$, indicates to the brain source space segmentation, as represented graphically in figure 3.1 (left part), and will be explained in detail in Section 3.4.

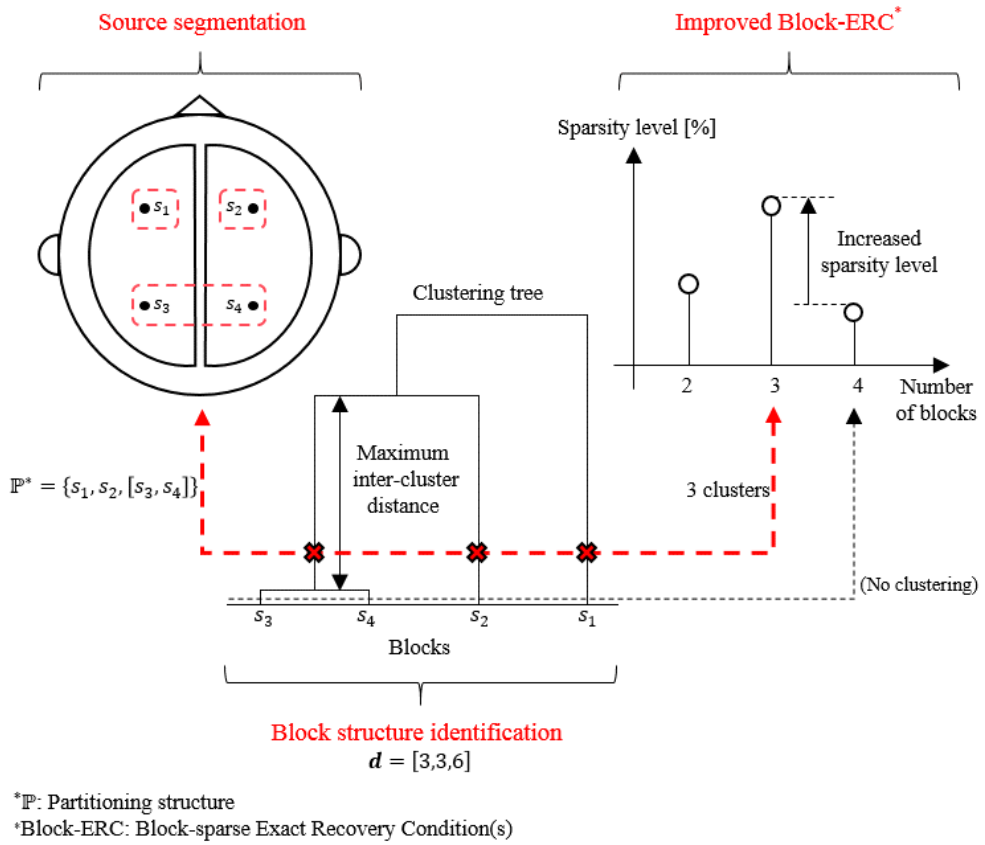



Figure 3.1: Clustering the coherent blocks using Block-MCC$_{q,p}$ determines the block structure, while improves Block-ERC and segments brain sources.

[6] *Sparsity Level*

### 3.2.1 Clustering coherent blocks of a general dictionary

Whether the block structure of a given dictionary is assumed to be known beforehand but it is not the optimal structure or it is unknown, by clustering coherent blocks in a bigger block, there would be the possibility of enhancing the structure of the dictionary as a preprocessing step.

In this chapter, we made use of standard agglomerative hierarchical clustering method, where, the proposed Block-MCC$_{q,p}$ is used as the measure of similarity or coherence. Hierarchical clustering provides a set of clustering structures in multilevel hierarchy manner, which allows to decide the level of clustering that is most suitable to the application of interest. The set of clustering structures is called clustering tree or *dendrogram*. The dendrogram consists of some nodes indicating the clusters, and usually the clustering level corresponding to the maximum inter-node distance is selected as the desired clustering level.

In figure 3.2, it has been assumed that before blocks clustering all of the blocks of the dictionary $\boldsymbol{\Phi}\in\mathbb{R}^{8\times 12}$ share the same length of two, i.e., $d_1=\cdots=d_6=2$. Then, by applying hierarchical clustering on initial blocks of dictionary and choosing the clustering level corresponding to the maximum inter-node distance in the clustering tree, i.e., three clusters, the clustered representation appears.

As it can be seen in figure 3.2, after blocks clustering, blocks $\boldsymbol{\Phi}[1]$, $\boldsymbol{\Phi}[4]$, and $\boldsymbol{\Phi}[6]$ are concatenated to form a new block $\boldsymbol{\Phi}_{cls}[1]$ in new representation, whereas $\boldsymbol{\Phi}[2]$ and $\boldsymbol{\Phi}[3]$ are concatenated to form $\boldsymbol{\Phi}_{cls}[2]$. Hence, the identified block structure is $\boldsymbol{d}=[6,4,2]$. Consequently, their corresponding blocks in the representation vector $\boldsymbol{\beta}$ will be clustered. The clustered representation of the new model, i.e., $\boldsymbol{y}=\boldsymbol{\Phi}_{cls}\boldsymbol{\beta}_{cls}$, leads to more relaxed Block-ERC, i.e., higher sparsity levels, which will be demonstrated in Section 3.5.

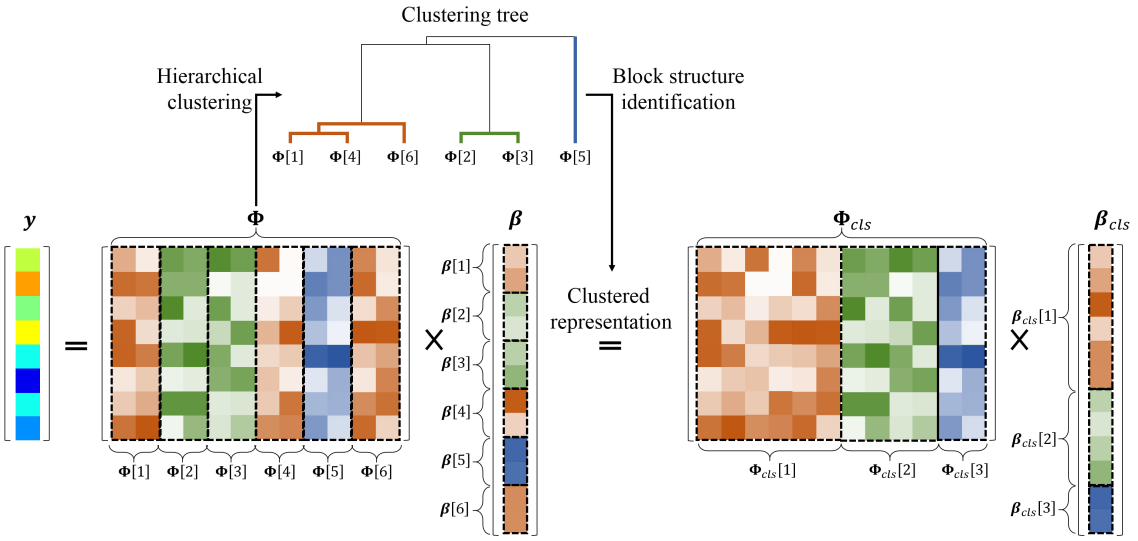



Figure 3.2: Clustered representation of general dictionary $\boldsymbol{\Phi}$ and representation vector $\boldsymbol{\beta}$.

## 3.3 Block-ERC in clustered representation

To investigate Block-ERC in clustered representation, we need to calculate Block-SL$_{q,p}$[7], which has inverse relationship with Block-MCC$_{q,p}$. Since blocks growing due to clustering can lead to a case where a big block covers whole space, we apply PCA[8] to clustered blocks before computing Block-MCC$_{q,p}$.

In order to compare the resulted Block-SL$_{q,p}$ to conventional sparsity level, we need to state the number of blocks of columns in terms of the number of columns. In a dictionary with equally-sized blocks, i.e., $d_1{=}\cdots{=}d_K{=}d$, the Block-SL$_{q,p}$ can be simply converted to sparsity level through multiplying Block-SL$_{q,p}$ by the length of each of the blocks, i.e., $d{\times}Block{-}SL_{q,p}(\mathbf{\Phi})$ or $\sum_{k=1}^{Block-SL_{q,p}(\mathbf{\Phi})} d_k$.

Next, we explain how to compute sparsity levels in a dictionary with differently-sized blocks.

### 3.3.1 Sparsity levels in differently-sized block structure of a dictionary

In a dictionary with differently-sized blocks, i.e., $d_1{\neq}\cdots{\neq}d_K$, it can be assumed a range of sparsity levels between $SL_{min}(\mathbf{\Phi}){=}\sum_{k=1}^{Block-SL_{q,p}(\mathbf{\Phi})} d_{(k)}$ and $SL_{max}(\mathbf{\Phi}){=}\sum_{k=K-Block-SL_{q,p}(\mathbf{\Phi})+1}^{K} d_{(k)}$, where, $d_{(k)}$ are elements of the ascendingly-sorted $\boldsymbol{d}$, i.e., $d_{(1)}{\leq}\cdots{\leq}d_{(K)}$. In other words, SL$_{min}$ and SL$_{max}$ are sum of the Block-SL$_{q,p}$ minimum and maximum elements of $\boldsymbol{d}$, respectively, as shown in figure 3.3.

**Remark 3.1.** For $d_1{=}\cdots{=}d_K$, the upper-bound and lower-bound of the range of sparsity level are equal, i.e., $SL_{max}(\mathbf{\Phi}){\equiv}SL_{min}(\mathbf{\Phi})$.

In some cases like figure 3.2, a block in clustered representation is in fact a concatenated version of a set of other blocks with the same size $d$, and not columns necessarily, i.e., $d{=}1$. This is the case in our real-world EEG and/or MEG source reconstruction problem, where, the

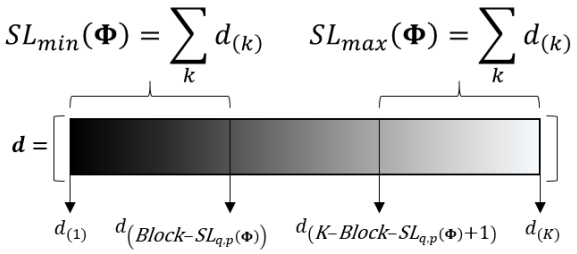



Figure 3.3: The computation of SL$_{min}$ and SL$_{max}$ from ascendingly-sorted $\boldsymbol{d}$.

[7] $(q,p)$-*Block-Sparsity Level*
[8] *Principal Component Analysis*

| Minimum | | Maximum | |
|---|---|---|---|
| $Block{-}SL_{min}(\mathbf{\Phi}) = 1$ | $SL_{min}(\mathbf{\Phi}) = 2$ | $Block{-}SL_{max}(\mathbf{\Phi}) = 3$ | $SL_{max}(\mathbf{\Phi}) = 6$ |

Table 3.1: Sparsity levels for the dictionary in figure 3.2, if $Block{-}SL_{q,p}(\mathbf{\Phi})$=1.

initial blocks are lead-fields with the same size of three. In this type of problem, the number of blocks of size $d$ or the blocks themselves have meaningful information. For instance, in our problem, each block of size $d$=3 represents a brain source.

Therefore, we can simply divide the number of elements to the length $d$ to provide another means of representing the results, which is the number of initial blocks. Then, considering maximum sparsity level in the most pessimistic case, i.e., $\text{SL}_{max}$, we can also represent the results in terms of maximum block-sparsity level, i.e., Block-$\text{SL}_{max}$, which is $SL_{max}(\mathbf{\Phi})/d$.

Therefore, we can consider the following basic types of sparsity level and Block-$\text{SL}_{q,p}$ for the maximum range:

- Block-$\text{SL}_{q,p}$: Number of differently-sized blocks resulted from $(1+(d_{max}M_{q,p}(\mathbf{\Phi}))^{-1}\min_k \min\{1, d_k^{1/q-1/p}\})/2$, (Theorem 2.3, page 73).
  - $\text{SL}_{min}$: Number of columns resulted from $\sum_{k=1}^{Block-SL_{q,p}(\mathbf{\Phi})} d_{(k)}$.
    - Block-$\text{SL}_{min}$: Number of equally-sized blocks of size $d$ resulted from $SL_{min}(\mathbf{\Phi})/d$.
  - $\text{SL}_{max}$: Number of columns resulted from $\sum_{k=K-Block-SL_{q,p}(\mathbf{\Phi})+1}^{K} d_{(k)}$.
    - Block-$\text{SL}_{max}$: Number of equally-sized blocks of size $d$ resulted from $SL_{max}(\mathbf{\Phi})/d$.

As an example, consider the dictionary in figure 3.2 with Block-$\text{SL}_{q,p}$ equal to one. Then, the mentioned different sparsity levels are shown in table 3.1.

## 3.4 EEG/MEG source reconstruction problem and USLE

It was explained in Section 1.2.5 that the solution of the EEG/MEG forward problem is called lead-field, denoted by $\boldsymbol{\Phi}_{\mathbf{EEG/MEG}} \in \mathbb{R}^{m \times n}$ , which linearly relates the activities in source, denoted by $\boldsymbol{\beta}$, and EEG/MEG sensor space, denoted by $\boldsymbol{y}_{\mathbf{EEG/MEG}}$. In other words, we have $\boldsymbol{y}_{\mathbf{EEG/MEG}} = \boldsymbol{\Phi}_{\mathbf{EEG/MEG}} \boldsymbol{\beta}$.

As mentioned earlier in Section 1.2.3, the inverse problem involves computing the parameters of the brain sources. Since the number of the sensors, i.e., $m$, is much less than the number of the brain sources, i.e., $K$, there would be infinitely many states of the sources that lead to the same electromagnetic fields, i.e., $\boldsymbol{y}_{\mathbf{EEG/MEG}} = \boldsymbol{\Phi}_{\mathbf{EEG/MEG}} \boldsymbol{\beta}_i$, where $i \in \{\mathbf{1}, \cdots, \boldsymbol{\infty}\}$. Hence, the EEG/MEG source reconstruction problem is ill-posed.

To make the problem solvable with a unique solution, a dipole source model can be applied to the brain current sources. In other words, a population of cerebral neurons can be represented by a single dipole. A dipole can be represented mathematically as a vector. Considering the 3D Cartesian coordination, for each dipole we have six parameters to be determined.

- Three parameters are related to the dipole position in $\boldsymbol{i}$, $\boldsymbol{j}$, and $\boldsymbol{k}$ direction, which by meshing the brain, these positions can be also limited to some predefined discrete values placed in the vertices of brain mesh.
- Three parameters characterise the power of the dipole in each of three directions, i.e., the magnitudes of the dipole moment in $\boldsymbol{i}$, $\boldsymbol{j}$, and $\boldsymbol{k}$ direction.

For instance in figure 3.4, the dipole located in $s_4$ is related to three consecutive entries in the source activity vector $\boldsymbol{\beta} \in \mathbb{R}^n$, where $K = n/3$.

Assuming only four brain source positions $s_1, \cdots, s_4$ and three electric or magnetic sensors in a simple case, figure 3.4 represents EEG and MEG linear models.

As it can be seen in figure 3.4, regardless of the head electromagnetic properties, and in terms of the parameters of the dipole source model, each element of EEG lead-field is a function $f$ of direction vectors $\boldsymbol{i}$, $\boldsymbol{j}$, or $\boldsymbol{k}$ and the distance between the position of corresponding source and sensor, e.g., the EEG lead-field for source $s_1$ is represented in figure 3.4. Each consecutive three elements in a row belongs to a certain sensor and source. Similarly, in terms of the parameters of the dipole source model, each element of MEG lead-field is a function $g$ of direction vectors $\boldsymbol{i}$, $\boldsymbol{j}$, or $\boldsymbol{k}$ and the distance between the position of corresponding source and sensor, e.g., the MEG lead-field for source $s_2$ is represented in figure 3.4.

Therefore, due to discretising and linearising effect of lead-field and having much less sensors than sources, an EEG/MEG source reconstruction problem can be stated in a form of USLE.

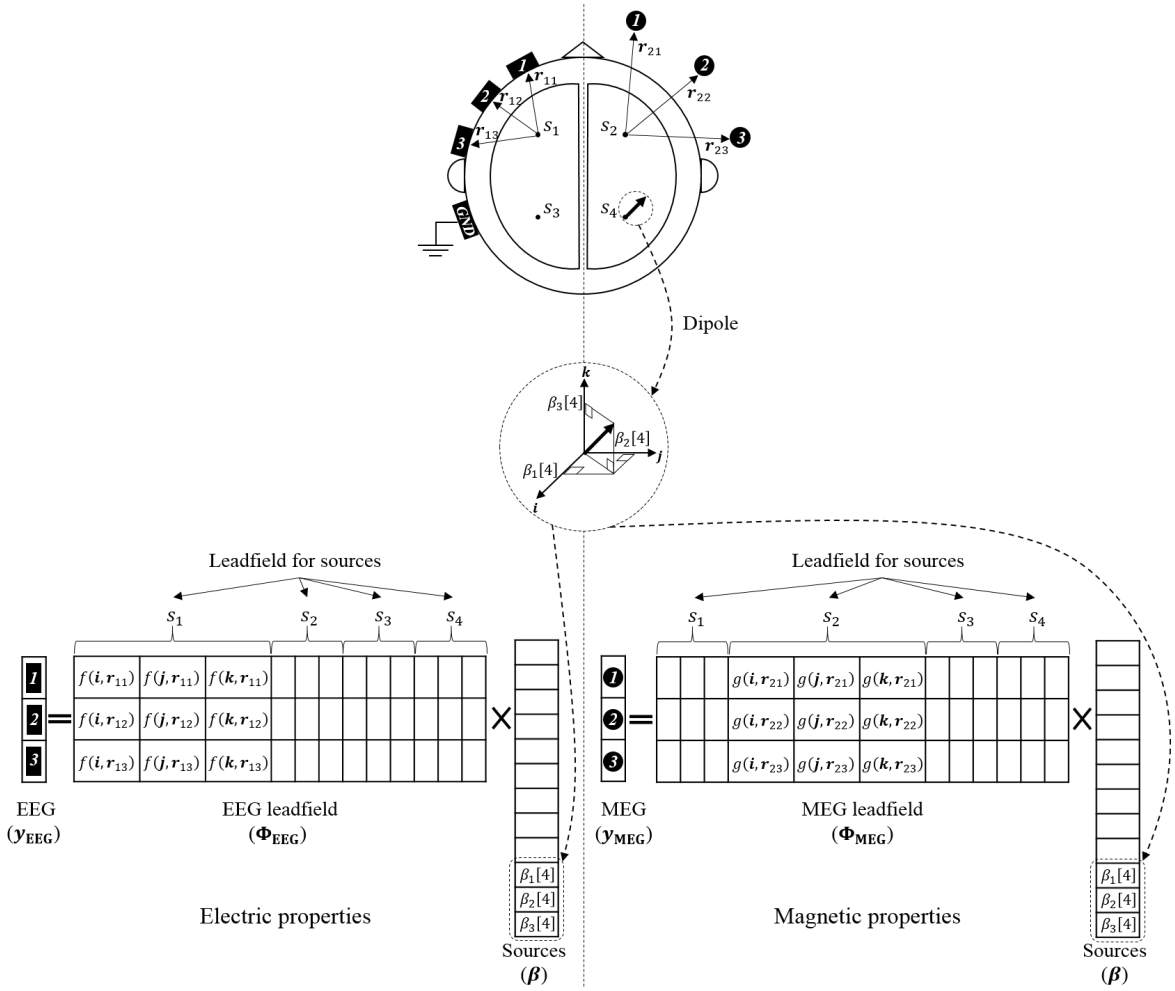



Figure 3.4: A simple EEG (left) and MEG (right) linear models for four brain sources $s_1, \cdots, s_4$, and three electric and magnetic sensors, respectively. A 3D dipole located in $s_4$ is shown and represented mathematically.

### 3.4.1 EEG/MEG source space segmentation

As mentioned earlier, the EEG/MEG source reconstruction problem can be modelled as an USLE. In our real-world problem, the dictionary in USLE is called the lead-field matrix. the whole lead-field matrix, is built by horizontal concatenation of lead-fields of all sources in the source space. As described before, the lead-field matrix of a single source is a $m$ by 3 matrix, where, $m$ is the number of sensors in the sensor space. Hence, a whole lead-field matrix is composed of consecutive three-column individual lead-field matrices.

On the other hands, as mentioned before, in the EEG/MEG linear model each of three-column individual lead-field matrices $\boldsymbol{\Phi}[k]$, $\forall k$, are multiplied to the activity of corresponding dipole in the source vector $\boldsymbol{\beta}$ to construct the activity of the sensor space, i.e., $\boldsymbol{y}=\sum_{k=1}^{K}\boldsymbol{\Phi}[k]\boldsymbol{\beta}[k]$. Then, each block of dictionary, $\boldsymbol{\Phi}[k]$, $\forall k$, is the direct coefficient of its corresponding source activity.

Therefore, by clustering the coherent lead-field matrices $\boldsymbol{\Phi}[k]$ of the whole lead-field matrix $\boldsymbol{\Phi}$, their corresponding blocks in the source activity vector $\boldsymbol{\beta}$ will also be clustered. Hence, the sources will be grouped and will form some segments in the source space.

To realise the mentioned idea, we propose the general Block-MCC$_{q,p}$ dictionary characterisation, which measures the similarity between the blocks of the dictionary. Then, Block-MCC$_{q,p}$ can be used in the clustering step to make similarity matrix.

As represented graphically in figure 3.5, the six sources in the brain are clustered in different three segments based on the clustering of their corresponding lead-fields in the lead-field matrix. In other words, clustering of the coherent lead-field matrices $\boldsymbol{\Phi}[1]$, $\boldsymbol{\Phi}[4]$, and $\boldsymbol{\Phi}[6]$ leads to clustering of their corresponding sources $\boldsymbol{\beta}[1]$, $\boldsymbol{\beta}[2]$, and $\boldsymbol{\beta}[3]$, and so on.

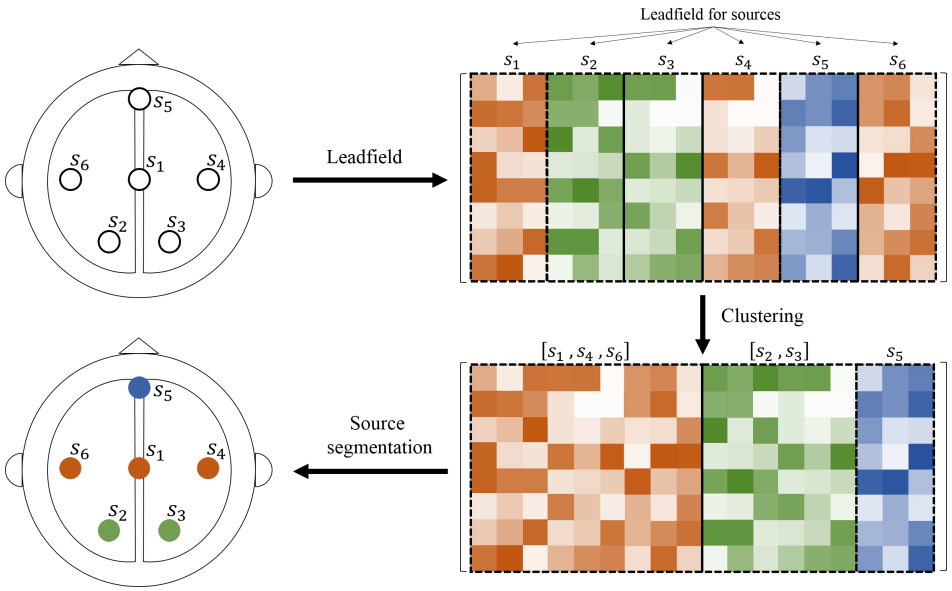



Figure 3.5: Segmenting the source space of brain using clustering the corresponding lead-fields.

## 3.5 Results

Assuming that the number of clusters is not known, a strategy will be needed to appropriately estimate it, which leads us to hierarchical clustering analysis. Hierarchical clustering can be visualized by the so-called clustering tree or dendrogram. The clustering tree is composed of nodes (clusters) and branches, which the length of each branch shows the distance between the two nodes. A scale with the largest distance between consecutive nodes or longest branch can be considered as an appropriate scale for clustering, which gives the estimated number of clusters. The agglomerative hierarchical clustering algorithm is used. In agglomerative as opposed to divisive method, each single data is considered as a cluster, then at each clustering level the clusters are successively merged until all the data are clustered into a single cluster.

Suppose that in figure 3.6(a), each block $\mathbf{\Phi}[k]$ of dictionary is represented by a single circle, and there are 100 blocks, i.e., $k \in \{1, \cdots, 100\}$. The relative distances between circles are determined by the proposed coherence measure Block-MCC$_{2,2}$. The random dictionary is generated in a way to have five equally-sized clusters of blocks of length 20. Then, $\mathbf{\Phi}[1], \cdots, \mathbf{\Phi}[20]$, construct cluster I and so on. The process of generating desired random dictionaries is explained in Section 3.5.1. A method called *complete* has been used to compute the distance between two clusters, which is equal to the longest distance between two points in the two clusters as illustrated in figure 3.6(a). In figure 3.6(b) the dendrogram resulted from clustering the random dictionary is shown. As expected, the magenta dotted horizontal line in the scale with the largest distance between consecutive nodes intersects the dendrogram five times, so the estimated number of clusters is five, which is true.

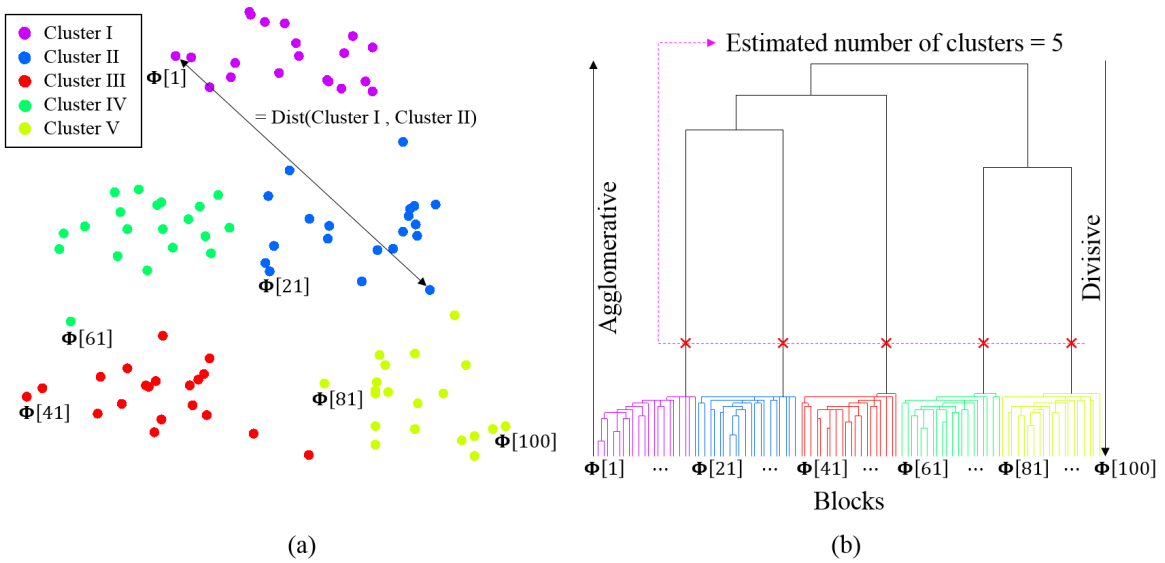



Figure 3.6: (a) The longest distance between two points in the two clusters determines inter-cluster distance; (b) Number of vertical branches at a clustering level with longest branch approximates the number of clusters.

### 3.5.1 Synthetic dictionary

Dictionaries $\boldsymbol{\Phi}\in\mathbb{R}^{m\times n}$, with equally-sized blocks of length $d_1=\cdots=d_K=d$, are generated from independent and identically distributed (i.i.d.) random variables and all columns of the dictionaries are normalized to have unit $\ell_2$ norm. The idea to generate the whole dictionary is first to start from an initial block, next some representative blocks of clusters are generated, then from each representative block of a cluster, the blocks belonging to the same cluster are produced. To construct a second block from a first one, the random matrix $\exp((\boldsymbol{C}-\boldsymbol{C}^T)\sqrt{2}\varepsilon/\|\boldsymbol{C}-\boldsymbol{C}^T\|_F)$ is multiplied to the first block, where, $\boldsymbol{C}$ is a square random matrix of dimension $m$. Depending on the role of $\varepsilon$ whether to generate representative blocks of clusters or blocks of dictionary, it is named $\varepsilon_{inter}$ or $\varepsilon_{intra}$, respectively.

In figure 3.7, the process of generating a random dictionary $\boldsymbol{\Phi}$ with $N$ clusters and $K$ blocks is illustrated. First, from a random initial block $\boldsymbol{\mu}^0\in\mathbb{R}^{m\times d}$, $N$ representatives of clusters $\boldsymbol{\mu}^1,\cdots,\boldsymbol{\mu}^N$ are created, using the mentioned multiplier and parameter $\varepsilon_{inter}$. Then, from each representative $\boldsymbol{\mu}^i$, $\forall i$, the blocks $\boldsymbol{\Phi}^i[j]$, $\forall j$, belonging to the same cluster $i$ are generated, using the mentioned multiplier and parameter $\varepsilon_{intra}$. The parameter $\varepsilon$ in the multiplier controls the amount of overlap and similarity between blocks. In fact, $\varepsilon_{inter}$ defines the overlap between the representative blocks of clusters, whereas $\varepsilon_{intra}$ controls the overlap between the blocks belonging to the same cluster. Finally, the results are shown in terms of average and standard deviation over 100 repetitions of random dictionary $\boldsymbol{\Phi}\in\mathbb{R}^{10\times 80}$ generation with $d_j=2$, $\forall j$.

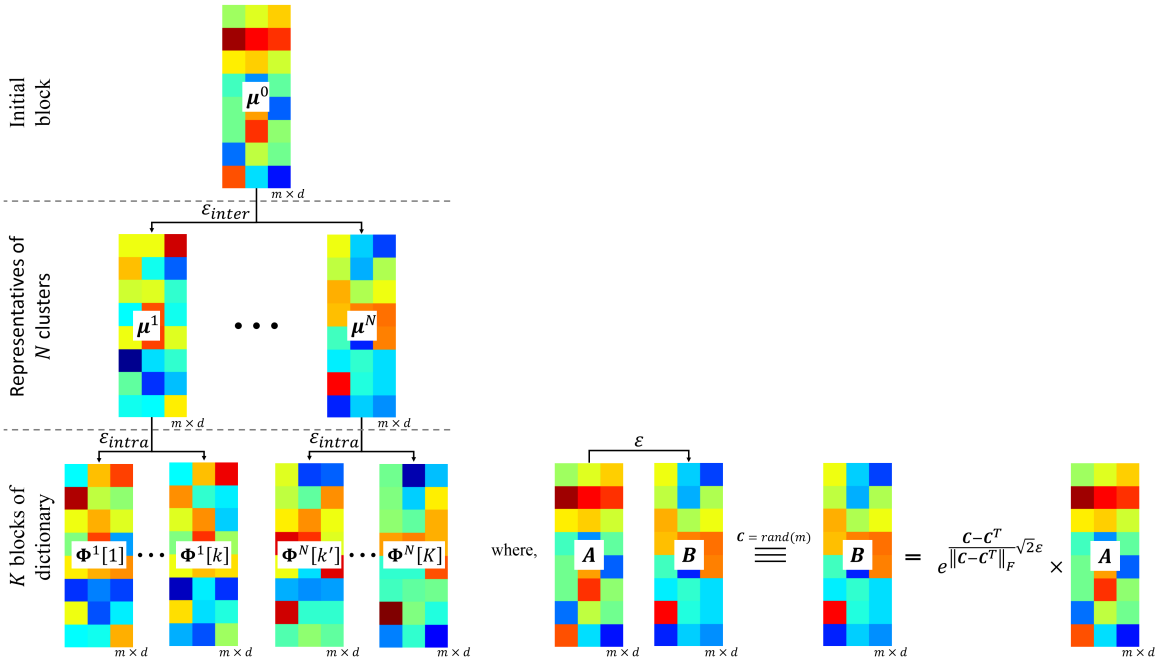



Figure 3.7: From initial block $\boldsymbol{\mu}^0$, $N$ representative blocks of clusters are computed. Then, from each representative block $\boldsymbol{\mu}^i$, all blocks $\boldsymbol{\Phi}^i[j]$ belonging to the same cluster $i$ are generated.

**Block-ERC in clustered representation and the number of clusters** In this experiment, the agglomerative hierarchical clustering algorithm is applied on the blocks of a random dictionary with certain number of clusters, whereas the Block-MCC$_{2,2}$ and complete method are used to measure the inter-blocks coherence and inter-clusters distance, respectively. The mentioned certain number of clusters is once set to four and once set to eight.

As it can be seen in figure 3.8, by applying hierarchical clustering on the blocks of the dictionary, i.e., by decreasing the number of clusters, there exists at least one clustering level in which the relative Block-SL$_{2,2}$ in the most pessimistic case, i.e., $Block{-}SL_{2,2}(\mathbf{\Phi})[\%]$, increases in comparison to when the clustering is not applied on the dictionary, i.e., the rightmost part of each diagram corresponding to 40 clusters. Therefore, clustering coherent blocks of the dictionary improves the Block-ERC through increasing the Block-SL$_{2,2}$.

In addition, for $\varepsilon_{inter}>1$ and $\varepsilon_{intra}<0.1$, the $Block{-}SL_{2,2}(\mathbf{\Phi})[\%]$ has a peak in a clustering level equal to the number of the clusters in the simulated dictionary. In fact, in figure 3.8 when there is four clusters in the simulated dictionary (blue curve, square markers), the maximum is at the fourth clustering level and also the consistent results are obtained for eight clusters in the dictionary (red curve, circle markers). For the clustering level lower than the optimal level, the space is under-sampled and there would be a block spanning the whole space, whereas for the clustering level higher than the optimal level, the space is over-sampled and the over-partitioning leads to high coherence measure.

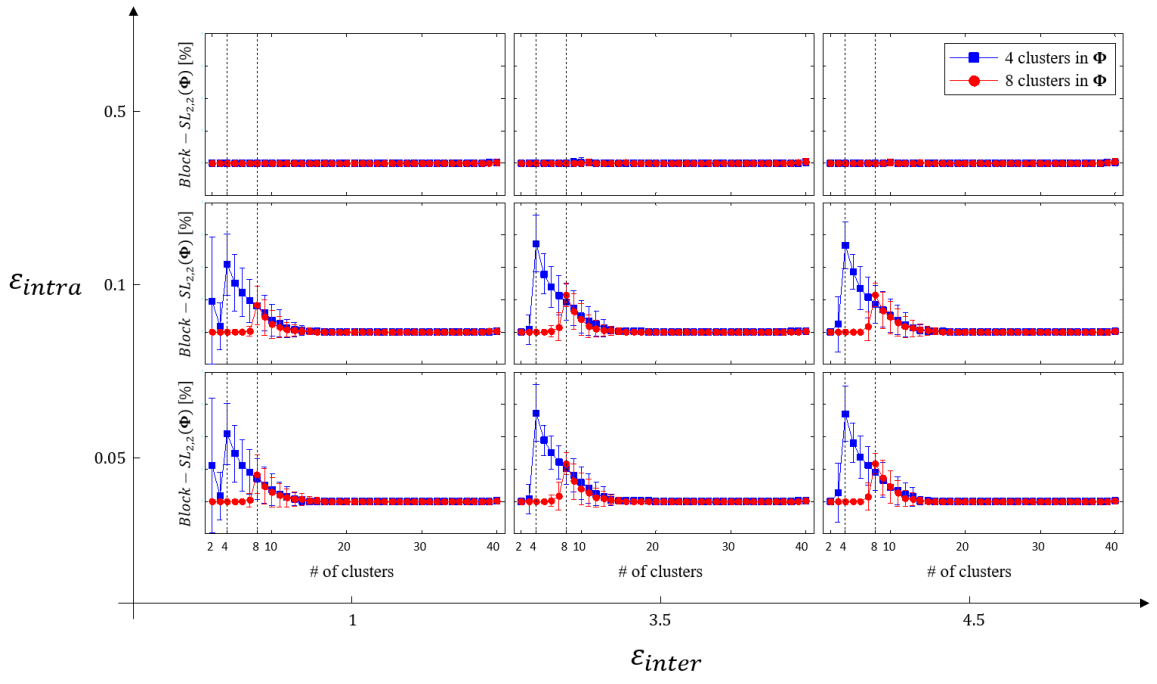


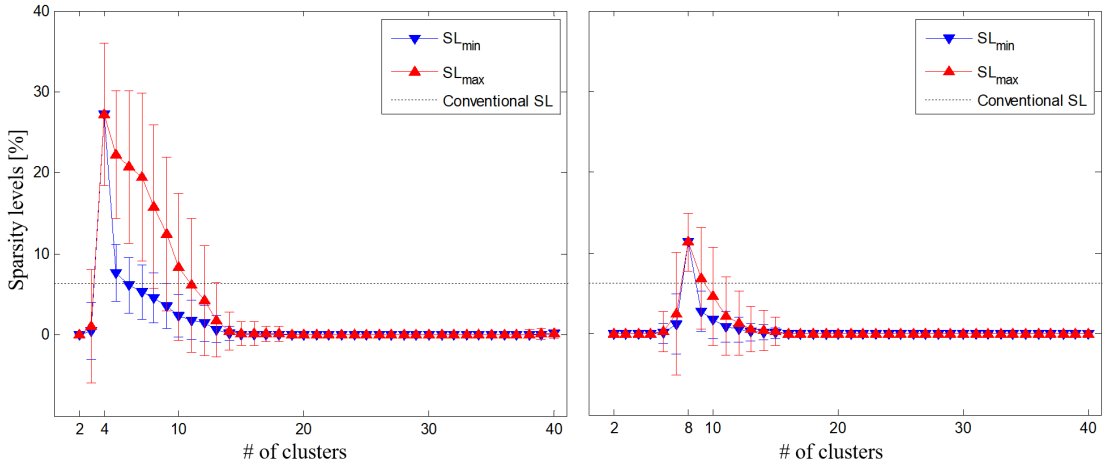

Figure 3.8: $Block{-}SL_{2,2}(\mathbf{\Phi})[\%]$ for each level of clustering computed for complete method, Block-MCC$_{2,2}$, $d$=2, $N$=$\{4,8\}$, and different values of $\varepsilon_{inter}$ and $\varepsilon_{intra}$ for simulating dictionary $\mathbf{\Phi}\in\mathbb{R}^{10\times 80}$.

**Block-ERC in clustered representation and the conventional ERC** In this experiment, the goal is to compare the sparsity level obtained in the proposed Block-ERC to the conventional sparsity level which is equal to the half of the number of rows of $\mathbf{\Phi}$.

First, we need to express the block-sparsity level in terms of the conventional sparsity level as described in Section 3.3, e.g., $\text{SL}_{min}$ and $\text{SL}_{max}$. Then, consider one of the cases in figure 3.8, e.g., $\varepsilon_{inter}$=3.5, $\varepsilon_{intra}$=0.1. So, the complete method is used to calculate inter-clusters distance, whereas the Block-$\text{MCC}_{2,2}$ measures the inter-blocks coherence of a dictionary $\mathbf{\Phi}\in\mathbb{R}^{10\times 80}$ with equally-sized blocks, i.e., $d_1=\cdots=d_{40}=d=2$.

In figure 3.9, it can be seen that by applying hierarchical clustering on the blocks of the dictionary, the increase in the sparsity levels even surpasses the conventional sparsity level which is marked with a black dotted line. It can be seen in figure 3.9 that close to the clustering level equal to the number of clusters in the simulated dictionary, even the proposed $\text{SL}_{min}$ can surpass the conventional sparsity level. Therefore, clustering the coherent blocks of a dictionary in addition to enhancing the Block-ERC through increasing the Block-$\text{SL}_{2,2}$, improves the conventional sparsity level. Hence, improves the conventional ERC.

At last, in order to make sure that the clustering structure in a clustering level, where, there is a peak in sparsity level is done properly, the average clustering accuracy over 100 repetitions is computed for fourth (for the case of four clusters in the dictionary) and eighth (for the case of eight clusters in the dictionary) clustering levels, which are equal to 99.65% and 100%, respectively. Therefore, the peak in the sparsity level diagram, which is more striking in figure 3.8 in addition to giving an estimation of the number of clusters in the dictionary, also with high probability (at least in this example) corresponds to the correct clustering structure of the dictionary.




(a) Four clusters in $\mathbf{\Phi}$ (b) Eight clusters in $\mathbf{\Phi}$

Figure 3.9: $\text{SL}_{min}$[%] and $\text{SL}_{max}$[%] for each level of clustering computed for complete method, Block-$\text{MCC}_{2,2}$, $d$=2, $\varepsilon_{inter}$=3.5 and $\varepsilon_{intra}$=0.1 for (a) $N$=4 and (b) $N$=8 in the dictionary $\mathbf{\Phi}\in\mathbb{R}^{10\times 80}$.

### 3.5.2 Real EEG/MEG lead-field

In this experiment the hierarchical clustering is used because of its characteristic in estimation of the number of clusters, as described in section 3.5. In order to simulate or import real-world sensor, source and volume conduction head model required to generate the real lead-field we used an open source software package MATLAB-based toolbox named FieldTrip [Oos+11].

EEG lead-field is obtained using 32 standard electrodes and utilising three different volume conduction head models of three-layer concentric spheres, realistic three-layer with inflated cortical sheet obtained by boundary element methods and realistic three-layer with highly-folded cortical sheet obtained by SPM8 software [Spm]. Realistic head models are obtained by imposing anatomical constraint, which is provided by MRI[9].

Similarly, for MEG lead-field generating, 32 standard sensors closest to the previously selected 32 EEG standard electrodes are selected and single sphere, realistic single layer with inflated and highly-folded cortical sheets are used as volume conduction head models. The sensor model, three-compartment (brain, skull, and scalp) volume conduction head model (analytical and realistic), and source model (spherical and realistic) are shown in figure 3.10.

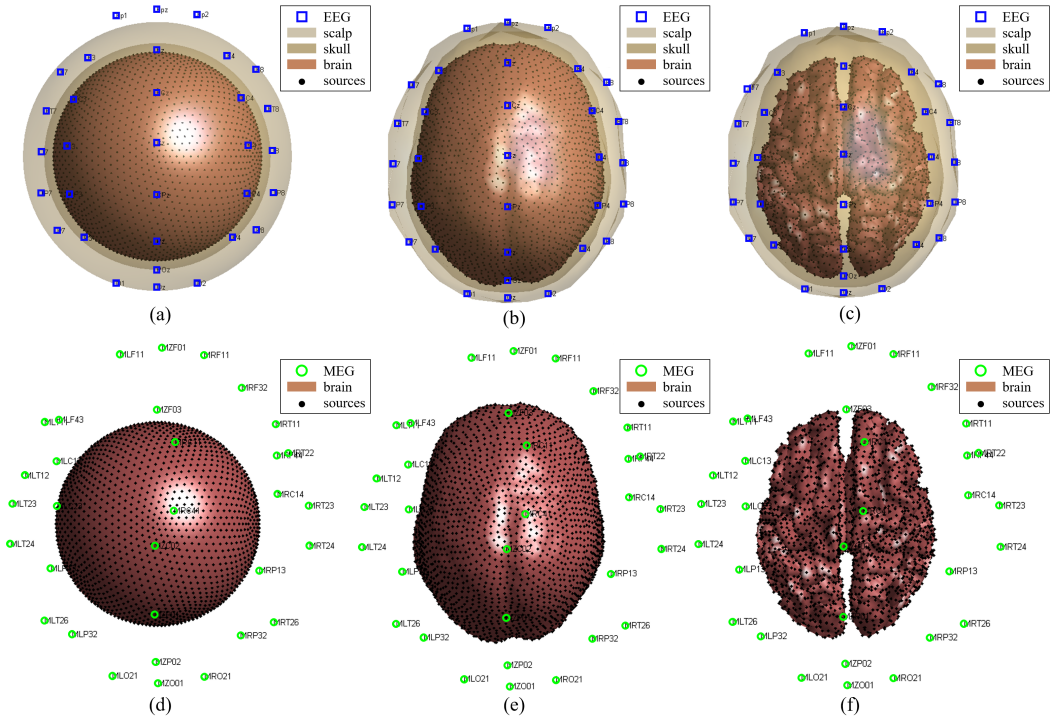



Figure 3.10: Different geometrical models of the EEG/MEG experiment: 1- Spherical (a,d), realistic inflated (b,e), and realistic highly-folded (c,f) source models; 2- EEG (a,b,c), and MEG (d,e,f) sensor models; 3- Three-layer concentric spheres (a), realistic three-layer (b,c), single sphere (d), and realistic single layer (e,f) volume conduction head models.

[9] *Magnetic Resonance Imaging*

**Block-ERC in clustered representation** As explained in Section 3.3, generally, in transforming the Block-$SL_{q,p}$, which is determined by Block-ERC, into the conventional sparsity level, different types of sparsity levels can be appeared, e.g., minimum sparsity level $SL_{min}$, and maximum sparsity level $SL_{max}$. On the other hand, as described earlier in the Section 3.5, number of branches in a clustering level with maximum inter-node distance in a clustering tree resulted from a hierarchical clustering algorithm can be used as an estimation of the number of clusters of a dictionary.

Assume the sparsity level computed in the clustering level corresponding to the maximum inter-node distance in the clustering tree is shown by $SL_2$, whereas the sparsity level computed for the lead-field without clustering is called $SL_1$. In order to investigate the effect of clustering the coherent blocks of the lead-field on sparsity levels, we compute the relative change quantity, i.e., $(SL_2(\mathbf{\Phi}) - SL_1(\mathbf{\Phi}))/SL_1(\mathbf{\Phi})$.

In figure 3.11, the relative change of sparsity levels $SL_{min}$, and $SL_{max}$ is shown. In addition, the experiment is repeated for two modalities of EEG and MEG, and three head models with spherical, realistic inflated and realistic highly-folded cortical sheets.

As it can be seen in figure 3.11, by clustering coherent blocks of the lead-field (whether EEG/MEG or spherical/realistic cortical sheet) using the proposed Block-$MCC_{2,2}$, all sparsity levels significantly increase, especially $SL_{max}$. Hence, clustering coherent blocks of lead-field using the proposed Block-$MCC_{q,p}$ coherence measure leads to improved Block-ERC.

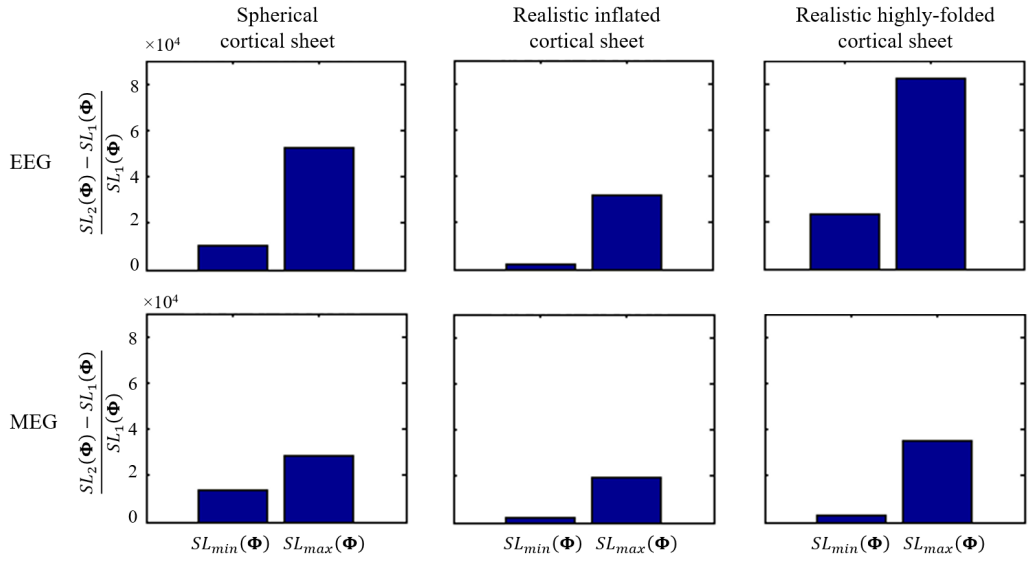



Figure 3.11: Relative increase of sparsity levels $SL_{min}$, $SL_{max}$, when clustering the coherent blocks of lead-field, repeated for two modalities of EEG and MEG, and three head models with spherical, realistic inflated and realistic highly-folded cortical sheets.

**EEG/MEG source space segmentation** As described in Section 3.4.1, clustering the coherent blocks of the lead-field matrix results in some brain regions, because each block of the lead-field correspond to a single source position in the source space. Then, coherent source positions form brain regions, where, coherency is defined by the Block-MCC$_{q,p}$ coherence measure applied on the bocks of the lead-field.

The results of clustering coherent source positions using EEG and MEG lead-field, and for three head models with spherical, realistic inflated and realistic highly-folded cortical sheets are shown in figure 3.12. Although, in this experiment the number of clusters or brain regions is estimated from the clustering tree, but since there is a series of clustering structures in the hierarchical clustering analysis, by introducing extra information to the problem about the number of brain regions, there would be the possibility of having different brain regions. In other words, the resulted brain regions in figure 3.12 are not fixed and can be adapted based on the number of regions. In addition, the resulted brain regions are a function of Block-MCC$_{q,p}$ and inter-cluster distance method.

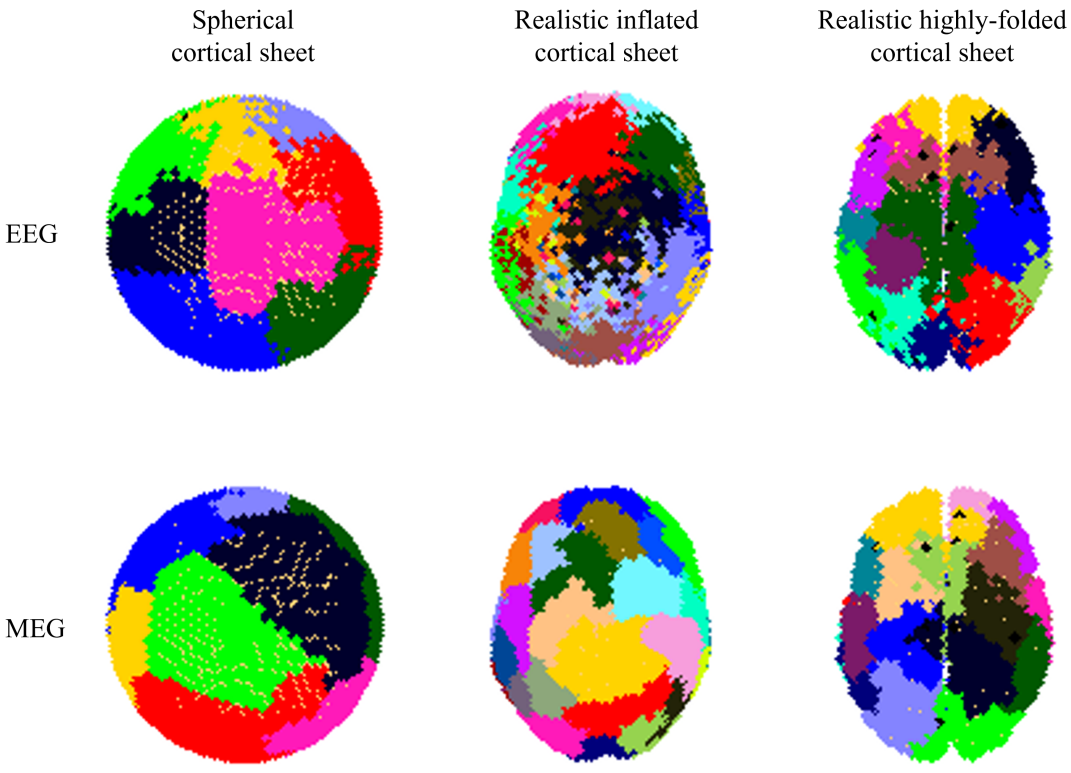



Figure 3.12: Brain source space segmentation, when clustering the coherent blocks of lead-field, repeated for two modalities of EEG and MEG, and three head models with spherical, realistic inflated and realistic highly-folded cortical sheets.

**Comparison to the conventional lobes of the brain** In order to compare the resulted brain source space segmentation in a specific clustering level to the conventional lobes of the brain, the closest boundaries to the conventional boundaries are shown in figure 3.13(a) and (b).

The brain segmentation and boundaries resulted from EEG, and MEG lead-fields are shown in figure 3.13(a), and (b), respectively. The anatomical lobes and the functional areas of the brain are shown in figure 3.13(c).

Therefore, using the proposed framework, i.e., clustering coherent blocks of the lead-field with the help of the proposed Block-MCC$_{q,p}$ coherence measure, the conventional structural and functional brain lobes can be divided into subject-specific refined regions.

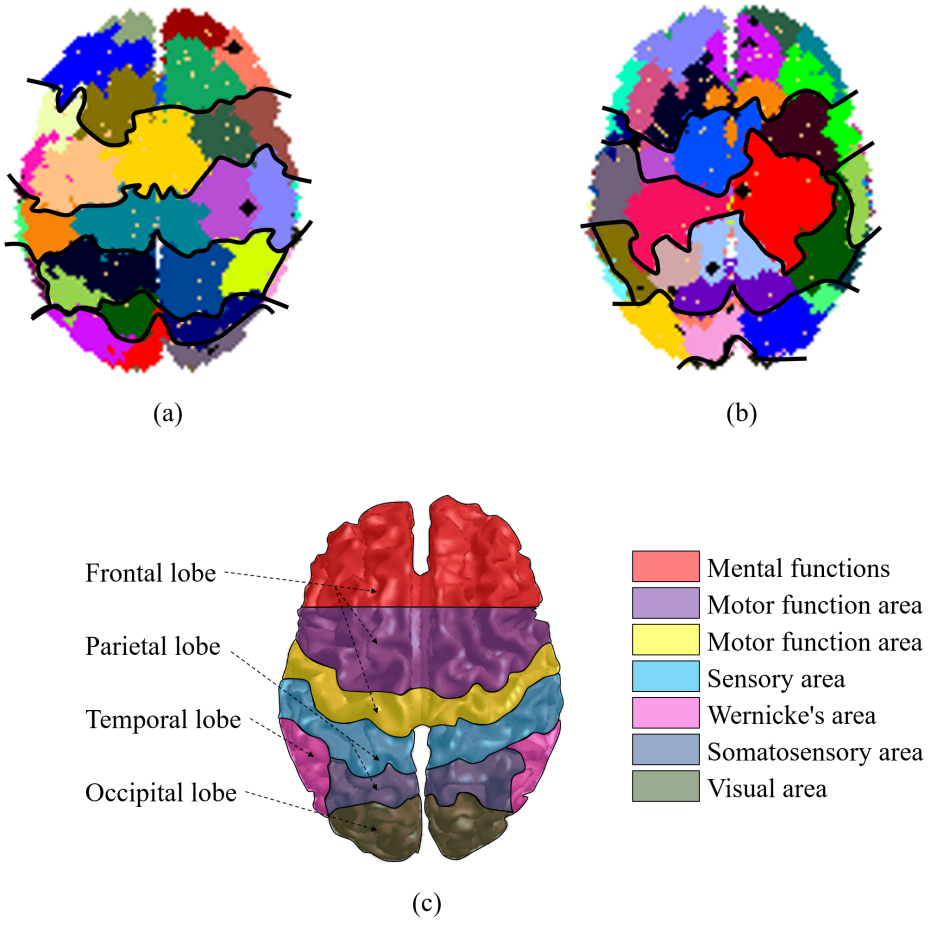



Figure 3.13: The proposed subject-specific brain source space segmentation, using (a) EEG, and (b) MEG lead-fields, compared to (c) the conventional anatomical and functional brain lobes.

## 3.6 Conclusion

In this chapter, we investigated the effect of clustering the coherent blocks of the dictionary on the Block-ERC and brain source space segmentation. The coherency is defined based on the proposed general characterisation of Block-MCC$_{q,p}$. To this aim, different experiments were done on synthetic dictionary and real EEG/MEG lead-field.

In experiments with synthetic dictionary, we demonstrated in figure 3.8 the positive effect of clustering the coherent blocks of the dictionary on the sparsity levels. Then, block-sparse exact recovery condition based on the block mutual coherence constant improves by clustering the coherent blocks. Furthermore, in figure 3.9 we showed that this improvement in the proposed block-sparse exact recovery conditions due to hierarchical clustering, even holds to the conventional exact recovery condition.

Finally, in experiments with real EEG/MEG lead-field in figure 3.11, we showed that the sparsity levels in different cases increase by clustering coherent lead-fields, hence, improves the corresponding block-sparse exact recovery condition. In figure 3.12, we showed that even by the mentioned clustering method, some consistent brain regions appear. By utilising the MRI of a person in calculating the realistic volume conduction head model, subject-specific brain regions appear. In fact, here we proposed a segmented structure of brain sources, making use of just the lead-field matrix in a general way, without being restricted to a special activity of dipoles $\boldsymbol{\beta_0}$ or EEG/MEG signals $\boldsymbol{y}$.

Ultimately, in this chapter for identifying the block structure of the dictionary or lead-field, a general framework based on the proposed block mutual coherence constant Block-MCC$_{q,p}$ is proposed, which at the same time improves the block-sparse exact recovery condition and provides a brain source space segmentation strategy.

Chapter 4

# EEG and MEG multi-modality

Contents



## 4.1 Introduction

As mentioned in Section 1.2, EEG[1] and MEG[2] source reconstruction problems solve a same brain source activity problem when acquired simultaneously [Mol+08]. In other words, for EEG or MEG signal $\boldsymbol{y}_{\mathbf{EEG/MEG}} \in \mathbb{R}^m$, EEG or MEG lead-field matrix $\boldsymbol{\Phi}_{\mathbf{EEG/MEG}} \in \mathbb{R}^{m \times n}$, and brain source activity $\boldsymbol{\beta}_{\mathbf{0}} \in \mathbb{R}^n$, we have the following EEG and MEG USLE [3]:

$$\boldsymbol{y}_{\mathbf{EEG}} = \boldsymbol{\Phi}_{\mathbf{EEG}} \boldsymbol{\beta}_{\mathbf{0}}, \qquad \text{and} \qquad \boldsymbol{y}_{\mathbf{MEG}} = \boldsymbol{\Phi}_{\mathbf{MEG}} \boldsymbol{\beta}_{\mathbf{0}}.$$

The key aim in this chapter is to integrate the complementary information of EEG and MEG modalities within the block structure identification framework introduced in Chapter 3 with application to brain source space segmentation.

In contrast to the brain source space segmentation studies in literature, which are integrating the information of EEG or MEG signals $\boldsymbol{y}_{\mathbf{EEG/MEG}}$ and lead-fields $\boldsymbol{\Phi}_{\mathbf{EEG/MEG}}$ into the problem, the proposed multi-modality strategy uses the block structure identification framework, which requires only the lead-field matrices, in order to demonstrate the benefit of merging EEG and MEG modalities at the same complexity or number of sensors.

[1] *ElectroEncephaloGraphy*
[2] *MagnetoEncephaloGraphy*
[3] *Underdetermined System(s) of Linear Equations*

## 4.2 Multi-modal lead-field

### 4.2.1 Structure and model

In order to investigate the impact of multi-modality, i.e., the integration of the complementary information of EEG and MEG modalities, we need to think of a strategy to combine the mono-modalities.

Suppose that the odd rows of USLE related to EEG generative model is represented as $\boldsymbol{y}^{O}_{\mathbf{EEG}}=\boldsymbol{\Phi}^{O}_{\mathbf{EEG}}\boldsymbol{\beta_0}$, and similarly the even rows of USLE related to MEG model as $\boldsymbol{y}^{E}_{\mathbf{MEG}}=\boldsymbol{\Phi}^{E}_{\mathbf{MEG}}\boldsymbol{\beta_0}$. Then, by concatenating the odd rows and even rows of EEG and MEG generative models, respectively, the EEG and MEG multi-modal model is as follows:

$$\begin{bmatrix} \boldsymbol{y}^{O}_{\mathbf{EEG}} \\ \boldsymbol{y}^{E}_{\mathbf{MEG}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Phi}^{O}_{\mathbf{EEG}} \\ \boldsymbol{\Phi}^{E}_{\mathbf{MEG}} \end{bmatrix} \boldsymbol{\beta_0}. \tag{4.1}$$

There are three reasons behind the mentioned strategy of integration of mono-modalities.

**1) Fixed number of sensors** Assuming that the number of sensors in both modalities is even, then each signal $\boldsymbol{y}^{O}_{\mathbf{EEG}}\in\mathbb{R}^{m/2}$ or $\boldsymbol{y}^{E}_{\mathbf{MEG}}\in\mathbb{R}^{m/2}$ has as half of observations as the original signal $\boldsymbol{y}_{\mathbf{EEG/MEG}}\in\mathbb{R}^{m}$, i.e., each modality contributes the same in terms of the number of sensors, so the comparison would be more fair compared to the odd number of one of original modalities' sensors. Then, by concatenating them, the number of equations in the resulted USLE in multi-modality would be equal to the the original USLE in mono-modality.

By keeping fixed the number of sensors in mono-modality and multi-modality, we assure that the probable added value is because of the multi-modality itself and not in reason of the increased observation. Because, in case of utilising all the equations in each of mono-modal USLE corresponding to EEG and MEG, the equations in final USLE would be more than the original USLE, i.e., $\boldsymbol{y}_{\mathbf{EMEG}}\in\mathbb{R}^{2m}$, where, $\boldsymbol{y}_{\mathbf{EMEG}}$ is the multi-modal observation. Therefore, we would have an increased input information in multi-modality in comparison to mono-modality, which would not lead to a fair comparison.

**2) Minimal change in sensors position** By drawing for example the odd rows of one modality and even rows of the other, i.e., without shared index for sensors position, we again ensure that the possible enhancement is due to the multi-modality itself and not because of major change in sensors position. Otherwise, if for example we select odd rows from both mono-modalities, the sensors layout in mono-modality would be significantly different from multi-modal layout. Because we are assuming that each MEG sensor is selected among the whole standard sensors, based on the proximity to an EEG sensor, to minimise the distance between each pair of EEG and MEG sensors. So, if the index set of selected sensors in one mono-modality does not have any shared sensor's index with the other mono-modality, the difference in sensors position in mono-modality and multi-modality will be minimised.

**3) Uniform sub-sampling in sensor space** By selecting even or odd rows of USLE, we are assuring that the sensors are uniformly distributed all over the head. Otherwise, by selecting for example the first half of one mono-modal sensors and the second half of the other mono-modal sensors, although the multi-modal sensors cover whole parts of the head, we would lose the information of one part of brain in mono-modality.

Therefore, considering the three mentioned reasons in multi-modality strategy, we are eliminating or reducing the side effects of increased observation, and optimum sensor position, so we can investigate the impact of multi-modality itself.

As mentioned before, the purpose of this chapter is to investigate the impact of multi-modal lead-field matrix within the block structure identification framework introduced in Chapter 3 with application to brain source space segmentation. Therefore, from the multi-modal USLE in (4.1), we only need to lead-field matrices $\mathbf{\Phi_{EEG}}$ and $\mathbf{\Phi_{MEG}}$, or precisely $\mathbf{\Phi}^{O}_{\mathbf{EEG}}$ and $\mathbf{\Phi}^{E}_{\mathbf{MEG}}$. In figure 4.1(c), the multi-modal lead-field $\mathbf{\Phi_{EMEG}}$ is built by concatenating the odd rows of $\mathbf{\Phi_{EEG}}$ (blue lead-field) and the even rows of $\mathbf{\Phi_{MEG}}$ (green lead-field).

As represented graphically in figure 4.1, number of multi-modal EMEG sensors in figure 4.1(c) is equal to the number of mono-modal EEG and MEG sensors in figure 4.1(a) and (b), which is equal to the number of rows of the lead-field matrices, i.e. $m$.

As mentioned before, the MEG sensors in figure 4.1 are the closest sensors to EEG sensors, e.g., MEG sensor #1 is the closest MEG sensor to EEG sensor #1, and so on.

The minimal change in sensors position in figure 4.1(c) in comparison to figure 4.1(a) and (b), also the uniform distribution of sensors, would be more evident by increasing the density of sensors.

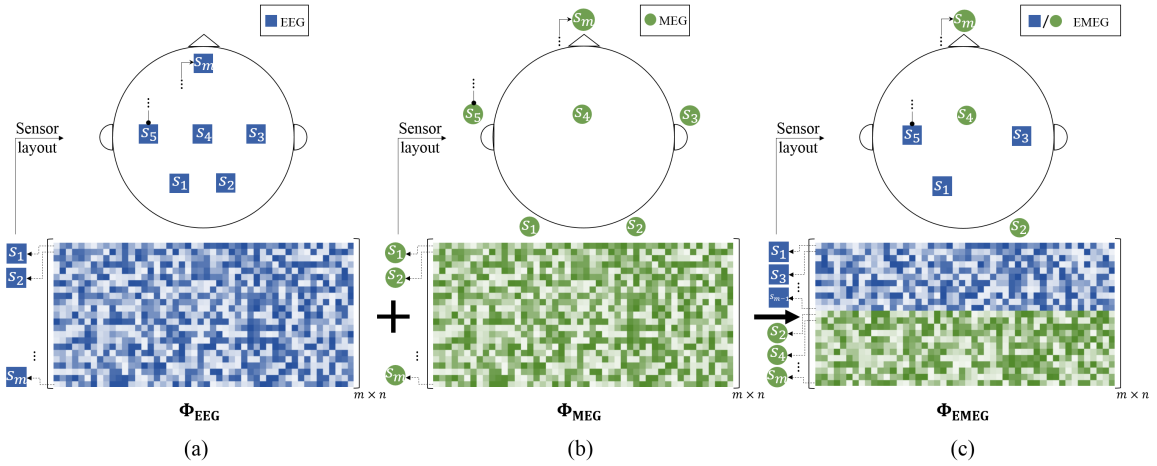



Figure 4.1: The rows of (a) EEG lead-field $\mathbf{\Phi_{EEG}}$ and (b) MEG lead-field $\mathbf{\Phi_{MEG}}$ can be combined together to form (c) the new multi-modal lead-field $\mathbf{\Phi_{EMEG}}$.

## 4.3 Results

In this section, the multi-modality strategy will be experimented on synthetic dictionaries and real EEG/MEG lead-fields. In all of the experiments, the agglomerative hierarchical clustering analysis based on the block structure identification framework proposed in Chapter 3 will be applied on the dictionaries.

The proposed inter-block coherence measure in a special case, Block-MCC$_{2,2}$[4], is used to compute the similarity matrix in the hierarchical clustering algorithm, i.e., $M_{2,2}(\mathbf{\Phi}) = \max_{k,k' \neq k} \|\mathbf{\Phi}^{\dagger}[k]\mathbf{\Phi}[k']\|_{2\to 2}/d_{max}$. The standard complete method is used to compute the inter-cluster distance required in the hierarchical clustering algorithm.

### 4.3.1 Synthetic dictionary

In this part, the set of two complementary random dictionaries $\mathbf{\Phi}_i \in \mathbb{R}^{10\times 80}$, $i=\{\mathbf{1},\mathbf{2}\}$, with equally-sized blocks of length $d_j=d=2$, $\forall j$, is generated from independent and identically distributed (i.i.d.) random variables and all columns of the dictionaries are normalized to have unit $\ell_2$ norm.

In order to generate two complementary random dictionaries, firstly, as described in Section 3.5.1, one of the dictionaries is generated by setting the related parameters $\varepsilon_{inter}$ and $\varepsilon_{intra}$. Then, utilising the same strategy and by setting the parameter $\varepsilon_{inter-dictionary}$, which controls the overlap between the two dictionaries, the second dictionary is produced. In figure 4.2(a) and (b), the role of different types of $\varepsilon$ in generation of random dictionaries $\mathbf{\Phi_1}$ and $\mathbf{\Phi_2}$ is shown. Then, as shown in figure 4.2(c) by drawing half of the rows of each of the dictionaries and finally by making a sufficient shift on one of the set of rows belonging to one dictionary, here $\mathbf{\Phi_2}$, the combined dictionary $\mathbf{\Phi_{12}}$ is obtained. All the results are shown in average and standard deviation over 100 repetitions of generation of complementary random dictionaries pairs.

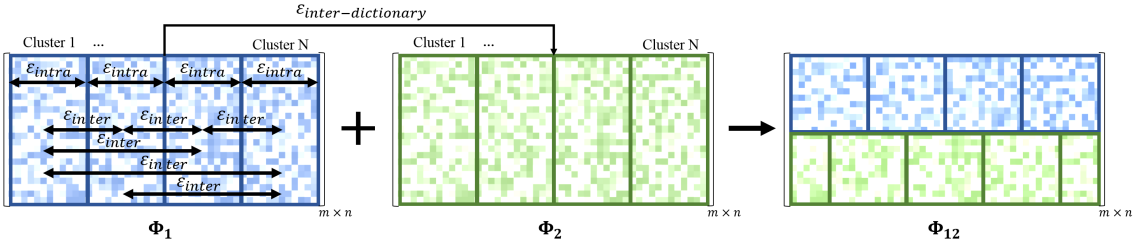



Figure 4.2: The role of (a) $\varepsilon_{inter}$ and $\varepsilon_{intra}$ in generation of random dictionary $\mathbf{\Phi_1}$, and (b) $\varepsilon_{inter-dictionary}$ in generation of $\mathbf{\Phi_2}$ to produce the multi-modal dictionary $\mathbf{\Phi_{12}}$.

[4] $(2,2)$-*Block Mutual Coherence Constant*

**Block-ERC in clustered representation for a combined dictionary** In Section 3.5.1, we demonstrated that in a random dictionary with certain values of $\varepsilon_{inter}$ and $\varepsilon_{intra}$, the Block-$SL_{2,2}$[5] computed in each clustering level of the hierarchical clustering algorithm have a peak value at a clustering level equal to the number of clusters in the generated dictionary.

Therefore, under certain values of $\varepsilon_{inter}$ and $\varepsilon_{intra}$, the argument of the peak in $Block{-}SL_{2,2}(\mathbf{\Phi})[\%]$ computed at each level of the hierarchical clustering algorithm is equal to the number of clusters in the dictionary, i.e., $N$.

In this part, in an experiment similar to Section 3.5.1, the block sparsity-level in the pessimistic case, i.e., $Block{-}SL_{2,2}(\mathbf{\Phi})[\%]$, is computed in each clustering level of the hierarchical clustering analysis, but here the dictionary is a combination of two complementary dictionaries instead of being a simple mono-modal random matrix. In Section 4.2.1, the procedure of combining two complementary dictionaries is described. Also, the complementarity of two dictionaries is defined based on their overlap, determined by $\varepsilon_{inter-dictionary}$.

In generating two complementary dictionaries, one of them is produced as explained in Section 3.5.1. Then the other dictionary is generated by multiplying the first dictionary to a random matrix $\exp((\boldsymbol{C}{-}\boldsymbol{C}^T)\sqrt{2}\varepsilon_{inter-dictionary}/\|\boldsymbol{C}{-}\boldsymbol{C}^T\|_F)$, where, $\boldsymbol{C}$ is a square random matrix of dimension $m$.

As it can be seen in figure 4.3, by applying hierarchical clustering on the blocks of the combined dictionary for $\varepsilon_{inter-dictionary}{=}1$, $\varepsilon_{inter}{\geq}3.5$ and $\varepsilon_{intra}{\leq}0.1$, a peak in $Block{-}SL_{2,2}(\mathbf{\Phi})[\%]$ appears in an argument equal to the sum of the clusters in each of the two original dictionaries.

For instance, when there are four clusters in each of the two complementary dictionaries i.e., $N{=}4$ (blue squares), which construct the combined one, the maximum $Block{-}SL_{2,2}(\mathbf{\Phi})[\%]$ for combined dictionary is at the eighth clustering level.

Similarly the consistent results are obtained for five clusters in each of the two complementary dictionaries, i.e., $N{=}5$ (red circles). In other words, there is a peak in $Block{-}SL_{2,2}(\mathbf{\Phi})[\%]$ at the tenth clustering level, when the multi-modal dictionary is composed of two mono-modal dictionaries, each has five clusters. For the clustering level lower than the optimal level, the space is under-sampled and there would be a block spanning the whole space, whereas for the clustering level higher than the optimal level, the space is over-sampled and the over-partitioning leads to high coherence measure.

On the other hand, when the average blocks of each cluster of blocks are not different enough, i.e., $\varepsilon_{inter}{<}3.5$, but still the blocks in each cluster of blocks are similar enough, i.e., $\varepsilon_{intra}{\leq}0.1$, we can observe that the number of clusters in the combined dictionary or the argument of the peak in $Block{-}SL_{2,2}(\mathbf{\Phi})[\%]$ is less than the sum of clusters in each of initial dictionaries but greater than the number of clusters in each of initial dictionaries.

[5] $(2,2)$-*Block-Sparsity Level*

Therefore, the number of clusters in the combined dictionary depending on the overlap between the blocks, which is determined by $\varepsilon$, is greater than the number of clusters in one of the initial dictionaries and at most is equal to the sum of the number of clusters in the initial dictionaries.

In addition, the amplitude of peak of $Block{-}SL_{2,2}(\mathbf{\Phi})[\%]$ increases when the number of clusters decreases. For instance, in figure 4.3, the amplitude of average peak of $Block{-}SL_{2,2}(\mathbf{\Phi})[\%]$ for multi-modal dictionaries composed of two four-clusters dictionaries (blue square) is higher than the corresponding average amplitude for five clusters in each of mono-modal dictionaries (red circle).

Therefore, by reducing the number of clusters in the dictionary, $Block{-}SL_{2,2}(\mathbf{\Phi})[\%]$ increases, hence, more weakened Block-ERC[6] are obtained.

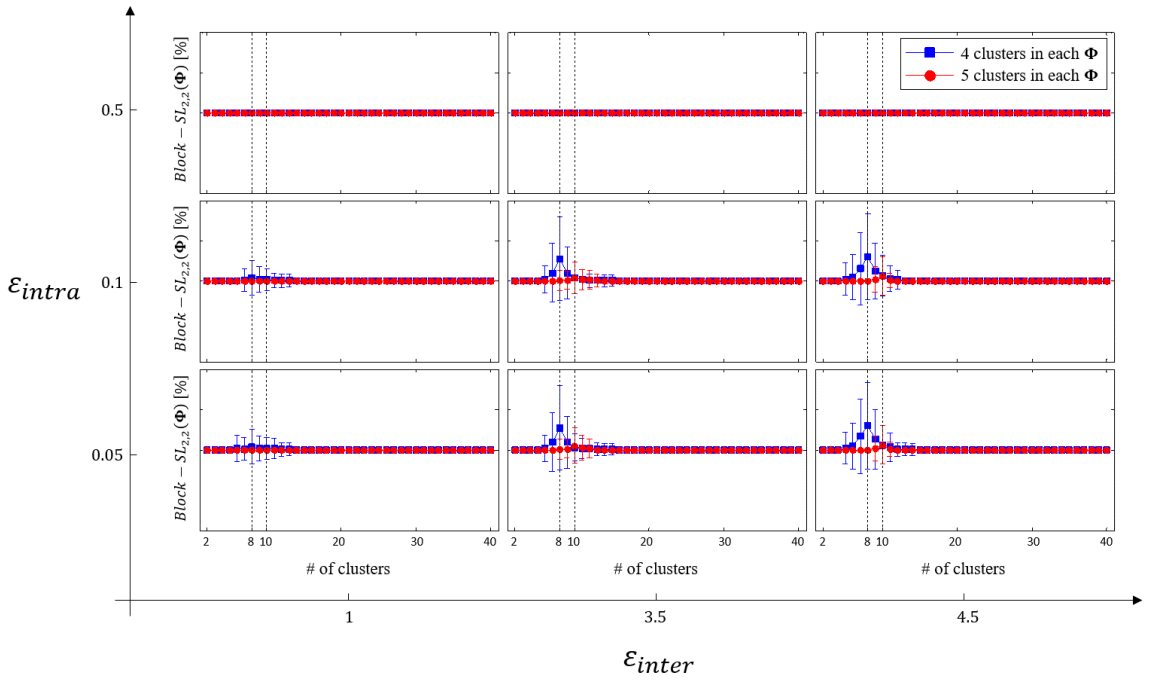



Figure 4.3: $Block{-}SL_{2,2}(\mathbf{\Phi})[\%]$ for each level of clustering is computed for the complete method, Block-MCC$_{2,2}$, $d$=2, $N$=$\{4,5\}$, $\varepsilon_{inter-dictionary}$=1, and different values of $\varepsilon_{inter}$ and $\varepsilon_{intra}$ for simulating dictionary $\mathbf{\Phi}\in\mathbb{R}^{10\times 80}$.

[6] *Block-sparse Exact Recovery Condition(s)*

### 4.3.2 Real EEG and MEG lead-field

In this part, the original lead-fields, which build the multi-modal lead-field are the ones explained in Section 3.5.2. Therefore, there are three head models with spherical, realistic inflated, and realistic highly-folded cortical sheets as shown in figure 4.4.

The 32 sensors shown in figure 4.4 are in multi-modal case, where, there are equal number of sensors from both modalities. By selecting the odd and even rows of EEG and MEG USLE, respectively, we can see that the sensor distribution is almost uniform, i.e., the sensors are not concentrated in a region of brain.

**EEG and MEG complementarity in brain source segmentation** In Section 1.2.6, some basic complementarities of EEG and MEG were discussed. In this section we show that in the brain source segmentation using the proposed strategy of clustering coherent lead-fields corresponding to brain sources utilising Block-MCC$_{q,p}$, we can observe some complementarities between brain source segmentation employing EEG or MEG lead-field.

In figure 4.5, the brain clusters or regions based on clustering of the EEG or MEG lead-field for clustering levels 3, 10, 15, 17, and 20 and three types of cortical sheets are shown.

It can be seen that for the all shown clustering levels, the previously mentioned complementarity also appears in our brain source segmentation strategy, which is more visible in low clustering levels in figure 4.5.

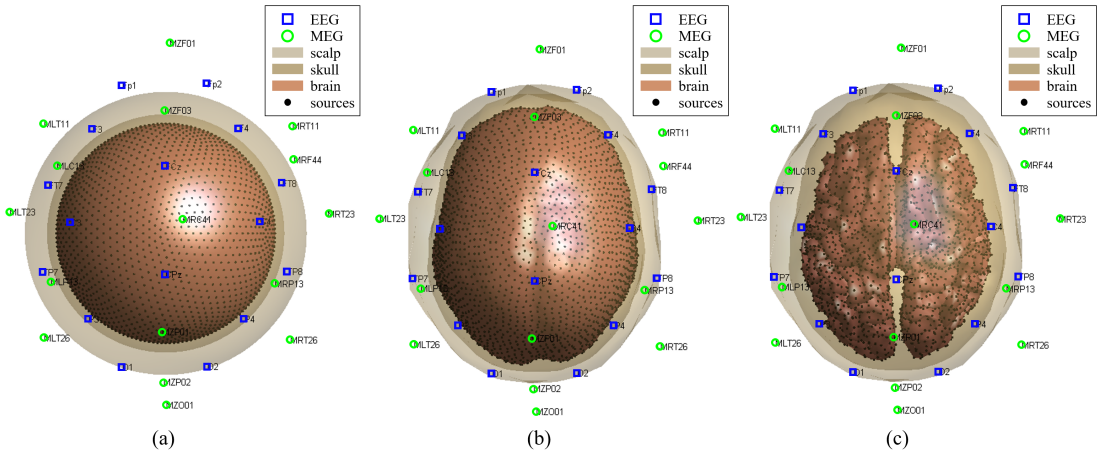



Figure 4.4: (a) spherical, (b) realistic inflated and (c) realistic highly-folded cortical sheets for the head models used in EEG and MEG multi-modality experiment.

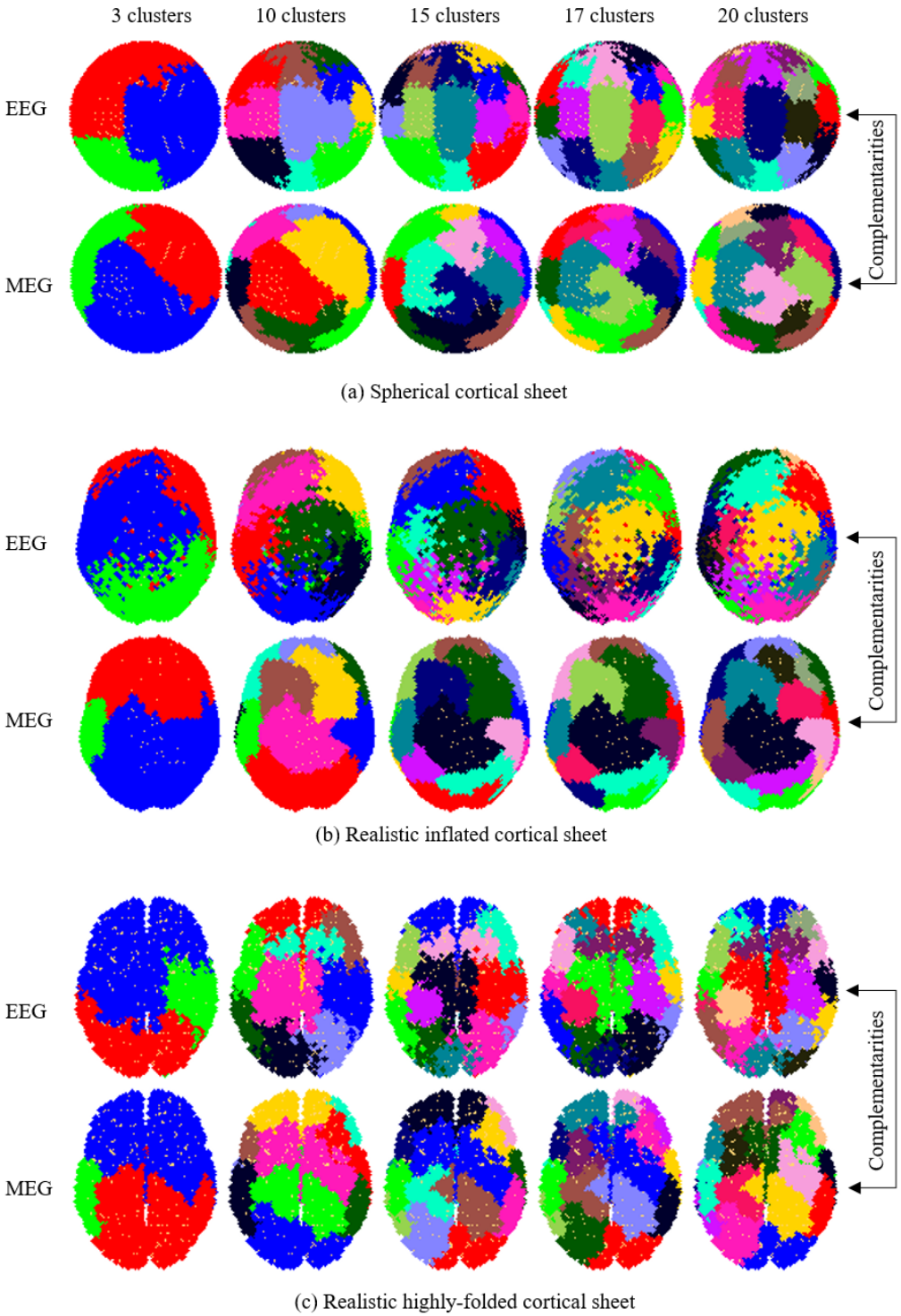



Figure 4.5: EEG and MEG complementarity in brain source segmentation using Block-MCC$_{2,2}$ and complete method for some clustering levels and for (a) spherical, (b) realistic inflated, and (c) realistic highly-folded cortical sheets.

**EEG and MEG multi-modality and the clustering tree** The agglomerative hierarchical cluster analysis, using the standard inter-cluster distance and our proposed inter-block similarity methods builds a hierarchy of clustering structures, which is represented in the clustering tree, as shown in figure 4.6.

The clustering trees built using EEG, MEG, and multi-modal EEG+MEG lead-fields for spherical, realistic inflated, and realistic highly-folded cortical sheets are shown in figure 4.6(a), (b), and (c), respectively.

As mentioned before, the advantage of using the hierarchical clustering algorithm is in estimating the number of clusters and in general the block structure, by selecting a block structure corresponding to the longest inter-node distance. Also, one can impose other constraints such as minimum number of clusters.

In figure 4.6, for each source model, the block structure is opted based on a minimum number of clusters and maximum inter-node distance. To make clusters distinguishable, each cluster is colour-coded.

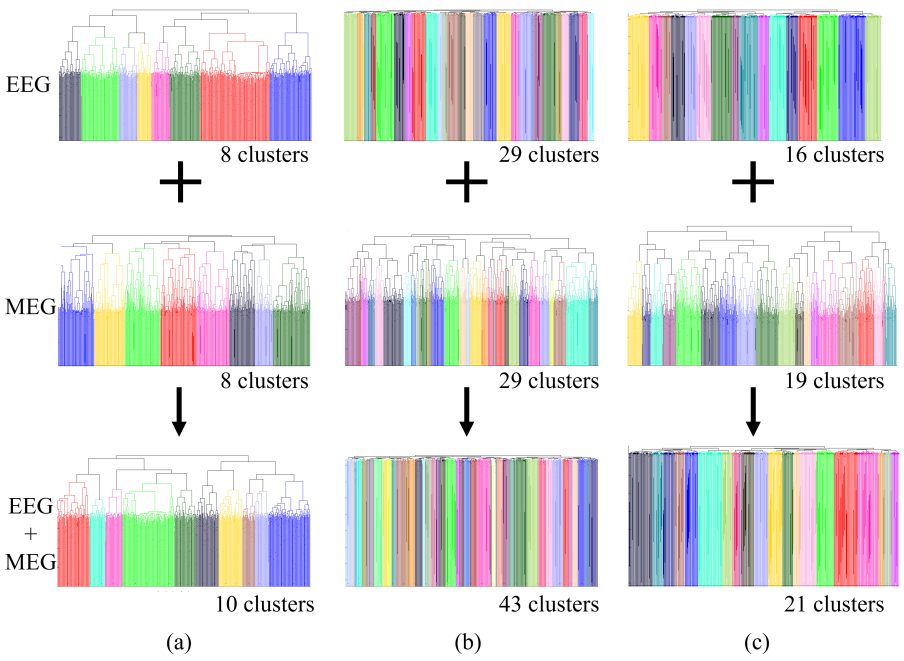



Figure 4.6: Clustering tree using EEG or MEG mono-modal lead-field and EEG+MEG multi-modal lead-field for (a) spherical, (b) realistic inflated, and (c) realistic highly-folded cortical sheets.

**EEG and MEG multi-modality and the number of regions** In order to estimate the number of clusters from the dendrogram as described in Section 3.5, in addition to the constraint of the maximum distance between consecutive nodes or clusters in dendrogram, the number of clusters is also set to be greater than a threshold.

As it can be seen in figure 4.7(a), the estimated number of clusters or brain regions resulted from the EEG and MEG multi-modal lead-field for spherical cortical sheet is 10, which is more than the estimated number of brain regions using EEG or MEG lead-field alone, which both are 8 clusters. The similar results are obtained for realistic head models as shown in figure 4.7(b) and (c).

Therefore, brain source space segmentation using the proposed strategy of clustering the coherent lead-fields based on the introduced Block-$MCC_{2,2}$ is more refined when utilising EEG and MEG multi-modal lead-field compared to EEG or MEG mono-modal lead-fields.

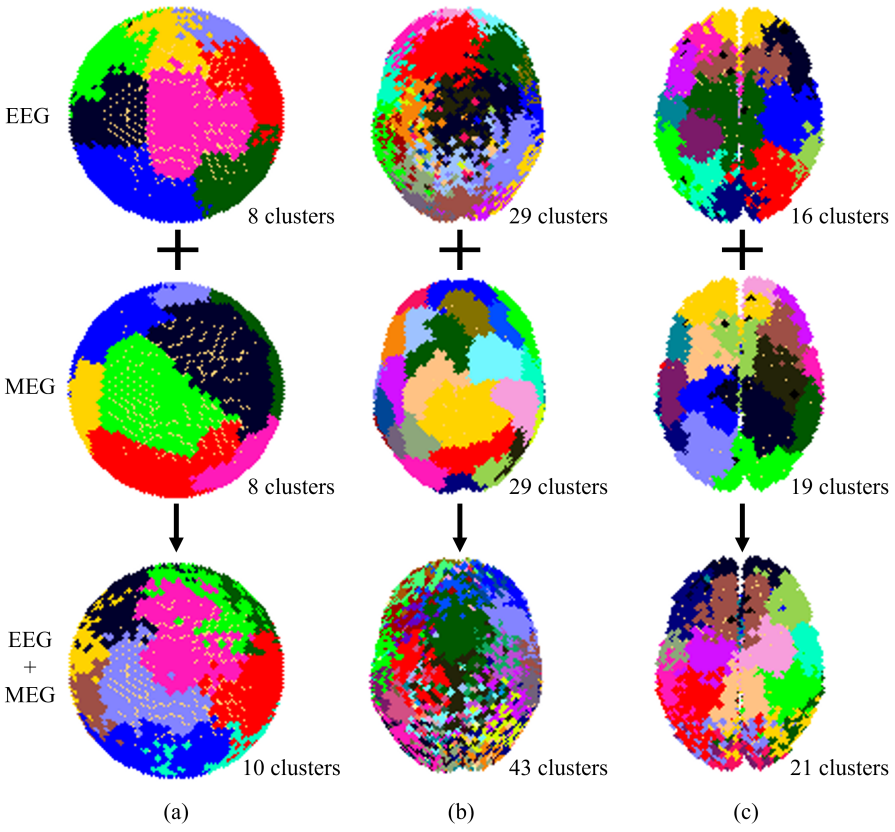



Figure 4.7: Number of clusters using EEG or MEG mono-modal lead-field and EEG+MEG multi-modal lead-field for (a) spherical, (b) realistic inflated, and (c) realistic highly-folded cortical sheets.

**EEG and MEG multi-modality and the area of regions** In this part, for a fixed number of regions or clusters, the area of brain regions is investigated for EEG/MEG mono-modal lead-fields and multi-modal EEG and MEG lead-field clustering.

In this experiment, to better demonstrate the impact of multi-modality on the area of brain regions, only the source model with spherical cortical sheet is used. In addition the standard *average* method is used to compute the inter-cluster distance. In this method, the distance between two clusters is defined as the average distance between each point in one cluster to every point in the other cluster. Furthermore, the Block-MCC$_{1,\infty}$ is used as coherence measure.

In the top row of figure 4.8, it can be seen that in each certain number of clusters, the biggest area (maximum value in bar chart) belongs to multi-modal EEG and MEG lead-field clustering and nearly for all the other regions the area for multi-modal EEG and MEG is less than EEG or MEG lead-field alone.

In addition, the area of brain regions are represented in pie charts. The largest area is coloured as blue in pie charts, which is not very obvious in brain source space in the current angle of view. Because, the region with maximum area represents brain sources at the bottom of the head, where, there is not any sensor on it, whereas the other regions represent brain sources under the sensors.

Therefore, for a fixed number of clusters, brain regions under sensors in multi-modal lead-field clustering are smaller than mono-modal lead-field clustering. Hence, EEG and MEG multi-modality leads to increased brain source space resolution.

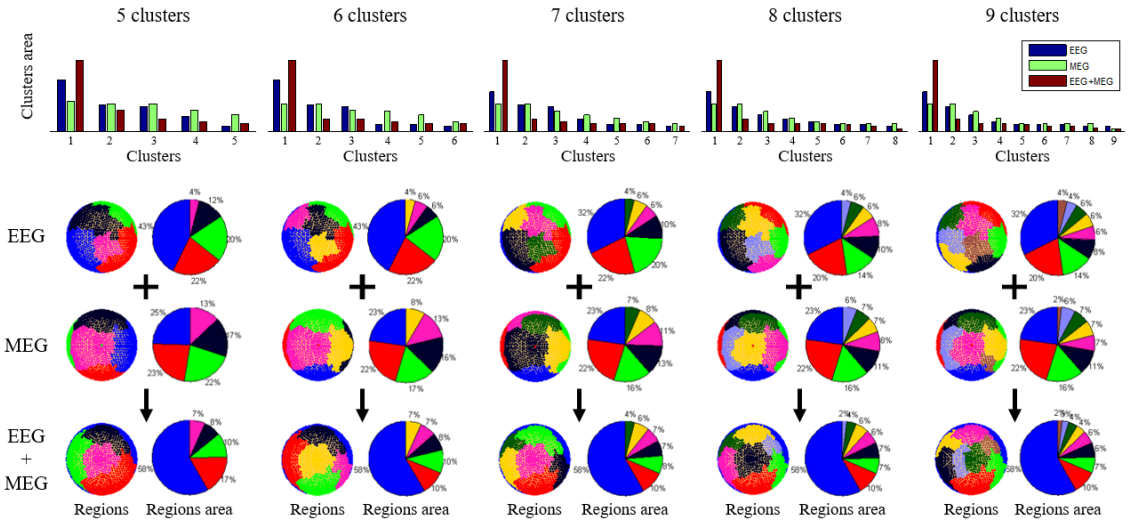



Figure 4.8: At each level of clustering, the area of brain regions are shown in pie and bar charts for EEG, MEG and their combination. At the top row, the descendingly sorted brain regions are shown for each one of three types of lead-fields. The hierarchical clustering method is average, and BMIC$_{1,\infty}$ is used as coherency measure.

## 4.4 Conclusion

In this chapter, first we showed the effectiveness of multi-modality on synthetic data through demonstrating that the number of clusters in a combined random dictionary is equal to the sum of the number of clusters in its two original random dictionaries.

Then, passing from synthetic data to real one, we showed that by clustering coherent brain sources through clustering their corresponding EEG or MEG lead-fields, some regions appear. The brain regions resulted from two mono-modal lead-fields are somehow complement each other.

Then, by applying the block structure identification framework introduced in Chapter 3, i.e., assuming that the number of clusters can be estimated by the largest distance between the adjacent nodes in dendrogram, we demonstrated that by combining EEG and MEG modalities, the number of clusters existing in the resulted multi-modal lead-field increases.

In addition, in another experiment we demonstrated that for a fixed number of clusters, the under-sensor brain regions are smaller in multi-modal lead-field compared to the mono-modal one. Therefore, more refined segmentation can be achieved in EEG and MEG multi-modality clustering using the proposed coherency measure.

In all the above experiments the number of sensors for mono-modal or multi-modal cases was kept to be equal.

Future research should focus on:

- Study on electromagnetic properties in 2D/3D mediums with different number of boundaries as a general case.
- Designing more scenarios (other than brain segmentation) to investigate the impact of multi-modality.
- Designing more methods to combine modalities.
- The impact of combining more modalities (greater than two).
- The optimum ratio of the number of measurements from each modality (not necessarily 50% in two modalities case).

# Contributions and perspectives

**Contents**



The contributions of this thesis can be categorized into three main research domains. The three classes of research contributions address the three challenges that have been mentioned and discussed in Section Introduction of the dissertation.

In order to explore and address the three mentioned challenges in Introduction, this dissertation is organized in three main parts. Since the first challenge uses the results of the second challenge, the second challenge is presented in the dissertation before the first challenge, and ultimately in the last part a partial answer to the third challenge is provided through experiments.

First challenge is about the ineffectiveness of some classic methods in high-dimensional problems. Second challenge raises when classic recovery conditions cannot be established for a new concept of constraint of corresponding optimisation problem, because of materials shortage in new framework. Finally, the combination of different information of a same phenomenon is the subject of the third challenge.

In this section, first we briefly recall the challenges and then mention the corresponding contributions to deal with them.

As mentioned in Section Introduction, all the three independent research orientations meet each other in our real-world application, which is distributed EEG/MEG source reconstruction problem, to partially respond to the main research question of the thesis, which is:

What is the added value of multi-modality, when solving inverse problems?

In the following the contributions and perspectives are organized into three sections:

## High-dimensional problem

**Challenge 1.** Many real-world inverse problems are vastly underdetermined, and classical sparse estimation techniques do no longer give acceptable recovery conditions. How can high-dimensional problems be adapted in favour of the coherence-based notion of conventional conditions?

For vastly underdetermined systems of linear equations, which the columns of the coefficient matrix are great in number, the classical dictionary characterisations such as mutual coherence constant are more likely to be high. Because the classical dictionary characterisations measure the coherency of the columns of the dictionary. Naturally, increasing in the number of columns of the dictionary would be equivalent to increasing the probability of coherency between the columns of the dictionary. On the other hands, due to the inverse relationship between the sparsity level defined in the exact recovery conditions and the classical dictionary characterisation, high coherency leads to low sparsity level. Therefore, being vastly underdetermined has a negative effect on the exact recovery conditions.

Assuming that the atomic entity in classical dictionary characterisations is *columns* of the dictionary, one possible solution would be to decrease the number of atomic entities by changing the concept of atomic entity from columns to *block of columns* of the dictionary. Therefore, in Chapter 3 we proposed to cluster the coherent entities (columns or block of columns) of the dictionary. By the idea of clustering the coherent entities to build new entities, we are improving the exact recovery conditions by gaining two advantages at the same time: (1) reducing the number of entities, (2) building more incoherent entities.

In the EEG/MEG source reconstruction problem, the initial entity is already a fixed-length block of columns (three columns), then we clustered the coherent blocks of columns. In order to determine the coherency between block of columns we utilised the dictionary characterisation block mutual coherence constant proposed in Chapter 2.

Then, for EEG/MEG source reconstruction problem we profited the idea of clustering the coherent blocks of columns, i.e., brain source lead-fields, to segment the brain source space. The proposed strategy for brain source space segmentation is in a general case, without being restricted to know any information about the sources activity and sensors measurement, only utilises EEG/MEG lead-field matrix. In addition, at the same time by clustering the coherent sources and forming the brain regions, the number of brain regions in which it is ensured to have a unique solution in the EEG or MEG source reconstruction problem is improved.

As perspective in this domain, different standard or customized clustering strategies can be investigated to find an optimum clustering strategy, which leads to improvement in the exact recovery conditions. One possible strategy in hierarchical clustering algorithm would be to update the similarity matrix at each clustering level, i.e., recompute the new distance of clusters and update the similarity matrix based on new clustering structure.

## Block-structured problem

**Challenge 2.** In classical assumption, the recovery conditions are generated from the columns of coefficient matrix, but clustered coefficient matrix consists of some differently-sized blocks, and not necessarily columns. Therefore, the initial assumption of classical recovery conditions does no longer hold true. How can appropriate recovery conditions be developed for block-structured problems?

In Chapter 2, we defined a general framework to cover all types of atomic entity including columns, and equally/differently-sized blocks of columns, in order to propose general theoretical exact recovery conditions. The proposed theoretical recovery conditions are based on *block-sparsity* constraint, which ensure the uniqueness of the block-sparse solution of corresponding weighted (pseudo-)mixed-norm optimisation problem in an underdetermined system of linear equations. The mentioned generality of the framework is in terms of the properties of the underdetermined system of linear equations, extracted characterisations, optimisation problems, and ultimately the recovery conditions. The mentioned theoretical exact recovery conditions are categorized in four different groups based on the utilised characterisations and properties, i.e., conditions based on (1) Block-Spark, (2) block null space property, (3) block mutual coherence constant, and (4) cumulative coherence constant.

On the other hand, the proposed framework is consistent with the base findings, since all the materials in the proposed infrastructure are a generalisation of the existing references Indeed, we investigated the theoretical relationship between the proposed infrastructure and the classical one, and showed that all the new materials reduce to the conventional ones in specific cases. We redemonstrated the benefit of block-sparsity assumption compared to conventional sparsity in the improvement of the recovery conditions. In addition, we theoretically proved the supremacy of the theoretical exact recovery conditions defined in the proposed general infrastructure over existing conditions, which are assuming the same block-sparsity constraint.

As perspective in this research direction, we could mention to the following subjects:

- Introducing block-sparse recovery conditions based on the proposed cumulative Block-MCC$_{q,p}$ defined in Definition 2.7.
- Generalising the conventional dictionary characterisation of $\mu$ defined in [DE03b], to establish block-sparse recovery conditions.
- Transforming all the previously mentioned block-sparse *exact* recovery conditions to block-sparse *stable* recovery conditions. In stable or robust recovery conditions we have, $\|\boldsymbol{y}-\boldsymbol{\Phi}\hat{\boldsymbol{\beta}}\|_2<e$, where, $e$ is a bounded noise.
- Study on block-sparse optimisation algorithms, and the relationship between the *theoretical* and *algorithmic* block-sparse recovery conditions.

## Multi-modality

**Challenge 3.** Is joining multiple modalities always beneficial, knowing that each modality provides us with different properties of the same phenomenon? How can the added value of multi-modality be demonstrated?

In order to approach the above-mentioned important and huge challenge, we simplify and limit the challenge to the conditions of our main problem, i.e., distributed EEG and MEG source reconstruction problem. In other words, we restrict the mentioned general challenge to have only two modalities, and try to partially address the challenge.

In Chapter 4, we propose a multi-modality framework based on the block structure identification framework proposed in Chapter 3 and block mutual coherence constant proposed in Chapter 2.

To this aim, we applied the block structure identification framework on multi-modal lead-field instead of mono-modal one, to segment brain source space. To investigate the impact of multi-modality, we defined a lead-field combining strategy, which reduces the impact of other factors such as change in the position and number of sensors.

First, we showed that brain regions resulted from clustering the coherent sources of EEG and MEG lead-field matrices separately, are complementary. Then, we made use of complementarities of EEG and MEG lead-field matrices to generated a combined EEG and MEG multi-modal lead-field matrix, and it turned out that in multi-modality case the number of clusters determined by the largest distance between adjacent nodes in dendrogram is higher than mono-modal cases. In addition, for a fixed number of clusters, the under-sensor brain regions in multi-modal lead-field is smaller than mono-modal lead-field clustering.

Therefore, it can be deduced that in multi-modality case, more refined and precise regions appear, hence, the resolution of identifying the active regions increases in comparison to the mono-modal cases.

As perspective in this research area, we could mention to the following points:

- Study on electromagnetic properties in 2D/3D mediums with different number of boundaries as a general case.
- Designing more scenarios (other than brain segmentation) to investigate the impact of multi-modality.
- Designing more methods to combine modalities.
- The impact of combining more modalities (greater than two).
- The optimum ratio of the number of measurements from each modality (not necessarily 50% in two modalities case).

Appendix A

# Proofs

## A.1 Proof of Property 2.2 (Bounds of division of two vector norms, page 54)

*Proof.* In order to demonstrate the bounds of $\|\boldsymbol{a}\|_p/\|\boldsymbol{a}\|_q$, first we prove it using the derivative method for $\forall(q,p)\in\mathbb{R}_{>1}$, next utilising the Hölder's inequality we show that the same bounds hold true for the wider range of $q$ and $p$, i.e., $\forall(q,p)\in\mathbb{R}_{>0}$.

1) To compute the critical point of $\|\boldsymbol{a}\|_p/\|\boldsymbol{a}\|_q$ using the derivative method, we need to compute its derivative with respect to the coordinates, knowing that $(f/g)'=(f'g-g'f)/g^2$, and $(|f|)'=f'\,f/|f|$, where $f'$ is derivative of $f$ with respect to $x$, i.e. $d\,f(x)/d\,x$, we have:

$$\forall(q,p)\in\mathbb{R}_{>1},\qquad \frac{\partial}{\partial a_i}\frac{\|\boldsymbol{a}\|_p}{\|\boldsymbol{a}\|_q}=\frac{\frac{\partial\|\boldsymbol{a}\|_p}{\partial a_i}\|\boldsymbol{a}\|_q-\frac{\partial\|\boldsymbol{a}\|_q}{\partial a_i}\|\boldsymbol{a}\|_p}{\|\boldsymbol{a}\|_q^2} \tag{A.1}$$

Then, we need to compute $\partial\|\boldsymbol{a}\|_p/\partial a_i$ and $\partial\|\boldsymbol{a}\|_q/\partial a_i$, so:

$$\frac{\partial\,\|\boldsymbol{a}\|_p}{\partial a_i}=\frac{\partial}{\partial a_i}\left(\sum_i|a_i|^p\right)^{\frac{1}{p}}=\frac{1}{p}\left(\sum_i|a_i|^p\right)^{\frac{1}{p}-1}\frac{\partial}{\partial a_i}\sum_i|a_i|^p=\|\boldsymbol{a}\|_p^{1-p}\,|a_i|^{p-2}\,a_i.$$

Therefore, substituting $\partial\|\boldsymbol{a}\|_p/\partial a_i$ and $\partial\|\boldsymbol{a}\|_q/\partial a_i$ in (A.1) and make it equal to zero, we have:

$$\begin{aligned}\frac{\partial}{\partial a_i}\frac{\|\boldsymbol{a}\|_p}{\|\boldsymbol{a}\|_q}&=\frac{\|\boldsymbol{a}\|_q\,\|\boldsymbol{a}\|_p^{1-p}\,|a_i|^{p-2}\,a_i-\|\boldsymbol{a}\|_p\,\|\boldsymbol{a}\|_q^{1-q}\,|a_i|^{q-2}\,a_i}{\|\boldsymbol{a}\|_q^2}\\&=a_i\frac{\|\boldsymbol{a}\|_p}{\|\boldsymbol{a}\|_q}\left(\frac{|a_i|^{p-2}}{\|\boldsymbol{a}\|_p^p}-\frac{|a_i|^{q-2}}{\|\boldsymbol{a}\|_q^q}\right)=0\qquad\Rightarrow|a_i|\in\left\{0,\left(\frac{\|\boldsymbol{a}\|_p^p}{\|\boldsymbol{a}\|_q^q}\right)^{\frac{1}{p-q}}\right\}.\end{aligned}$$

Hence, the derivative cancels for all $\boldsymbol{a}$ with $1\leq m\leq d$ non-zero elements that all are equal, i.e., $|a_i|=(\|\boldsymbol{a}\|_p^p/\|\boldsymbol{a}\|_q^q)^{1/(p-q)}=C\in\mathbb{R}_{>0}$, whereas all other elements are identically zero. Then, the fraction in the critical point evaluates to

$$\forall(q,p)\in\mathbb{R}_{>1},\qquad\frac{\|\boldsymbol{a}\|_p}{\|\boldsymbol{a}\|_q}=\frac{(mC^p)^{\frac{1}{p}}}{(mC^q)^{\frac{1}{q}}}=m^{\frac{1}{p}-\frac{1}{q}},$$

which is minimised for $m=\begin{cases} d^{1/p-1/q}, & \text{if } q\leq p \\ 1, & \text{if } q>p \end{cases} = \min\{1, d^{1/p-1/q}\}$, and maximised for $m=\begin{cases} 1, & \text{if } q\leq p \\ d^{1/p-1/q}, & \text{if } q>p \end{cases} = \max\{1, d^{1/p-1/q}\}$.

2) Knowing that the $\ell_p$ norm is a decreasing function in $p$, the following lower-bound is resulted for $p\leq q$:

$$\forall \boldsymbol{a} \in \mathbb{R}^d, \forall (q,p) \in \mathbb{R}^2_{>0}, \qquad \text{if } p \leq q \Rightarrow \|\boldsymbol{a}\|_q \leq \|\boldsymbol{a}\|_p \Rightarrow 1 \leq \frac{\|\boldsymbol{a}\|_p}{\|\boldsymbol{a}\|_q}.$$

Now to demonstrate the upper-bound, the following Hölder's inequality is used [GL13]:

$$\forall r \in \mathbb{R}_{\geq 1}, \qquad \sum_{i=1}^{d} |x_i y_i| \leq \|\boldsymbol{x}\|_r \, \|\boldsymbol{y}\|_{\frac{r}{r-1}} = \left(\sum_{i=1}^{d} |x_i|^r\right)^{\frac{1}{r}} \left(\sum_{i=1}^{d} |y_i|^{\frac{r}{r-1}}\right)^{\frac{r-1}{r}}.$$

Next, assuming $x_i = {a_i}^p$, $y_i = 1$, and $r = q/p$ (the above condition on $r$ is met, because $p\leq q$), we have $\sum_{i=1}^{d} |x_i y_i| = \sum_{i=1}^{d} |{a_i}^p| = \sum_{i=1}^{d} |a_i|^p$, and the above equation turns into:

$$\begin{aligned} \forall \boldsymbol{a} \in \mathbb{R}^d, \forall (q,p) \in \mathbb{R}^2_{>0}, p \leq q, \qquad \sum_{i=1}^{d} |a_i|^p &\leq \left(\sum_{i=1}^{d} (|a_i|^p)^{\frac{q}{p}}\right)^{\frac{p}{q}} \left(\sum_{i=1}^{d} 1^{\frac{q}{q-p}}\right)^{\frac{q-p}{q}} \\ &= \left(\sum_{i=1}^{d} |a_i|^q\right)^{\frac{p}{q}} d^{\frac{q-p}{q}}. \end{aligned}$$

Then, taking the both sides to the power of $1/p$, we get:

$$\begin{aligned} \forall \boldsymbol{a} \in \mathbb{R}^d, \forall (q,p) \in \mathbb{R}^2_{>0}, p \leq q, \qquad \left(\sum_{i=1}^{d} |a_i|^p\right)^{\frac{1}{p}} &\leq \left(\sum_{i=1}^{d} |a_i|^q\right)^{\frac{1}{q}} d^{\frac{1}{p}-\frac{1}{q}} \\ \Rightarrow \frac{\|\boldsymbol{a}\|_p}{\|\boldsymbol{a}\|_q} &\leq d^{\frac{1}{p}-\frac{1}{q}}. \end{aligned}$$

Therefore, for $p\leq q$, we have $1 \leq \|\boldsymbol{a}\|_p / \|\boldsymbol{a}\|_q \leq d^{1/p-1/q}$. Similarly, for $q\leq p$, we get $d^{1/p-1/q} \leq \|\boldsymbol{a}\|_p / \|\boldsymbol{a}\|_q \leq 1$, which proves the property. □

## A.2 Proof of Property 2.4 ($\ell_{q\to p}$ operator-norm properties, page 56)

*Proof.* Based on the definition of the $\|\boldsymbol{A}\|_{q\to p}$, i.e., $\max_{\boldsymbol{a}\neq\boldsymbol{0}} \|\boldsymbol{A}\boldsymbol{a}\|_p/\|\boldsymbol{a}\|_q$ or $\max_{\|\boldsymbol{a}\|_q\leq 1} \|\boldsymbol{A}\boldsymbol{a}\|_p$, we prove the following properties, for $\boldsymbol{A}\in\mathbb{R}^{m\times n}$, $\boldsymbol{B}\in\mathbb{R}^{m\times n}$, and $\boldsymbol{C}\in\mathbb{R}^{n\times l}$:

- Nonnegativity: In definition $\|\boldsymbol{A}\|_{q\to p}=\max_{\|\boldsymbol{a}\|_q\leq 1} \|\boldsymbol{A}\boldsymbol{a}\|_p$, $\|\boldsymbol{A}\boldsymbol{a}\|_p$ is greater than or equal to zero, then the nonnegativity of $\|\boldsymbol{A}\|_{q\to p}$ is obvious $\forall(q,p)\in\mathbb{R}^2_{\geq 0}$.
- Positivity: $\forall(q,p)\in\mathbb{R}^2_{\geq 0}$ if $\|\boldsymbol{A}\|_{q\to p}=0$, then $\|\boldsymbol{A}\boldsymbol{a}\|_p=0$ for each $\boldsymbol{a}\in\mathbb{R}^n$, i.e., each column in $\boldsymbol{A}$ is zero. Hence, $\boldsymbol{A}=\boldsymbol{0}$.
- Homogeneity: $\forall q\geq 0$ and $\forall p>0$, we have

$$\begin{aligned}\|\alpha\boldsymbol{A}\|_{q\to p} = \max_{\boldsymbol{a}\neq\boldsymbol{0}} \frac{\|\alpha\boldsymbol{A}\boldsymbol{a}\|_p}{\|\boldsymbol{a}\|_q} &= \max_{\boldsymbol{a}\neq\boldsymbol{0}} |\alpha| \frac{\|\boldsymbol{A}\boldsymbol{a}\|_p}{\|\boldsymbol{a}\|_q} \\ &= |\alpha| \, \|\boldsymbol{A}\|_{q\to p}\,.\end{aligned}$$

  The above second equality in the first line is resulted from the homogeneity property of norm of vectors, which holds true only for $p>0$ [Ela10]; [GL13].

- Triangle inequality: $\forall q\geq 0$, $p=0$ or $\forall p\geq 1$, it is obtained as follows

$$\begin{aligned}\|\boldsymbol{A}+\boldsymbol{B}\|_{q\to p} = \max_{\boldsymbol{a}\neq\boldsymbol{0}} \frac{\|(\boldsymbol{A}+\boldsymbol{B})\,\boldsymbol{a}\|_p}{\|\boldsymbol{a}\|_q} &\leq \max_{\boldsymbol{a}\neq\boldsymbol{0}} \frac{\|\boldsymbol{A}\boldsymbol{a}\|_p + \|\boldsymbol{B}\boldsymbol{a}\|_p}{\|\boldsymbol{a}\|_q} \\ &\leq \max_{\boldsymbol{a}\neq\boldsymbol{0}} \frac{\|\boldsymbol{A}\boldsymbol{a}\|_p}{\|\boldsymbol{a}\|_q} + \max_{\boldsymbol{a}\neq\boldsymbol{0}} \frac{\|\boldsymbol{B}\boldsymbol{a}\|_p}{\|\boldsymbol{a}\|_q} \\ &= \|\boldsymbol{A}\|_{q\to p} + \|\boldsymbol{B}\|_{q\to p}\,.\end{aligned}$$

  The above inequality in the first line is resulted from the triangle inequality property of norm of vectors, which holds true only for $p=0$ and $p\geq 1$ [Ela10]; [GL13].

- Submultiplicativity: It is obtained as follows:

$$\begin{aligned}\|\boldsymbol{A}\boldsymbol{C}\|_{q\to p} = \max_{\boldsymbol{a}\neq\boldsymbol{0}} \frac{\|\boldsymbol{A}\boldsymbol{C}\boldsymbol{a}\|_p}{\|\boldsymbol{a}\|_q} &= \max_{\boldsymbol{a}\neq\boldsymbol{0}} \frac{\|\boldsymbol{A}\boldsymbol{C}\boldsymbol{a}\|_p}{\|\boldsymbol{a}\|_q} \frac{\|\boldsymbol{C}\boldsymbol{a}\|_q}{\|\boldsymbol{C}\boldsymbol{a}\|_q} \frac{\|\boldsymbol{C}\boldsymbol{a}\|_p}{\|\boldsymbol{C}\boldsymbol{a}\|_p} \\ &\leq \max_{\boldsymbol{C}\boldsymbol{a}\neq\boldsymbol{0}} \frac{\|\boldsymbol{A}\boldsymbol{C}\boldsymbol{a}\|_p}{\|\boldsymbol{C}\boldsymbol{a}\|_q} \max_{\boldsymbol{a}\neq\boldsymbol{0}} \frac{\|\boldsymbol{C}\boldsymbol{a}\|_p}{\|\boldsymbol{a}\|_q} \max_{\boldsymbol{a}\neq\boldsymbol{0}} \frac{\|\boldsymbol{C}\boldsymbol{a}\|_q}{\|\boldsymbol{C}\boldsymbol{a}\|_p} \\ &= \|\boldsymbol{A}\|_{q\to p} \, \|\boldsymbol{C}\|_{q\to p} \max\left\{1, n^{\frac{1}{q}-\frac{1}{p}}\right\}.\end{aligned} \tag{A.2}$$

  The last term of (A.2) is obtained using the upper-bound in Property 2.2 (bounds of two vector norms division), i.e.,:

$$\forall\,(q,p)\in\mathbb{R}^2_{>0}, \forall\boldsymbol{a}\in\mathbb{R}^d, \qquad \min\left\{1, d^{\frac{1}{q}-\frac{1}{p}}\right\} \leq \frac{\|\boldsymbol{a}\|_q}{\|\boldsymbol{a}\|_p} \leq \max\left\{1, d^{\frac{1}{q}-\frac{1}{p}}\right\}.$$

  Therefore, from (A.2) for $q\geq p>0$, we see that the submultiplicativity holds true, i.e., $\|\boldsymbol{A}\boldsymbol{C}\|_{q\to p} \leq \|\boldsymbol{A}\|_{q\to p} \, \|\boldsymbol{C}\|_{q\to p}$.

- Bounds: 1) Considering the definition of $\ell_{q\to p}$ operator-norm, i.e., $\max_{\boldsymbol{a}\neq\boldsymbol{0}} \|\boldsymbol{A}\boldsymbol{a}\|_p/\|\boldsymbol{a}\|_q$, and the lower-bound of division of two vector norms introduced in Property 2.2 (bounds of two vector norms division), i.e., $\forall(q,p)\in\mathbb{R}^2_{>0}, \forall\boldsymbol{a}\in\mathbb{R}^d : \|\boldsymbol{a}\|_q/\|\boldsymbol{a}\|_p \geq \min\left\{1, d^{1/q-1/p}\right\}$, we conclude the following lower-bounds for $\boldsymbol{A}\in\mathbb{R}^{m\times n}$, $\boldsymbol{a}\in\mathbb{R}^n$, and $\forall\,(q,p,q',p')\in\mathbb{R}^4_{>0}$:

$$\begin{aligned}
&\forall \boldsymbol{a}\neq\boldsymbol{0}, \qquad \frac{\|\boldsymbol{A}\boldsymbol{a}\|_p}{\|\boldsymbol{a}\|_q} \geq \min\left\{1, m^{\frac{1}{p}-\frac{1}{p'}}\right\}\frac{\|\boldsymbol{A}\boldsymbol{a}\|_{p'}}{\|\boldsymbol{a}\|_q},\\
&\forall \boldsymbol{a}\neq\boldsymbol{0}, \qquad \frac{\|\boldsymbol{A}\boldsymbol{a}\|_p}{\|\boldsymbol{a}\|_q} \geq \min\left\{1, n^{\frac{1}{q'}-\frac{1}{q}}\right\}\frac{\|\boldsymbol{A}\boldsymbol{a}\|_p}{\|\boldsymbol{a}\|_{q'}}.
\end{aligned}$$

Therefore, we have

$$\forall \boldsymbol{a}\neq\boldsymbol{0}, \qquad \frac{\|\boldsymbol{A}\boldsymbol{a}\|_p}{\|\boldsymbol{a}\|_q} \geq \max\left\{\min\left\{1, m^{\frac{1}{p}-\frac{1}{p'}}\right\}\frac{\|\boldsymbol{A}\boldsymbol{a}\|_{p'}}{\|\boldsymbol{a}\|_q}, \min\left\{1, n^{\frac{1}{q'}-\frac{1}{q}}\right\}\frac{\|\boldsymbol{A}\boldsymbol{a}\|_p}{\|\boldsymbol{a}\|_{q'}}\right\}.$$

Then, taking into account that the max operator is order preserving, i.e., $\forall x$, if $f(x)\leq g(x)$, then $\max_x f(x)\leq \max_x g(x)$, the lower-bound of $\|\boldsymbol{A}\|_{q\to p}$ is achieved.

Similarly, utilising the upper-bound of division of two vector norms introduced in Property 2.2 (bounds of two vector norms division), i.e., $\forall(q,p)\in\mathbb{R}^2_{>0}, \forall\boldsymbol{a}\in\mathbb{R}^d$ : $\|\boldsymbol{a}\|_q/\|\boldsymbol{a}\|_p \leq \max\{1, d^{1/q-1/p}\}$, results in the following inequalities, which conclude the upper-bound of $\|\boldsymbol{A}\|_{q\to p}$ for $\boldsymbol{A}\in\mathbb{R}^{m\times n}$, $\boldsymbol{a}\in\mathbb{R}^n$, and $\forall\,(q,p,q',p')\in\mathbb{R}^4_{>0}$:

$$\begin{aligned}
&\forall \boldsymbol{a}\neq\boldsymbol{0}, \qquad \frac{\|\boldsymbol{A}\boldsymbol{a}\|_p}{\|\boldsymbol{a}\|_q} \leq \max\left\{1, m^{\frac{1}{p}-\frac{1}{p'}}\right\}\frac{\|\boldsymbol{A}\boldsymbol{a}\|_{p'}}{\|\boldsymbol{a}\|_q},\\
&\forall \boldsymbol{a}\neq\boldsymbol{0}, \qquad \frac{\|\boldsymbol{A}\boldsymbol{a}\|_p}{\|\boldsymbol{a}\|_q} \leq \max\left\{1, n^{\frac{1}{q'}-\frac{1}{q}}\right\}\frac{\|\boldsymbol{A}\boldsymbol{a}\|_p}{\|\boldsymbol{a}\|_{q'}}.
\end{aligned}$$

Therefore,

$$\forall \boldsymbol{a}\neq\boldsymbol{0}, \qquad \frac{\|\boldsymbol{A}\boldsymbol{a}\|_p}{\|\boldsymbol{a}\|_q} \leq \min\left\{\max\left\{1, m^{\frac{1}{p}-\frac{1}{p'}}\right\}\frac{\|\boldsymbol{A}\boldsymbol{a}\|_{p'}}{\|\boldsymbol{a}\|_q}, \max\left\{1, n^{\frac{1}{q'}-\frac{1}{q}}\right\}\frac{\|\boldsymbol{A}\boldsymbol{a}\|_p}{\|\boldsymbol{a}\|_{q'}}\right\}.$$

2) To prove the second set of bounds, from Property 2.2 (bounds of two vector norms division), i.e., $\forall(q,p)\in\mathbb{R}^2_{>0}, \forall\boldsymbol{a}\in\mathbb{R}^d : \|\boldsymbol{a}\|_q/\|\boldsymbol{a}\|_p \geq \min\{1, d^{1/q-1/p}\}$, we have the following inequalities for $\boldsymbol{A}\in\mathbb{R}^{m\times n}$, $\boldsymbol{a}\in\mathbb{R}^n$, and $\forall\,(q,p,q',p')\in\mathbb{R}^4_{>0}$:

$$\begin{aligned}
&\forall \boldsymbol{a}\neq\boldsymbol{0}, \qquad \|\boldsymbol{A}\boldsymbol{a}\|_p \quad \geq \min\left\{1, m^{\frac{1}{p}-\frac{1}{p'}}\right\}\|\boldsymbol{A}\boldsymbol{a}\|_{p'},\\
&\forall \boldsymbol{a}\neq\boldsymbol{0}, \qquad \frac{1}{\|\boldsymbol{a}\|_q} \quad \geq \min\left\{1, n^{\frac{1}{q'}-\frac{1}{q}}\right\}\frac{1}{\|\boldsymbol{a}\|_{q'}}.
\end{aligned}$$

Considering that both sides of the above inequalities are positive, by multiplying the same sides we get:

$$\forall \boldsymbol{a}\neq\boldsymbol{0}, \qquad \frac{\|\boldsymbol{A}\boldsymbol{a}\|_p}{\|\boldsymbol{a}\|_q} \geq \min\left\{1, m^{\frac{1}{p}-\frac{1}{p'}}\right\}\min\left\{1, n^{\frac{1}{q'}-\frac{1}{q}}\right\}\frac{\|\boldsymbol{A}\boldsymbol{a}\|_{p'}}{\|\boldsymbol{a}\|_{q'}}.$$

Then taking into account that the max function is order preserving, the proof is done for lower-bound.

Similarly, for the upper-bound, from Property 2.2 (bounds of two vector norms division), i.e., $\forall(q,p)\in\mathbb{R}^2_{>0}, \forall \boldsymbol{a}\in\mathbb{R}^d : \|\boldsymbol{a}\|_q/\|\boldsymbol{a}\|_p \leq \max\{1, d^{1/q-1/p}\}$, we have the following inequalities for $\boldsymbol{A}\in\mathbb{R}^{m\times n}$, $\boldsymbol{a}\in\mathbb{R}^n$, and $\forall\,(q,p,q',p')\in\mathbb{R}^4_{>0}$:

$$\begin{aligned} \forall \boldsymbol{a}\neq\boldsymbol{0}, \qquad \|\boldsymbol{A}\boldsymbol{a}\|_p &\leq \max\left\{1, m^{\frac{1}{p}-\frac{1}{p'}}\right\}\|\boldsymbol{A}\boldsymbol{a}\|_{p'}\,, \\ \forall \boldsymbol{a}\neq\boldsymbol{0}, \qquad \frac{1}{\|\boldsymbol{a}\|_q} &\leq \max\left\{1, n^{\frac{1}{q'}-\frac{1}{q}}\right\}\frac{1}{\|\boldsymbol{a}\|_{q'}}. \end{aligned}$$

Considering that both sides of the above inequalities are positive, by multiplying the same sides we get:

$$\forall \boldsymbol{a}\neq\boldsymbol{0}, \qquad \frac{\|\boldsymbol{A}\boldsymbol{a}\|_p}{\|\boldsymbol{a}\|_q} \leq \max\left\{1, m^{\frac{1}{p}-\frac{1}{p'}}\right\}\max\left\{1, n^{\frac{1}{q'}-\frac{1}{q}}\right\}\frac{\|\boldsymbol{A}\boldsymbol{a}\|_{p'}}{\|\boldsymbol{a}\|_{q'}}.$$

Then taking into account that the max function is order preserving, the proof is done for upper-bound.

In order to prove the $\ell_{q\to p}$ operator-norm inequalities shown schematically in figure 2.3, utilising the lower-bound of second set of bounds in Property 2.4 ($\ell_{q\to p}$ operator-norm properties), i.e., $\forall\,(q,p,q',p')\in\mathbb{R}^4_{>0}, \forall\boldsymbol{A}\in\mathbb{R}^{m\times n}$ : $\|\boldsymbol{A}\|_{q\to p}\geq \min\{1, m^{1/p-1/p'}\}\min\{1, n^{1/q'-1/q}\}\|\boldsymbol{A}\|_{q'\to p'}$, it is straightforward that for a fixed $p{=}p'$ if $q'{\leq}q$ (or similarly, for a fixed $q{=}q'$ if $p{\leq}p'$) we have $\|\boldsymbol{A}\|_{q\to p}\geq\|\boldsymbol{A}\|_{q'\to p'}$.

3) In order to prove the lower-bound of the third set of bounds, using the lower-bound of second set of bounds in Property 2.4 ($\ell_{q\to p}$ operator-norm properties), i.e., $\forall\,(q,p,q',p')\in\mathbb{R}^4_{>0}, \forall\boldsymbol{A}\in\mathbb{R}^{m\times n} : \|\boldsymbol{A}\|_{q\to p}\geq \min\{1, m^{1/p-1/p'}\}\min\{1, n^{1/q'-1/q}\}\|\boldsymbol{A}\|_{q'\to p'}$, we will have the following first line of inequalities. Then, by utilising the same inequality once again, this time for $q'{=}p'{=}2$, we get the following second line of inequalities. The reason that first the relation with $\|\boldsymbol{A}\|_{q'\to p'}$ is stated and then the relation with $\|\boldsymbol{A}\|_{2\to 2}$, is that if in a problem the $q$ and $p$ values are bounded based on some thresholds such as $q'$ and $p'$, respectively, e.g., $q{\leq/\geq}q'$ and $p{\leq/\geq}p'$, then there could be a unique solution to the related min and max operators. Whereas, if directly the relation with $\|\boldsymbol{A}\|_{2\to 2}$ was stated, and supposing that $p{\geq}1$, then form instance $\min\{1, m^{1/p-1/2}\}$ would have two values based on the value of $p$ compared to 1. Then from [GL13], we have $\|\boldsymbol{A}\|_{2\to 2}\geq\|\boldsymbol{A}\|_F/\sqrt{Rank(\boldsymbol{A})}$, where the Frobenius norm is defined as $(\sum_{i=1}^m\sum_{j=1}^n|a_{i,j}|^2)^{1/2}$, which produces the following third line. But $Rank(\boldsymbol{A})$ is upper-bounded by $\min\{m,n\}$ [GL13], then $1/\sqrt{Rank(\boldsymbol{A})}$ is lower-bounded

by $1/\sqrt{\min\{m,n\}}$, which produces the following last line:

$$\begin{aligned}
\|\boldsymbol{A}\|_{q\to p} &\geq \min\left\{1, m^{\frac{1}{p}-\frac{1}{p'}}\right\} \min\left\{1, n^{\frac{1}{q'}-\frac{1}{q}}\right\} \|\boldsymbol{A}\|_{q'\to p'} \\
&\geq \min\left\{1, m^{\frac{1}{p}-\frac{1}{p'}}\right\} \min\left\{1, n^{\frac{1}{q'}-\frac{1}{q}}\right\} \min\left\{1, m^{\frac{1}{p'}-\frac{1}{2}}\right\} \min\left\{1, n^{\frac{1}{2}-\frac{1}{q'}}\right\} \|\boldsymbol{A}\|_{2\to 2} \\
&\geq \frac{\min\left\{1, m^{\frac{1}{p}-\frac{1}{p'}}\right\} \min\left\{1, n^{\frac{1}{q'}-\frac{1}{q}}\right\} \min\left\{1, m^{\frac{1}{p'}-\frac{1}{2}}\right\} \min\left\{1, n^{\frac{1}{2}-\frac{1}{q'}}\right\}}{\sqrt{Rank(\boldsymbol{A})}} \|\boldsymbol{A}\|_F \\
&\geq \frac{\min\left\{1, m^{\frac{1}{p}-\frac{1}{p'}}\right\} \min\left\{1, n^{\frac{1}{q'}-\frac{1}{q}}\right\} \min\left\{1, m^{\frac{1}{p'}-\frac{1}{2}}\right\} \min\left\{1, n^{\frac{1}{2}-\frac{1}{q'}}\right\}}{\sqrt{\min\{m,n\}}} \|\boldsymbol{A}\|_F .
\end{aligned}$$

In order to prove the upper-bound of the last set of bounds, using the upper-bound of second set of bounds in Property 2.4 ($\ell_{q\to p}$ operator-norm properties), i.e., $\forall\,(q,p,q',p')\in\mathbb{R}^4_{>0}, \forall \boldsymbol{A}\in\mathbb{R}^{m\times n}$ : $\|\boldsymbol{A}\|_{q\to p}\leq \max\{1, m^{1/p-1/p'}\}\max\{1, n^{1/q'-1/q}\}\|\boldsymbol{A}\|_{q'\to p'}$, we will have the following first line of inequalities. Then, by utilising the same inequality once again, this time for $q'=p'=2$, we get the following second line of inequalities. But from [GL13], we have $\|\boldsymbol{A}\|_{2\to 2}\leq\|\boldsymbol{A}\|_F$, which produces the following last line:

$$\begin{aligned}
\|\boldsymbol{A}\|_{q\to p} &\leq \max\left\{1, m^{\frac{1}{p}-\frac{1}{p'}}\right\} \max\left\{1, n^{\frac{1}{q'}-\frac{1}{q}}\right\} \|\boldsymbol{A}\|_{q'\to p'} \\
&\leq \max\left\{1, m^{\frac{1}{p}-\frac{1}{p'}}\right\} \max\left\{1, n^{\frac{1}{q'}-\frac{1}{q}}\right\} \max\left\{1, m^{\frac{1}{p'}-\frac{1}{2}}\right\} \max\left\{1, n^{\frac{1}{2}-\frac{1}{q'}}\right\} \|\boldsymbol{A}\|_{2\to 2} \\
&\leq \max\left\{1, m^{\frac{1}{p}-\frac{1}{p'}}\right\} \max\left\{1, n^{\frac{1}{q'}-\frac{1}{q}}\right\} \max\left\{1, m^{\frac{1}{p'}-\frac{1}{2}}\right\} \max\left\{1, n^{\frac{1}{2}-\frac{1}{q'}}\right\} \|\boldsymbol{A}\|_F .
\end{aligned}$$

4) In order to prove the relation between the $\ell_{q\to p}$ operator-norm of two matrices $\boldsymbol{A}\in\mathbb{R}^{m\times n}$ and $\boldsymbol{B}\in\mathbb{R}^{m\times n}$, having the condition $\forall i,j,\ |a_{i,j}|\leq b_{i,j}=\max_{i,j}|a_{i,j}|$, or even when the on-diagonal entries are set to zero, we can utilise the second bounds proposed in the current property. First, using the lower-bound of the mentioned set of bounds, i.e., $\forall\,(q,p,q',p')\in\mathbb{R}^4_{>0}, \forall \boldsymbol{A}\in\mathbb{R}^{m\times n}$ : $\|\boldsymbol{A}\|_{q\to p}\geq \min\{1, m^{1/p-1/p'}\}\min\{1, n^{1/q'-1/q}\}\|\boldsymbol{A}\|_{q'\to p'}$, for $q'=1$ and $p'=\infty$, the following first line is produced. Then considering $\min\{1, m^{1/p}\}=1$ and also $\|\boldsymbol{A}\|_{1\to\infty}=\max_{i,j}|a_{i,j}|=\max_{i,j}|b_{i,j}|=\|\boldsymbol{B}\|_{1\to\infty}$, we can substitute $\|\boldsymbol{A}\|_{1\to\infty}$ by $\|\boldsymbol{B}\|_{1\to\infty}$ to have the following second line. Again, using the same lower-bound, this time for $q=1$, $p=\infty$, $q'=q$ and $p'=p$, we reach the following third line:

$$\begin{aligned}
\|\boldsymbol{A}\|_{q\to p} &\geq \min\left\{1, m^{\frac{1}{p}}\right\} \min\left\{1, n^{1-\frac{1}{q}}\right\} \|\boldsymbol{A}\|_{1\to\infty} \\
&= \min\left\{1, n^{1-\frac{1}{q}}\right\} \|\boldsymbol{B}\|_{1\to\infty} \\
&\geq \min\left\{1, n^{1-\frac{1}{q}}\right\} \min\left\{1, m^{-\frac{1}{p}}\right\} \min\left\{1, n^{\frac{1}{q}-1}\right\} \|\boldsymbol{B}\|_{q\to p} \\
&= m^{-\frac{1}{p}} n^{-\left|1-\frac{1}{q}\right|} \|\boldsymbol{B}\|_{q\to p} .
\end{aligned}$$

Therefore,

$$\frac{\|\boldsymbol{A}\|_{q\to p}}{\|\boldsymbol{B}\|_{q\to p}} \geq m^{-\frac{1}{p}} n^{-\left|1-\frac{1}{q}\right|} \xrightarrow{m^{-\frac{1}{p}} n^{-\left|1-\frac{1}{q}\right|}\leq 1} \|\boldsymbol{A}\|_{q\to p} \leq \|\boldsymbol{B}\|_{q\to p} .$$

Similarly, using the upper-bound of the mentioned set of bounds, i.e., $\forall\,(q,p,q',p')\in\mathbb{R}^4_{>0}, \forall \boldsymbol{A}\in\mathbb{R}^{m\times n}$ : $\|\boldsymbol{A}\|_{q\to p}\leq \max\{1, m^{1/p-1/p'}\}\max\{1, n^{1/q'-1/q}\}\|\boldsymbol{A}\|_{q'\to p'}$, and following above steps, we obtain:

$$\frac{\|\boldsymbol{B}\|_{q\to p}}{\|\boldsymbol{A}\|_{q\to p}} \leq m^{\frac{1}{p}} n^{\left|1-\frac{1}{q}\right|} \xRightarrow{m^{\frac{1}{p}} n^{\left|1-\frac{1}{q}\right|}\geq 1} \|\boldsymbol{A}\|_{q\to p} \leq \|\boldsymbol{B}\|_{q\to p}\,.$$

□

## A.3 Proof of Property 2.5 (Block-MCC$_{q,p}$ inequalities, page 58)

*Proof.* From Block-MCC$_{q,p}$ (Definition 2.6, page 54) we have:

$$\forall (q,p) \in \mathbb{R}^2_{>0}, \qquad M_{q,p}(\mathbf{\Phi}) \overset{\text{def}}{=} \max_{k,k' \neq k} \frac{d_k^{-\frac{1}{p}} d_{k'}^{\frac{1}{q}}}{d_{max}} \left\| \mathbf{\Phi}^\dagger [k] \, \mathbf{\Phi} \left[k'\right] \right\|_{q \to p}.$$

Next, considering the relationship between the $\ell_{1\to1}$ and $\ell_{1\to2}$, and also $\ell_{1\to2}$ and $\ell_{1\to\infty}$ operator-norms, which are presented in general in Property 2.4 ($\ell_{q\to p}$ operator-norm properties) and for certain $(q,p)$ pairs in table 2.3, the proof is straightforward, as follows:

$$\begin{aligned}
M_{1,1}(\mathbf{\Phi}) &\overset{\text{def}}{=} \max_{k,k' \neq k} && \frac{d_k^{-1} d_{k'}}{d_{max}} \left\| \mathbf{\Phi}^\dagger [k] \, \mathbf{\Phi} \left[k'\right] \right\|_{1\to1} \\
&\leq \max_{k,k' \neq k} && \frac{d_k^{-1} d_{k'}}{d_{max}} d_k^{\frac{1}{2}} \left\| \mathbf{\Phi}^\dagger [k] \, \mathbf{\Phi} \left[k'\right] \right\|_{1\to2} \\
&= \max_{k,k' \neq k} && \frac{d_k^{-\frac{1}{2}} d_{k'}}{d_{max}} \left\| \mathbf{\Phi}^\dagger [k] \, \mathbf{\Phi} \left[k'\right] \right\|_{1\to2} \\
&= M_{1,2}(\mathbf{\Phi}) && \leq \max_{k,k' \neq k} \frac{d_k^{-\frac{1}{2}} d_{k'}}{d_{max}} d_k^{\frac{1}{2}} \left\| \mathbf{\Phi}^\dagger [k] \, \mathbf{\Phi} \left[k'\right] \right\|_{1\to\infty} \\
& && = \max_{k,k' \neq k} \frac{d_{k'}}{d_{max}} \left\| \mathbf{\Phi}^\dagger [k] \, \mathbf{\Phi} \left[k'\right] \right\|_{1\to\infty} \\
& && = M_{1,\infty}(\mathbf{\Phi}).
\end{aligned}$$

Similarly, for the second inequality taking into account the relationship between the $\ell_{2\to2}$ and $\ell_{1\to2}$ operator-norms, we have:

$$\begin{aligned}
M_{2,2}(\mathbf{\Phi}) &\overset{\text{def}}{=} \max_{k,k' \neq k} \frac{d_k^{-\frac{1}{2}} d_{k'}^{\frac{1}{2}}}{d_{max}} \left\| \mathbf{\Phi}^\dagger [k] \, \mathbf{\Phi} \left[k'\right] \right\|_{2\to2} \\
&\leq \max_{k,k' \neq k} \frac{d_k^{-\frac{1}{2}} d_{k'}^{\frac{1}{2}}}{d_{max}} d_{k'}^{\frac{1}{2}} \left\| \mathbf{\Phi}^\dagger [k] \, \mathbf{\Phi} \left[k'\right] \right\|_{1\to2} \\
&= \max_{k,k' \neq k} \frac{d_k^{-\frac{1}{2}} d_{k'}}{d_{max}} \left\| \mathbf{\Phi}^\dagger [k] \, \mathbf{\Phi} \left[k'\right] \right\|_{1\to2} \\
&= M_{1,2}(\mathbf{\Phi}).
\end{aligned}$$

Then, for the third inequality considering the relationship between the $\ell_{2\to2}$ and $\ell_{2\to\infty}$,

and also $\ell_{2\to\infty}$ and $\ell_{1\to\infty}$ operator-norms, we have:

$$\begin{aligned}
M_{2,2}\left(\boldsymbol{\Phi}\right) &\overset{\text{def}}{=} \max_{k,k'\neq k} && \frac{d_k^{-\frac{1}{2}} d_{k'}^{\frac{1}{2}}}{d_{max}} \left\| \boldsymbol{\Phi}^{\dagger}\left[k\right] \boldsymbol{\Phi}\left[k'\right] \right\|_{2\to 2} \\
&\leq \max_{k,k'\neq k} && \frac{d_k^{-\frac{1}{2}} d_{k'}^{\frac{1}{2}}}{d_{max}} d_k^{\frac{1}{2}} \left\| \boldsymbol{\Phi}^{\dagger}\left[k\right] \boldsymbol{\Phi}\left[k'\right] \right\|_{2\to\infty} \\
&= \max_{k,k'\neq k} && \frac{d_{k'}^{\frac{1}{2}}}{d_{max}} \left\| \boldsymbol{\Phi}^{\dagger}\left[k\right] \boldsymbol{\Phi}\left[k'\right] \right\|_{2\to\infty} \\
&= M_{2,\infty}\left(\boldsymbol{\Phi}\right) && \leq \max_{k,k'\neq k} \frac{d_{k'}^{\frac{1}{2}}}{d_{max}} d_{k'}^{\frac{1}{2}} \left\| \boldsymbol{\Phi}^{\dagger}\left[k\right] \boldsymbol{\Phi}\left[k'\right] \right\|_{1\to\infty} \\
& && = \max_{k,k'\neq k} \frac{d_{k'}}{d_{max}} \left\| \boldsymbol{\Phi}^{\dagger}\left[k\right] \boldsymbol{\Phi}\left[k'\right] \right\|_{1\to\infty} \\
& && = M_{1,\infty}\left(\boldsymbol{\Phi}\right).
\end{aligned}$$

Finally, for the last inequality we need to know the relationship between the $\ell_{\infty\to\infty}$ and $\ell_{2\to\infty}$ operator-norms:

$$\begin{aligned}
M_{\infty,\infty}\left(\boldsymbol{\Phi}\right) &\overset{\text{def}}{=} \max_{k,k'\neq k} \frac{1}{d_{max}} \left\| \boldsymbol{\Phi}^{\dagger}\left[k\right] \boldsymbol{\Phi}\left[k'\right] \right\|_{\infty\to\infty} \\
&\leq \max_{k,k'\neq k} \frac{1}{d_{max}} d_{k'}^{\frac{1}{2}} \left\| \boldsymbol{\Phi}^{\dagger}\left[k\right] \boldsymbol{\Phi}\left[k'\right] \right\|_{2\to\infty} \\
&= \max_{k,k'\neq k} \frac{d_{k'}^{\frac{1}{2}}}{d_{max}} \left\| \boldsymbol{\Phi}^{\dagger}\left[k\right] \boldsymbol{\Phi}\left[k'\right] \right\|_{2\to\infty} \\
&= M_{2,\infty}\left(\boldsymbol{\Phi}\right).
\end{aligned}$$

For the proof of general relationships of figure 2.4, utilising the upper-bound of second set of bounds in Property 2.4 ($\ell_{q\to p}$ operator-norm properties), i.e., $\forall\left(q,p,q',p'\right)\in\mathbb{R}^4_{>0}, \forall \boldsymbol{A}\in\mathbb{R}^{m\times n}$ : $\|\boldsymbol{A}\|_{q\to p}\leq \max\{1, m^{1/p-1/p'}\}\max\{1, n^{1/q'-1/q}\}\|\boldsymbol{A}\|_{q'\to p'}$, and for a fixed $p{=}p'$, and $q'{\leq}q$, which yields $\|\boldsymbol{A}\|_{q\to p}{\leq} n^{1/q'-1/q}\|\boldsymbol{A}\|_{q'\to p}$, we get the following inequalities:

$$\begin{aligned}
\forall(q,p,q')\in\mathbb{R}^3_{>0}, q'\leq q, \qquad M_{q,p}\left(\boldsymbol{\Phi}\right) &\overset{\text{def}}{=} \max_{k,k'\neq k} \frac{d_k^{-\frac{1}{p}} d_{k'}^{\frac{1}{q}}}{d_{max}} \left\| \boldsymbol{\Phi}^{\dagger}\left[k\right] \boldsymbol{\Phi}\left[k'\right] \right\|_{q\to p} \\
&\leq \max_{k,k'\neq k} \frac{d_k^{-\frac{1}{p}} d_{k'}^{\frac{1}{q}}}{d_{max}} d_{k'}^{\frac{1}{q'}-\frac{1}{q}} \left\| \boldsymbol{\Phi}^{\dagger}\left[k\right] \boldsymbol{\Phi}\left[k'\right] \right\|_{q'\to p} \\
&= \max_{k,k'\neq k} \frac{d_k^{-\frac{1}{p}} d_{k'}^{\frac{1}{q'}}}{d_{max}} \left\| \boldsymbol{\Phi}^{\dagger}\left[k\right] \boldsymbol{\Phi}\left[k'\right] \right\|_{q'\to p} \\
&= M_{q',p}\left(\boldsymbol{\Phi}\right).
\end{aligned}$$

Similarly, using the same upper-bound of Property 2.4, this time for a fixed $q{=}q'$, and $p{\leq}p'$,

which yields $\|\boldsymbol{A}\|_{q\to p}\leq m^{1/p-1/p'}\|\boldsymbol{A}\|_{q\to p'}$, we get the following inequalities:

$$\begin{aligned}
\forall(q,p,p')\in\mathbb{R}^3_{>0}, p\leq p', \qquad M_{q,p}\left(\boldsymbol{\Phi}\right) &\overset{\text{def}}{=} \max_{k,k'\neq k}\frac{d_k^{-\frac{1}{p}}d_{k'}^{\frac{1}{q}}}{d_{max}}\left\|\boldsymbol{\Phi}^\dagger\left[k\right]\boldsymbol{\Phi}\left[k'\right]\right\|_{q\to p}\\
&\leq \max_{k,k'\neq k}\frac{d_k^{-\frac{1}{p}}d_{k'}^{\frac{1}{q}}}{d_{max}}d_k^{\frac{1}{p}-\frac{1}{p'}}\left\|\boldsymbol{\Phi}^\dagger\left[k\right]\boldsymbol{\Phi}\left[k'\right]\right\|_{q\to p'}\\
&= \max_{k,k'\neq k}\frac{d_k^{-\frac{1}{p'}}d_{k'}^{\frac{1}{q}}}{d_{max}}\left\|\boldsymbol{\Phi}^\dagger\left[k\right]\boldsymbol{\Phi}\left[k'\right]\right\|_{q\to p'}\\
&= M_{q,p'}\left(\boldsymbol{\Phi}\right).
\end{aligned}$$

□

## A.4 Proof of Property 2.7 (Block-MCC$_{q,p}$ bounds with intra-block orthonormality I, page 59)

*Proof.* Using the definition of $M_{q,p}(\boldsymbol{\Phi})$ in Property 2.6, and submultiplicativity property of operator-norm introduced in Property 2.4 ($\ell_{q\to p}$ operator-norm properties), we have:

$$\forall\,(q,p)\in\mathbb{R}^2_{>0},\qquad M_{q,p}\left(\boldsymbol{\Phi}\right)=\max_{k,k'\neq k}\frac{d_k^{-\frac{1}{p}}d_{k'}^{\frac{1}{q}}}{d_{max}}\left\|\boldsymbol{\Phi}^T\left[k\right]\boldsymbol{\Phi}\left[k'\right]\right\|_{q\to p}$$
$$\leq\max_{k,k'\neq k}\frac{d_k^{-\frac{1}{p}}d_{k'}^{\frac{1}{q}}}{d_{max}}\left\|\boldsymbol{\Phi}^T\left[k\right]\right\|_{q\to p}\left\|\boldsymbol{\Phi}[k']\right\|_{q\to p}\max\left\{1,m^{\frac{1}{q}-\frac{1}{p}}\right\}. \tag{A.3}$$

On the other hand, from the following property, $\|\cdot\|_{q\to p}$ can be bounded in terms of $\|\cdot\|_{2\to 2}$, which is often called the spectral norm:

**Property A.1 (Bounds of $\ell_{q\to p}$ operator-norm in terms of $\ell_{2\to 2}$).** The bounds of the $\ell_{q\to p}$ operator-norm of a matrix $\boldsymbol{A}\in\mathbb{R}^{m\times n}$ in terms of its $\ell_{2\to 2}$ operator-norm based on the second set of bounds in Property 2.4 ($\ell_{q\to p}$ operator-norm properties) for $q'=p'=2$ is:

$$\|\boldsymbol{A}\|_{q\to p}\geq\min\left\{1,m^{\frac{1}{p}-\frac{1}{2}}\right\}\min\left\{1,n^{\frac{1}{2}-\frac{1}{q}}\right\}\|\boldsymbol{A}\|_{2\to 2},$$
$$\|\boldsymbol{A}\|_{q\to p}\leq\max\left\{1,m^{\frac{1}{p}-\frac{1}{2}}\right\}\max\left\{1,n^{\frac{1}{2}-\frac{1}{q}}\right\}\|\boldsymbol{A}\|_{2\to 2}.$$

These bounds for different values of $q$ and $p$ are shown in table A.1.

Therefore, based on Property A.1, we upper-bound the $\ell_{q\to p}$ operator-norms of the upper-bound of Block-MCC in equation (A.3):

$$M_{q,p}\left(\boldsymbol{\Phi}\right)\leq\max_{k,k'\neq k}\frac{d_k^{-\frac{1}{p}}d_{k'}^{\frac{1}{q}}}{d_{max}}\max\left\{1,d_k^{\frac{1}{p}-\frac{1}{2}}\right\}\max\left\{1,m^{\frac{1}{2}-\frac{1}{q}}\right\}\left\|\boldsymbol{\Phi}^T\left[k\right]\right\|_{2\to 2}\times$$
$$\max\left\{1,m^{\frac{1}{p}-\frac{1}{2}}\right\}\max\left\{1,d_{k'}^{\frac{1}{2}-\frac{1}{q}}\right\}\left\|\boldsymbol{\Phi}[k']\right\|_{2\to 2}\max\left\{1,m^{\frac{1}{q}-\frac{1}{p}}\right\}. \tag{A.4}$$

But $\forall k,\|\boldsymbol{\Phi}^T[k]\|_{2\to 2}=\|\boldsymbol{\Phi}[k]\|_{2\to 2}=1$, because for a typical matrix $\boldsymbol{A}\in\mathbb{R}^{m\times n}$ we have $\|\boldsymbol{A}\|_{2\to 2}=\sigma_{max}(\boldsymbol{A})=\sqrt{\lambda_{max}(\boldsymbol{A}^T\boldsymbol{A})}$, where, $\sigma$ and $\lambda$ are singular value and eigenvalue, respectively, but here $\boldsymbol{A}$ is an orthonormal matrix, then $\boldsymbol{A}^T\boldsymbol{A}=\boldsymbol{I}_n$, and then $\lambda_{max}(\boldsymbol{I}_n)=1$.

| | $q\,\&\,p\leq 2$ | $q\,\&\,p\geq 2$ | $q\leq 2\,\&\,p\geq 2$ | $q\geq 2\,\&\,p\leq 2$ |
|---|---|---|---|---|
| $\frac{\|\boldsymbol{A}\|_{q\to p}}{\|\boldsymbol{A}\|_{2\to 2}}\leq$ | $m^{\frac{1}{p}-\frac{1}{2}}$ | $n^{-\frac{1}{q}+\frac{1}{2}}$ | $1$ | $m^{\frac{1}{p}-\frac{1}{2}}n^{-\frac{1}{q}+\frac{1}{2}}$ |
| $\frac{\|\boldsymbol{A}\|_{q\to p}}{\|\boldsymbol{A}\|_{2\to 2}}\geq$ | $n^{-\frac{1}{q}+\frac{1}{2}}$ | $m^{\frac{1}{p}-\frac{1}{2}}$ | $m^{\frac{1}{p}-\frac{1}{2}}n^{-\frac{1}{q}+\frac{1}{2}}$ | $1$ |

Table A.1: Bounds of $\|\boldsymbol{A}\|_{q\to p}/\|\boldsymbol{A}\|_{2\to 2}$ for different values of $q$ and $p$, where, $\boldsymbol{A}$ is a $m$ by $n$ matrix.

| | $\max M_{q,p}(\mathbf{\Phi})$ | | $\max M_{q,p}(\mathbf{\Phi})$ |
|---|---|---|---|
| $0 < q \,\&\, p \leq 2$ | $\max\limits_{k,k' \neq k} d_k^{-\frac{1}{2}}\, d_{k'}^{\frac{1}{q}}\, d_{max}^{-1}\, m^{\frac{1}{p}-\frac{1}{2}}\, \max\left\{1, m^{\frac{1}{q}-\frac{1}{p}}\right\}$ | $\leq$ | $d_{min}^{-\frac{1}{2}}\, d_{max}^{\frac{1}{q}-1}\, m^{\frac{1}{p}-\frac{1}{2}}\, \max\left\{1, m^{\frac{1}{q}-\frac{1}{p}}\right\}$ |
| $q \,\&\, p \geq 2$ | $\max\limits_{k,k' \neq k} d_k^{-\frac{1}{p}}\, d_{k'}^{\frac{1}{2}}\, d_{max}^{-1}\, m^{\frac{1}{2}-\frac{1}{q}}\, \max\left\{1, m^{\frac{1}{q}-\frac{1}{p}}\right\}$ | $\leq$ | $d_{min}^{-\frac{1}{p}}\, d_{max}^{-\frac{1}{2}}\, m^{\frac{1}{2}-\frac{1}{q}}\, \max\left\{1, m^{\frac{1}{q}-\frac{1}{p}}\right\}$ |
| $0 < q \leq 2 \,\&\, p \geq 2$ | $\max\limits_{k,k' \neq k} d_k^{-\frac{1}{p}}\, d_{k'}^{\frac{1}{q}}\, d_{max}^{-1}\, m^{\frac{1}{q}-\frac{1}{p}}$ | $\leq$ | $d_{min}^{-\frac{1}{p}}\, d_{max}^{\frac{1}{q}-1}\, m^{\frac{1}{q}-\frac{1}{p}}$ |
| $q \geq 2 \,\&\, 0 < p \leq 2$ | $\max\limits_{k,k' \neq k} d_k^{-\frac{1}{2}}\, d_{k'}^{\frac{1}{2}}\, d_{max}^{-1}\, m^{\frac{1}{p}-\frac{1}{q}}$ | $\leq$ | $d_{min}^{-\frac{1}{2}}\, d_{max}^{-\frac{1}{2}}\, m^{\frac{1}{p}-\frac{1}{q}}$ |

Table A.2: Upper-bound of Block-MCC$_{q,p}$ for different ranges of $q$ and $p$ and for a dictionary with intra-block orthonormality.

Therefore, the inequality (A.4) is developed into table A.2. In each case in the last column of the table A.2, the maximum value is achieved for minimum block length $d_{min}$ with negative power, and maximum block length $d_{max}$ with positive power. The table 2.4 shows the upper-bound of $M_{q,p}(\mathbf{\Phi})$ based on table A.2, for the basic tractable values of $q$ and $p$ based on table 2.1. □

## A.5 Proof of Property 2.8 (Block-MCC$_{q,p}$ bounds, page 59)

*Proof.* 1) Since the operator-norm is nonnegative (Property 2.4), it can be seen that $\forall k,k'\neq k, \forall (q,p)\in\mathbb{R}^2_{\geq 0} : \|\boldsymbol{\Phi}^{\dagger}[k]\boldsymbol{\Phi}[k']\|_{q\to p}\geq 0 \Rightarrow \forall (q,p)\in\mathbb{R}^2_{>0} : M_{q,p}(\boldsymbol{\Phi})\geq 0$. Considering that for all $k$ the blocks $\boldsymbol{\Phi}[k]$ are full column rank, using the Moore-Penrose pseudo-inverse property of matrices, $\forall (q,p)\in\mathbb{R}^2_{>0}$ we have:

$$\begin{aligned} M_{q,p}\left(\boldsymbol{\Phi}\right) &= \max_{k,k'\neq k} \frac{d_k^{-\frac{1}{p}} d_{k'}^{\frac{1}{q}}}{d_{max}} \left\| \boldsymbol{\Phi}^{\dagger}\left[k\right] \boldsymbol{\Phi}\left[k'\right] \right\|_{q\to p} \\ &= \max_{k,k'\neq k} \frac{d_k^{-\frac{1}{p}} d_{k'}^{\frac{1}{q}}}{d_{max}} \left\| \left(\boldsymbol{\Phi}^T\left[k\right] \boldsymbol{\Phi}\left[k\right]\right)^{-1} \boldsymbol{\Phi}^T\left[k\right] \boldsymbol{\Phi}\left[k'\right] \right\|_{q\to p}. \end{aligned}$$

Applying the general format of submultiplicativity property of operator-norm of a matrix, which is introduced in the Property 2.4 ($\ell_{q\to p}$ operator-norm properties) $\forall (q,p)\in\mathbb{R}^2_{>0}$, on the above last equality, $\forall (q,p)\in\mathbb{R}^2_{>0}$ we have:

$$M_{q,p}\left(\boldsymbol{\Phi}\right) \leq \max_{k,k'\neq k} \frac{d_k^{-\frac{1}{p}} d_{k'}^{\frac{1}{q}}}{d_{max}} \left\| \left(\boldsymbol{\Phi}^T\left[k\right] \boldsymbol{\Phi}\left[k\right]\right)^{-1} \right\|_{q\to p} \left\| \boldsymbol{\Phi}^T\left[k\right] \boldsymbol{\Phi}\left[k'\right] \right\|_{q\to p} \max\left\{1, d_k^{\frac{1}{q}-\frac{1}{p}}\right\}. \quad \text{(A.5)}$$

Since each element of the $\boldsymbol{\Phi}^T[k]\boldsymbol{\Phi}[k']$ is the pairwise correlation between the columns of the $k$ and $k'\neq k$ blocks, for any $k$ and $k'\neq k$ the $\boldsymbol{\Phi}^T[k]\boldsymbol{\Phi}[k']$ can be upper bounded by $M(\boldsymbol{\Phi})\,\mathbf{1}_{d_k\times d_{k'}}$, so it is true for their operator-norms, because using Property 2.4 ($\ell_{q\to p}$ operator-norm properties) for comparison of $\ell_{q\to p}$ operator-norm of two matrices (the forth property of the part Bounds), we have the following first line (based on table 2.1, the maximum absolute entry of a matrix is represented by its $\ell_{1\to\infty}$ operator-norm), whereas by the homogeneity property of $\ell_{q\to p}$ operator-norm defined in Property 2.4 $\forall q\geq 0$ and $\forall p>0$, the following second line is obtained $\forall (q,p)\in\mathbb{R}^2_{>0}$:

$$\begin{aligned} \forall k,k'\neq k, \qquad \left\| \boldsymbol{\Phi}^T\left[k\right] \boldsymbol{\Phi}\left[k'\right] \right\|_{q\to p} &\leq \left\| \left\| \boldsymbol{\Phi}^T\left[k\right] \boldsymbol{\Phi}\left[k'\right] \right\|_{1\to\infty} \mathbf{1}_{d_k\times d_{k'}} \right\|_{q\to p} \\ &= \left\| \boldsymbol{\Phi}^T\left[k\right] \boldsymbol{\Phi}\left[k'\right] \right\|_{1\to\infty} \left\| \mathbf{1}_{d_k\times d_{k'}} \right\|_{q\to p} \\ &\leq M\left(\boldsymbol{\Phi}\right) \left\| \mathbf{1}_{d_k\times d_{k'}} \right\|_{q\to p}. \end{aligned}$$

The above last line comes from the fact that $\forall k,k'\neq k$, the maximum absolute value of multiplication of blocks $k$ and $k'$ is less than or equal to $M(\boldsymbol{\Phi})$, which is by definition the maximum off-diagonal absolute value of multiplication of the whole dictionary to itself, i.e., $M(\boldsymbol{\Phi})\stackrel{\text{def}}{=}\max_{k,k'\neq k}|\boldsymbol{G}_{k,k'}(\boldsymbol{\Phi})|$, where $\boldsymbol{G}(\boldsymbol{\Phi})\stackrel{\text{def}}{=}\boldsymbol{\Phi}^T\boldsymbol{\Phi}$.

Then, using the upper-bound of the third set of bounds in Property 2.4, we will have the following second line. Then, by computing the Frobenius norm in the second line, $\forall k,k'\neq k$ and $\forall (q,p,q',p')\in\mathbb{R}^4_{>0}$ we get the following third line:

$$\begin{aligned}\left\|\boldsymbol{\Phi}^T[k]\,\boldsymbol{\Phi}\left[k'\right]\right\|_{q\to p} &\le M(\boldsymbol{\Phi})\left\|\mathbf{1}_{d_k\times d_{k'}}\right\|_{q\to p}\\ &\le M(\boldsymbol{\Phi})\max\left\{1,d_k^{\frac{1}{p}-\frac{1}{p'}}\right\}\max\left\{1,d_{k'}^{\frac{1}{q'}-\frac{1}{q}}\right\}\max\left\{1,d_k^{\frac{1}{p'}-\frac{1}{2}}\right\}\max\left\{1,d_{k'}^{\frac{1}{2}-\frac{1}{q'}}\right\}\left\|\mathbf{1}_{d_k\times d_{k'}}\right\|_F\\ &= (d_k\,d_{k'})^{\frac{1}{2}}\;M(\boldsymbol{\Phi})\max\left\{1,d_k^{\frac{1}{p}-\frac{1}{p'}}\right\}\max\left\{1,d_{k'}^{\frac{1}{q'}-\frac{1}{q}}\right\}\max\left\{1,d_k^{\frac{1}{p'}-\frac{1}{2}}\right\}\max\left\{1,d_{k'}^{\frac{1}{2}-\frac{1}{q'}}\right\}.\end{aligned} \tag{A.6}$$

Therefore, substituting $\|\boldsymbol{\Phi}^T[k]\boldsymbol{\Phi}[k']\|_{q\to p}$ with its upper-bound in the (A.5), $\forall(q,p,q',p')\in\mathbb{R}^4_{>0}$ we get:

$$\begin{aligned}M_{q,p}(\boldsymbol{\Phi}) \le M(\boldsymbol{\Phi})\max_{k,k'\neq k}\frac{d_k^{\frac{1}{2}-\frac{1}{p}}\,d_{k'}^{\frac{1}{q}+\frac{1}{2}}}{d_{max}}\left\|\left(\boldsymbol{\Phi}^T[k]\,\boldsymbol{\Phi}[k]\right)^{-1}\right\|_{q\to p}\times\\ \max\left\{1,d_k^{\frac{1}{p}-\frac{1}{p'}}\right\}\max\left\{1,d_{k'}^{\frac{1}{q'}-\frac{1}{q}}\right\}\max\left\{1,d_k^{\frac{1}{p'}-\frac{1}{2}}\right\}\max\left\{1,d_{k'}^{\frac{1}{2}-\frac{1}{q'}}\right\}\max\left\{1,d_k^{\frac{1}{q}-\frac{1}{p}}\right\}.\end{aligned} \tag{A.7}$$

Now an upper bound for $\|(\boldsymbol{\Phi}^T[k]\boldsymbol{\Phi}[k])^{-1}\|_{q\to p}$ should be developed. As part of the required tools to reach this aim we will utilise the following property:

**Property A.2.** For $q\ge p\ge 1$, if $\|\boldsymbol{A}\|_{q\to p}<1$, then [HR12]

$$(\boldsymbol{I}+\boldsymbol{A})^{-1}=\sum_{i=0}^{\infty}(-\boldsymbol{A})^i\,.$$

*Proof.* It follows from Corollary 5.6.16 of [HR12]. Based on the Corollary 5.6.16 of [HR12], the *matrix norm* of $\boldsymbol{A}$ should be less than one. Therefore, we imposed the constraint $q\ge p\ge 1$, because according to Remark 2.5 (matrix norm) on page 56 the $\ell_{q\to p}$ operator-norm for $q\ge p\ge 1$ satisfies all the required properties to be conceived as a matrix norm. □

Therefore, according to Property A.2, we need to decompose the $\boldsymbol{\Phi}^T[k]\boldsymbol{\Phi}[k]$ into a sum of two matrices, which one of them is an identity matrix. For now we continue with the current constraints on $q$ and $p$, i.e., $\forall(q,p)\in\mathbb{R}^2_{>0}$, but as soon as we use the above property the constraints will change to $q\ge p\ge 1$.

$\boldsymbol{\Phi}^T[k]\boldsymbol{\Phi}[k]$ is a $d_k\times d_k$ matrix which all of its diagonal elements are one, and its off-diagonal elements measure pairwise correlation between the columns of the same block $k$. Therefore, $\boldsymbol{\Phi}^T[k]\boldsymbol{\Phi}[k]$ can be decomposed as $\boldsymbol{\Phi}^T[k]\boldsymbol{\Phi}[k]=\boldsymbol{I}_{d_k}+\boldsymbol{F}[k]$, where:

$$\forall k,\qquad f_{i,j}[k]=\begin{cases}0, & \text{if } i=j\\ \boldsymbol{\varphi}_i^T[k]\boldsymbol{\varphi}_j[k], & \text{if } i\neq j.\end{cases}$$

But we are interested in the upper-bound of $\|(\boldsymbol{\Phi}^T[k]\boldsymbol{\Phi}[k])^{-1}\|_{q\to p}$. By the mentioned decomposition we have the following first line of inequalities. Then, utilising the Property A.2,

assuming $q{\geq}p{\geq}1$ and $\|\boldsymbol{F}[k]\|_{q\to p}{<}1$ (its veracity will be investigated) the following second line is obtained. Next utilising the generalisation (from two matrices to more than two matrices) of the triangle inequality property of the $\ell_{q\to p}$ operator-norm introduced in Property 2.4 ($\ell_{q\to p}$ operator-norm properties) $\forall q{\in}\mathbb{R}_{\geq 0}, \forall p{\in}\mathbb{R}_{\geq 1}$, or for $p{=}0$, the following third line is obtained. Then utilising the generalisation (from two matrices to more than two matrices) of the submultiplicativity property of the $\ell_{q\to p}$ operator-norm introduced in Property 2.4 for $q{\geq}p{>}0$, the following forth line is obtained. Finally, utilising the sum of an infinite geometric series, i.e., $\sum_{i=0}^{\infty} r^i{=}1/(1-r)$, where $|r|{<}1$, the following last line is obtained, considering the assumption $\|\boldsymbol{F}[k]\|_{q\to p}{<}1$ (to be investigated).

$$\begin{aligned}
\forall k, q \geq p \geq 1, \qquad \left\| \left( \boldsymbol{\Phi}^T[k]\, \boldsymbol{\Phi}[k] \right)^{-1} \right\|_{q\to p} &= \left\| \left( \boldsymbol{I}_{d_k} + \boldsymbol{F}[k] \right)^{-1} \right\|_{q\to p} \\
&= \left\| \sum_{i=0}^{\infty} \left( -\boldsymbol{F}[k] \right)^i \right\|_{q\to p} \\
&\leq \sum_{i=0}^{\infty} \left\| \left( -\boldsymbol{F}[k] \right)^i \right\|_{q\to p} \\
&\leq \sum_{i=0}^{\infty} \|\boldsymbol{F}[k]\|_{q\to p}^i \\
&= \frac{1}{1 - \|\boldsymbol{F}[k]\|_{q\to p}}.
\end{aligned} \tag{A.8}$$

Now we investigate the veracity of the assumption $\|\boldsymbol{F}[k]\|_{q\to p}{<}1$. Since each off-diagonal entry of $\boldsymbol{F}[k]$ is the pairwise correlation between the columns of the $k^{th}$ block and the on-diagonal values are all zero, it can be upper bounded by $M(\boldsymbol{\Phi})$ $(\boldsymbol{1}_{d_k}{-}\boldsymbol{I}_{d_k})$, so it is true for their operator-norms, because using Property 2.4 ($\ell_{q\to p}$ operator-norm properties) for comparison of $\ell_{q\to p}$ operator-norm of two matrices $\forall (q,p){\in}\mathbb{R}^2_{>0}$ (the forth property of the part Bounds), we have:

$$\begin{aligned}
\forall k, q \geq p \geq 1, \qquad \|\boldsymbol{F}[k]\|_{q\to p} &\leq \| \|\boldsymbol{F}[k]\|_{1\to\infty} \, (\boldsymbol{1}_{d_k} - \boldsymbol{I}_{d_k}) \|_{q\to p} \\
&= \|\boldsymbol{F}[k]\|_{1\to\infty} \, \|(\boldsymbol{1}_{d_k} - \boldsymbol{I}_{d_k})\|_{q\to p} \\
&\leq M(\boldsymbol{\Phi}) \, \|(\boldsymbol{1}_{d_k} - \boldsymbol{I}_{d_k})\|_{q\to p}.
\end{aligned}$$

The $\ell_{1\to\infty}$ operator-norm of a matrix computes the maximum absolute entry of the matrix (table 2.1). The above second line uses the homogeneity property of $\ell_{q\to p}$ operator-norm defined in Property 2.4 ($\ell_{q\to p}$ operator-norm properties) $\forall q{\geq}0$ and $\forall p{>}0$. The above last line comes from the fact that $\forall k$, the maximum off-diagonal absolute value of multiplication of block $k$ to itself is less than or equal to $M(\boldsymbol{\Phi})$, because the off-diagonal values of $\boldsymbol{F}[k]$ are a subset of off-diagonal values of the Gram matrix $\boldsymbol{G}(\boldsymbol{\Phi}){\stackrel{\text{def}}{=}}\boldsymbol{\Phi}^T\boldsymbol{\Phi}$, which the $M(\boldsymbol{\Phi})$ is derived from, i.e., $M(\boldsymbol{\Phi}){\stackrel{\text{def}}{=}}\max_{k,k'\neq k} |\boldsymbol{G}_{k,k'}(\boldsymbol{\Phi})|{=}\max_{k,k'\neq k} |\boldsymbol{\varphi}_k^T\boldsymbol{\varphi}_{k'}|$, hence $\|\boldsymbol{F}[k]\|_{1\to\infty}{\leq}M(\boldsymbol{\Phi})$.

Then, using the upper-bound of the third set of bounds in Property 2.4 ($\ell_{q\to p}$ operator-norm properties), we will have the following second line. By computing the Frobenius norm in the second line, the third line is obtained. Considering the obtained constraint on $q$

and $p$, i.e., $q{\geq}p{\geq}1$, we choose $q'{=}p'{=}1$ in the following third line, because both $q$ and $p$ are lower-bounded by one.

$$\begin{aligned}
\|\boldsymbol{F}[k]\|_{q\to p} &\leq M(\boldsymbol{\Phi})\,\|(\mathbf{1}_{d_k}-\boldsymbol{I}_{d_k})\|_{q\to p}\\
&\leq M(\boldsymbol{\Phi})\,\max\left\{1,d_k^{\frac{1}{p}-\frac{1}{p'}}\right\}\max\left\{1,d_k^{\frac{1}{q'}-\frac{1}{q}}\right\}\max\left\{1,d_k^{\frac{1}{p'}-\frac{1}{2}}\right\}\max\left\{1,d_k^{\frac{1}{2}-\frac{1}{q'}}\right\}\|(\mathbf{1}_{d_k}-\boldsymbol{I}_{d_k})\|_F\\
&= \left(d_k^2-d_k\right)^{\frac{1}{2}}M(\boldsymbol{\Phi})\,\max\left\{1,d_k^{\frac{1}{p}-\frac{1}{p'}}\right\}\max\left\{1,d_k^{\frac{1}{q'}-\frac{1}{q}}\right\}\max\left\{1,d_k^{\frac{1}{p'}-\frac{1}{2}}\right\}\max\left\{1,d_k^{\frac{1}{2}-\frac{1}{q'}}\right\}\\
&= \left(d_k^2-d_k\right)^{\frac{1}{2}}\,d_k^{1-\frac{1}{q}}\,d_k^{\frac{1}{2}}\,M(\boldsymbol{\Phi})\\
&= d_k^{2-\frac{1}{q}}\,(d_k-1)^{\frac{1}{2}}\,M(\boldsymbol{\Phi}).
\end{aligned}$$

Therefore, for all $k$ we have $\|\boldsymbol{F}[k]\|_{q\to p}{\leq}d_{max}^{2-1/q}(d_{max}{-}1)^{1/2}M(\boldsymbol{\Phi})$. On the other hand, in the Property 2.8, it is assumed that $d_{max}^{2-1/q}(d_{max}{-}1)^{1/2}M(\boldsymbol{\Phi}){<}1$, so the required condition to use the Property A.2 in (A.8), i.e., $\|\boldsymbol{F}[k]\|_{q\to p}{<}1$, is satisfied. Substituting the obtained upper-bound of $\|\boldsymbol{F}[k]\|_{q\to p}$ in (A.8), we get:

$$\begin{aligned}
\forall k, q\geq p\geq 1,\qquad \left\|\left(\boldsymbol{\Phi}^T[k]\,\boldsymbol{\Phi}[k]\right)^{-1}\right\|_{q\to p} &\leq \frac{1}{1-\|\boldsymbol{F}[k]\|_{q\to p}}\\
&\leq \frac{1}{1-d_k^{2-\frac{1}{q}}\,(d_k-1)^{\frac{1}{2}}\,M(\boldsymbol{\Phi})}.
\end{aligned}$$

On the other hand, the previously obtained main equation (A.7) for new constraints $q{\geq}p{\geq}1$ ($\max\{1,d_k^{1/q-1/p}\}{=}1$) and $q'{=}p'{=}1$ becomes:

$$\begin{aligned}
M_{q,p}(\boldsymbol{\Phi}) &\leq M(\boldsymbol{\Phi})\max_{k,k'\neq k}\ \frac{d_k^{\frac{1}{2}-\frac{1}{p}}\,d_{k'}^{\frac{1}{q}+\frac{1}{2}}}{d_{max}}\left\|\left(\boldsymbol{\Phi}^T[k]\,\boldsymbol{\Phi}[k]\right)^{-1}\right\|_{q\to p}\times\\
&\qquad \max\left\{1,d_k^{\frac{1}{p}-\frac{1}{p'}}\right\}\max\left\{1,d_{k'}^{\frac{1}{q'}-\frac{1}{q}}\right\}\max\left\{1,d_k^{\frac{1}{p'}-\frac{1}{2}}\right\}\max\left\{1,d_{k'}^{\frac{1}{2}-\frac{1}{q'}}\right\}\max\left\{1,d_k^{\frac{1}{q}-\frac{1}{p}}\right\}\\
&= M(\boldsymbol{\Phi})\max_{k,k'\neq k}\ \frac{d_k^{\frac{1}{2}-\frac{1}{p}}\,d_{k'}^{\frac{1}{q}+\frac{1}{2}}}{d_{max}}\left\|\left(\boldsymbol{\Phi}^T[k]\,\boldsymbol{\Phi}[k]\right)^{-1}\right\|_{q\to p}\times\\
&\qquad \max\left\{1,d_k^{\frac{1}{p}-1}\right\}\max\left\{1,d_{k'}^{1-\frac{1}{q}}\right\}\max\left\{1,d_k^{\frac{1}{2}}\right\}\max\left\{1,d_{k'}^{-\frac{1}{2}}\right\}\\
&= M(\boldsymbol{\Phi})\max_{k,k'\neq k}\ \frac{d_k^{1-\frac{1}{p}}\,d_{k'}^{\frac{3}{2}}}{d_{max}}\left\|\left(\boldsymbol{\Phi}^T[k]\,\boldsymbol{\Phi}[k]\right)^{-1}\right\|_{q\to p}.
\end{aligned}\tag{A.9}$$

Substituting the upper-bound of $\|(\boldsymbol{\Phi}^T[k]\boldsymbol{\Phi}[k])^{-1}\|_{q\to p}$ obtained in equation (A.5) into the

updated main equation (A.9), we get:

$$q \geq p \geq 1, \qquad M_{q,p}(\mathbf{\Phi}) \leq \frac{M(\mathbf{\Phi})}{d_{max}} \max_{k,k' \neq k} \frac{d_k^{1-\frac{1}{p}} d_{k'}^{\frac{3}{2}}}{1 - d_k^{2-\frac{1}{q}} (d_k - 1)^{\frac{1}{2}} M(\mathbf{\Phi})}$$
$$\leq \frac{M(\mathbf{\Phi})}{d_{max}} \max_k d_k^{1-\frac{1}{p}} \max_{k' \neq k} d_{k'}^{\frac{3}{2}} \max_k \frac{1}{1 - d_k^{2-\frac{1}{q}} (d_k - 1)^{\frac{1}{2}} M(\mathbf{\Phi})}$$
$$= \frac{d_{max}^{\frac{3}{2}-\frac{1}{p}} M(\mathbf{\Phi})}{1 - d_{max}^{2-\frac{1}{q}} (d_{max} - 1)^{\frac{1}{2}} M(\mathbf{\Phi})}.$$

2) Starting from the definition of coherence in Property 2.6 (block-MCC$_{q,p}$ for intra-block orthonormality) we have the following first line for $\forall (q,p) \in \mathbb{R}^2_{>0}$. Then considering the upper-bound of $\|\mathbf{\Phi}^T[k]\mathbf{\Phi}[k']\|_{q \to p}$ for any $k$ and $k' \neq k$, obtained in (A.6), we will have the following second line $\forall (q,p,q',p') \in \mathbb{R}^4_{>0}$:

$$M_{q,p}(\mathbf{\Phi}) = \max_{k,k' \neq k} \frac{d_k^{-\frac{1}{p}} d_{k'}^{\frac{1}{q}}}{d_{max}} \left\| \mathbf{\Phi}^T[k]\, \mathbf{\Phi}\left[k'\right] \right\|_{q \to p}$$
$$\leq \frac{M(\mathbf{\Phi})}{d_{max}} \max_{k,k' \neq k} d_k^{\frac{1}{2}-\frac{1}{p}} d_{k'}^{\frac{1}{q}+\frac{1}{2}} \max\left\{1, d_k^{\frac{1}{p}-\frac{1}{p'}}\right\} \max\left\{1, d_{k'}^{\frac{1}{q'}-\frac{1}{q}}\right\} \max\left\{1, d_k^{\frac{1}{p'}-\frac{1}{2}}\right\} \max\left\{1, d_{k'}^{\frac{1}{2}-\frac{1}{q'}}\right\}.$$

□

## A.6 Proof of Lemma 2.1 (Block-UP based on Block-Spark, page 65)

*Proof.* Here, we used the following triangle inequality:

$$\|\boldsymbol{\beta_0}\|_{p,0} + \|\boldsymbol{\beta_1}\|_{p,0} \geq \|\boldsymbol{\beta_0} - \boldsymbol{\beta_1}\|_{p,0}\,.$$

which is the the block-structured generalisation of one of the properties of the $\ell_0$ pseudo-norm operator for vector space. The above relation simply indicates that, if two representations have no overlapped supports, the sum of their number of active blocks are equal to the number of active blocks of their subtraction (equality), while in case of overlapped supports, the sum of their number of active blocks is greater than the number of active blocks of their subtraction (inequality).

On the other hand, we have $\boldsymbol{\Phi}(\boldsymbol{\beta_0}-\boldsymbol{\beta_1})=\mathbf{0}$. Therefore $(\boldsymbol{\beta_0}-\boldsymbol{\beta_1})$ is in the Kernel of the dictionary, and based on the definition of Block-Spark in Definition 2.5 (page 52), we have:

$$\|\boldsymbol{\beta_0} - \boldsymbol{\beta_1}\|_{p,0} \geq Block{-}Spark(\boldsymbol{\Phi}),$$

which proves the lemma. □

## A.7 Proof of Theorem 2.2 (Block-NSP, page 67)

*Proof.* Under the assumption of $Q_{\boldsymbol{w};p_1,p_2}(S_b(\boldsymbol{\beta}),\boldsymbol{\Phi})<1/2$ and $S_b(\boldsymbol{\beta_0})\subset S_b(\boldsymbol{\beta})$, to show that $\boldsymbol{\beta_0}$ is the unique solution to the $P_{\boldsymbol{w};p_1,p_2}$ minimisation problem, we need to prove that:

$$\forall \boldsymbol{x}\in Ker(\boldsymbol{\Phi}),\qquad \|\boldsymbol{\beta_0}\|^{p_2}_{\boldsymbol{w};p_1,p_2}<\|\boldsymbol{\beta_0}+\boldsymbol{x}\|^{p_2}_{\boldsymbol{w};p_1,p_2}\,,$$

where as defined in Definition 2.2 (weighted (pseudo-)mixed-norm, page 48), the $\ell^{\boldsymbol{w}}_{p_1,p_2}$ weighted (pseudo-)mixed-norm of a vector $\boldsymbol{\beta_0}$ is:

$$\|\boldsymbol{\beta_0}\|_{\boldsymbol{w};p_1,p_2}\overset{\text{def}}{=}\begin{cases}\displaystyle\sum_k I\left(\frac{\|\boldsymbol{\beta_0}\,[k]\|_{p_1}}{d_k^{\frac{1}{p_1}}}\right), & \text{for}\ \ p_2=0\\[2ex] \displaystyle\left(\sum_k\frac{\|\boldsymbol{\beta_0}\,[k]\|^{p_2}_{p_1}}{d_k^{\frac{p_2}{p_1}}}\right)^{\frac{1}{p_2}}, & \text{for}\ \ 0<p_2<+\infty\\[2ex] \displaystyle\max_k\left\{\frac{\|\boldsymbol{\beta_0}\,[k]\|_{p_1}}{d_k^{\frac{1}{p_1}}}\right\}, & \text{for}\ \ p_2=\infty.\end{cases}$$

Then, dividing the whole blocks to on-Block-Support ($\in S_b(\boldsymbol{\beta})$) and off-Block-Support ($\notin S_b(\boldsymbol{\beta})$), we have

$$0<\sum_{k\in S_b(\boldsymbol{\beta})}\left(\left|\frac{\sum_{j=1}^{d_k}\left|\beta_{0_j}\,[k]+x_j\,[k]\right|^{p_1}}{d_k}\right|^{\frac{p_2}{p_1}}-\left|\frac{\sum_{j=1}^{d_k}\left|\beta_{0_j}\,[k]\right|^{p_1}}{d_k}\right|^{\frac{p_2}{p_1}}\right)$$
$$+\sum_{k\notin S_b(\boldsymbol{\beta})}\left|\frac{\sum_{j=1}^{d_k}|x_j\,[k]|^{p_1}}{d_k}\right|^{\frac{p_2}{p_1}}.$$

It can be rewritten in terms of the norm operator, so

$$0<\sum_{k\in S_b(\boldsymbol{\beta})}w_k^{p_2}\left(\|\boldsymbol{\beta_0}\,[k]+\boldsymbol{x}\,[k]\|^{p_2}_{p_1}-\|\boldsymbol{\beta_0}\,[k]\|^{p_2}_{p_1}\right)+\sum_{k\notin S_b(\boldsymbol{\beta})}w_k^{p_2}\,\|\boldsymbol{x}\,[k]\|^{p_2}_{p_1}\,.$$

where, in a block-structured vector $\boldsymbol{a}=[\boldsymbol{a}^T[1],\cdots,\boldsymbol{a}^T[K]]^T$ with length vector $\boldsymbol{d}=[d_1,\cdots,d_K]$, $w_k$ in $w_k\|\boldsymbol{a}[k]\|_p$ is equal to $d_k^{-1/p}$.

From quasi-triangle inequality for scalars, we generalised it to vector space and derived the corresponding triangle inequality $\|\boldsymbol{a}+\boldsymbol{b}\|^{p_2}_{p_1}-\|\boldsymbol{a}\|^{p_2}_{p_1}\geq-\|\boldsymbol{b}\|^{p_2}_{p_1}$ for $0\leq p_2\leq 1\leq p_1$. Therefore, it is sufficient to prove:

$$0<\sum_{k\notin S_b(\boldsymbol{\beta})}w_k^{p_2}\,\|\boldsymbol{x}\,[k]\|^{p_2}_{p_1}-\sum_{k\in S_b(\boldsymbol{\beta})}w_k^{p_2}\,\|\boldsymbol{x}\,[k]\|^{p_2}_{p_1}\,.$$

Adding $2\sum_{k\in S_b(\boldsymbol{\beta})} w_k^{p_2}\|\boldsymbol{x}[k]\|_{p_1}^{p_2}$ to both sides, we have

$$2\sum_{k\in S_b(\boldsymbol{\beta})} w_k^{p_2}\left\|\boldsymbol{x}\left[k\right]\right\|_{p_1}^{p_2} < \sum_{k\notin S_b(\boldsymbol{\beta})} w_k^{p_2}\left\|\boldsymbol{x}\left[k\right]\right\|_{p_1}^{p_2} + \sum_{k\in S_b(\boldsymbol{\beta})} w_k^{p_2}\left\|\boldsymbol{x}\left[k\right]\right\|_{p_1}^{p_2} = \sum_{k} w_k^{p_2}\left\|\boldsymbol{x}\left[k\right]\right\|_{p_1}^{p_2},$$

or equivalently,

$$\frac{\sum_{k\in S_b(\boldsymbol{\beta})} w_k^{p_2}\left\|\boldsymbol{x}\left[k\right]\right\|_{p_1}^{p_2}}{\sum_{k} w_k^{p_2}\left\|\boldsymbol{x}\left[k\right]\right\|_{p_1}^{p_2}} < \frac{1}{2}.$$

But the above left-hand side expression is exactly $Q_{\boldsymbol{w};p_1,p_2}(S_b(\boldsymbol{\beta}),\boldsymbol{\Phi})$ and the inequality is the initial assumption of the proof, i.e., $Q_{\boldsymbol{w};p_1,p_2}(S_b(\boldsymbol{\beta}),\boldsymbol{\Phi})<1/2$.

Similarly, for the case of equally-sized blocks, i.e., $d_1=\cdots=d_K=d$, under the assumption of $Q_{p_1,p_2}(S_b(\boldsymbol{\beta}),\boldsymbol{\Phi})<1/2$ and $S_b(\boldsymbol{\beta_0})\subset S_b(\boldsymbol{\beta})$, to show that $\boldsymbol{\beta_0}$ is the unique solution to the $P_{p_1,p_2}$ minimisation problem, we need to prove that:

$$\forall \boldsymbol{x}\in Ker(\boldsymbol{\Phi}), \qquad \|\boldsymbol{\beta_0}\|_{p_1,p_2}^{p_2} < \|\boldsymbol{\beta_0}+\boldsymbol{x}\|_{p_1,p_2}^{p_2},$$

where as defined on page 34, the $\ell_{p_1,p_2}$ (pseudo-)mixed-norm of a vector $\boldsymbol{\beta_0}$ is:

$$\|\boldsymbol{\beta_0}\|_{p_1,p_2} \stackrel{\text{def}}{=} \begin{cases} \sum_k I\left(\left\|\boldsymbol{\beta_0}\left[k\right]\right\|_{p_1}\right), & \text{for } p_2=0 \\ \left(\sum_k \left\|\boldsymbol{\beta_0}\left[k\right]\right\|_{p_1}^{p_2}\right)^{\frac{1}{p_2}}, & \text{for } 0<p_2<+\infty \\ \max_k\left\{\left\|\boldsymbol{\beta_0}\left[k\right]\right\|_{p_1}\right\}, & \text{for } p_2=\infty, \end{cases}$$

Then, the above-mentioned inequality is equivalent to showing

$$\sum_{k=1}^{K}\left|\sum_{j=1}^{d_k}\left|\beta_{0_j}\left[k\right]\right|^{p_1}\right|^{\frac{p_2}{p_1}} < \sum_{k=1}^{K}\left|\sum_{j=1}^{d_k}\left|\beta_{0_j}\left[k\right]+x_j\left[k\right]\right|^{p_1}\right|^{\frac{p_2}{p_1}}.$$

Then, dividing the whole blocks to on-Block-Support ($\in S_b(\boldsymbol{\beta})$) and off-Block-Support ($\notin S_b(\boldsymbol{\beta})$), we have

$$0 < \sum_{k\in S_b(\boldsymbol{\beta})}\left(\left|\sum_{j=1}^{d_k}\left|\beta_{0_j}\left[k\right]+x_j\left[k\right]\right|^{p_1}\right|^{\frac{p_2}{p_1}} - \left|\sum_{j=1}^{d_k}\left|\beta_{0_j}\left[k\right]\right|^{p_1}\right|^{\frac{p_2}{p_1}}\right) + \sum_{k\notin S_b(\boldsymbol{\beta})}\left|\sum_{j=1}^{d_k}\left|x_j\left[k\right]\right|^{p_1}\right|^{\frac{p_2}{p_1}}.$$

It can be rewritten in terms of the norm operator, so

$$0 < \sum_{k \in S_b(\boldsymbol{\beta})} \|\boldsymbol{\beta_0}[k] + \boldsymbol{x}[k]\|_{p_1}^{p_2} - \|\boldsymbol{\beta_0}[k]\|_{p_1}^{p_2} + \sum_{k \notin S_b(\boldsymbol{\beta})} \|\boldsymbol{x}[k]\|_{p_1}^{p_2}.$$

Utilising $\|\boldsymbol{a}+\boldsymbol{b}\|_{p_1}^{p_2} - \|\boldsymbol{a}\|_{p_1}^{p_2} \geq -\|\boldsymbol{b}\|_{p_1}^{p_2}$ for $0 \leq p_2 \leq 1 \leq p_1$, it is sufficient to prove:

$$0 < \sum_{k \notin S_b(\boldsymbol{\beta})} \|\boldsymbol{x}[k]\|_{p_1}^{p_2} - \sum_{k \in S_b(\boldsymbol{\beta})} \|\boldsymbol{x}[k]\|_{p_1}^{p_2}.$$

Adding $2\sum_{k \in S_b(\boldsymbol{\beta})} \|\boldsymbol{x}[k]\|_{p_1}^{p_2}$ to both sides, we have

$$\begin{aligned} 2 \sum_{k \in S_b(\boldsymbol{\beta})} \|\boldsymbol{x}[k]\|_{p_1}^{p_2} &< \sum_{k \notin S_b(\boldsymbol{\beta})} \|\boldsymbol{x}[k]\|_{p_1}^{p_2} + \sum_{k \in S_b(\boldsymbol{\beta})} \|\boldsymbol{x}[k]\|_{p_1}^{p_2} \\ &= \sum_{k} \|\boldsymbol{x}[k]\|_{p_1}^{p_2}, \end{aligned}$$

or equivalently,

$$\frac{\sum_{k \in S_b(\boldsymbol{\beta})} \|\boldsymbol{x}[k]\|_{p_1}^{p_2}}{\sum_{k} \|\boldsymbol{x}[k]\|_{p_1}^{p_2}} < \frac{1}{2}.$$

But the above left-hand side expression is exactly $Q_{p_1,p_2}(S_b(\boldsymbol{\beta}), \boldsymbol{\Phi})$ and the inequality is the initial assumption of the proof, i.e., $Q_{p_1,p_2}(S_b(\boldsymbol{\beta}), \boldsymbol{\Phi}) < 1/2$. □

## A.8 Proof of Lemma 2.2 (Basic Block-UP, page 69)

*Proof.* The following proof generalises the corresponding proofs in [EB01]; [EB02]; [EB09]; [EKB10b] and [EKB10a]. Without loss of generality, assume that the non-zero signal $\boldsymbol{y}$ is normalized to have unit squared Euclidean norm, i.e., $\|\boldsymbol{y}\|_2^2{=}\boldsymbol{y}^T\boldsymbol{y}{=}1$, and suppose that $p$ is Hölder conjugate to $p'$, i.e., $1/p{+}1/p'{=}1$ [GL13]. Then,

$$\begin{aligned}
\forall(p,p')\in\mathbb{R}^2_{\geq 1},\forall q\in\mathbb{R}_{>0},\forall r\in\mathbb{R}_{\geq 0}, \qquad 1 &= \sum_{k,k'=1}^{K} \boldsymbol{\beta_1}^T[k]\,\boldsymbol{\Phi_1}^T[k]\,\boldsymbol{\Phi_2}\left[k'\right]\boldsymbol{\beta_2}\left[k'\right] \\
&\leq \sum_{k,k'=1}^{K} \|\boldsymbol{\beta_1}[k]\|_p \left\|\boldsymbol{\Phi_1}^T[k]\,\boldsymbol{\Phi_2}\left[k'\right]\boldsymbol{\beta_2}\left[k'\right]\right\|_{p'} \\
&\leq d_{max}\overline{M}_{q,p'}(\boldsymbol{\Phi_1},\boldsymbol{\Phi_2}) \sum_{k,k'=1}^{K} d_k^{\frac{1}{p'}} \|\boldsymbol{\beta_1}[k]\|_p\, d_{k'}^{-\frac{1}{q}} \left\|\boldsymbol{\beta_2}\left[k'\right]\right\|_q \\
&= d_{max}\overline{M}_{q,p'}(\boldsymbol{\Phi_1},\boldsymbol{\Phi_2}) \sum_{k=1}^{K} d_k^{\frac{1}{p'}} \|\boldsymbol{\beta_1}[k]\|_p \sum_{k'=1}^{K} d_{k'}^{-\frac{1}{q}} \left\|\boldsymbol{\beta_2}\left[k'\right]\right\|_q \\
&= d_{max}\overline{M}_{q,p'}(\boldsymbol{\Phi_1},\boldsymbol{\Phi_2}) \sum_{k=1}^{\|\boldsymbol{\beta_1}\|_{r,0}} d_k^{\frac{1}{p'}} \|\boldsymbol{\beta_1}[k]\|_p \sum_{k'=1}^{\|\boldsymbol{\beta_2}\|_{r,0}} d_{k'}^{-\frac{1}{q}} \left\|\boldsymbol{\beta_2}\left[k'\right]\right\|_q .
\end{aligned} \tag{A.10}$$

For vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ we have $\boldsymbol{a}^T\boldsymbol{b}{\leq}\sum_i|a_ib_i|$, and using Hölder's inequality [GL13], i.e., $\sum_i|a_ib_i|{\leq}\|\boldsymbol{a}\|_p\|\boldsymbol{b}\|_{p'}$, where, $\forall(p,p'){\in}\mathbb{R}^2_{\geq 1}$ : $1/p{+}1/p'{=}1$, we have $\boldsymbol{a}^T\boldsymbol{b}{\leq}\|\boldsymbol{a}\|_p\|\boldsymbol{b}\|_{p'}$, which results the above first inequality. Considering the definition of basic Block-MCC$_{q,p}$, i.e., $\overline{M}_{q,p'}(\boldsymbol{\Phi_1},\boldsymbol{\Phi_2}){=}\max_{k,k'} d_k^{-1/p'}\, d_{k'}^{1/q}/d_{max}\|\boldsymbol{\Phi_1}^T[k]\boldsymbol{\Phi_2}[k']\|_{q\to p'}$, where $p'{=}p/(p{-}1)$, we have $\|\boldsymbol{\Phi_1}^T[k]\boldsymbol{\Phi_2}[k']\boldsymbol{x}\|_{p'}{\leq}\overline{M}_{q,p'}(\boldsymbol{\Phi_1},\boldsymbol{\Phi_2})d_{max}d_k^{1/p'}d_{k'}^{-1/q}\|\boldsymbol{x}\|_q$, which produces the above third line with second inequality. The above last equality follows from summation only over non-zero blocks for any $r{\geq}0$.

Before finding the upper-bound for the above last inequality (A.10), it should be taken into account that according to the Parseval's theorem, we have $\|\boldsymbol{y}\|_2^2{=}\|\boldsymbol{\beta_1}\|_2^2{=}\|\boldsymbol{\beta_2}\|_2^2{=}1$, which will be used in the following optimisation problem. Therefore, in order to upper-bound the inequality (A.10), it is sufficient to solve the following optimisation problem:

$$\begin{aligned}
\forall(q,p,p')\in\mathbb{R}^3_{\geq 1},\forall r\in\mathbb{R}_{\geq 0}, \qquad \max_{k,k'} & \sum_{k=1}^{\|\boldsymbol{\beta_1}\|_{r,0}} d_k^{\frac{1}{p'}} \left(\sum_{j=1}^{d_k}\left|\beta_{1_j}[k]\right|^p\right)^{\frac{1}{p}} \sum_{k'=1}^{\|\boldsymbol{\beta_2}\|_{r,0}} d_{k'}^{-\frac{1}{q}} \left(\sum_{j=1}^{d_{k'}}\left|\beta_{2_j}\left[k'\right]\right|^q\right)^{\frac{1}{q}} \\
s.t. \quad & \sum_{k=1}^{\|\boldsymbol{\beta_1}\|_{r,0}}\sum_{j=1}^{d_k}\beta_{1_j}^2[k] = \sum_{k'=1}^{\|\boldsymbol{\beta_2}\|_{r,0}}\sum_{j=1}^{d_{k'}}\beta_{2_j}^2\left[k'\right] = 1.
\end{aligned} \tag{A.11}$$

Although the equation (A.10) holds true for $\forall q{\in}\mathbb{R}_{>0}$, we are going to upper-bound the (A.10) for $\forall q{\in}\mathbb{R}_{\geq 1}$, as it is mentioned in (A.11). On the other hand, since during the procedure of finding the upper-bound, the derivative of the $q$ and $p$ norms of a vector should be computed,

we solve (A.11) in two parts: 1) $\forall(q,p)\in\mathbb{R}^2_{>1}$ (differentiable $q$ and $p$ norms), and 2) $q{=}p{=}1$ (following the related proof in [EB02], page 3).

Above optimisation problem is separable, so we need to maximise $\forall(q,p)\in\mathbb{R}^2_{>1}$, $\sum_{k=1}^{\|\boldsymbol{\beta_1}\|_{r,0}} d_k^{1/p'}(\sum_{j=1}^{d_k}|\beta_{1_j}[k]|^p)^{1/p}$ subject to $\sum_{k=1}^{\|\boldsymbol{\beta_1}\|_{r,0}}\sum_{j=1}^{d_k}\beta_{1_j}^2[k]{=}1$, and $\sum_{k'=1}^{\|\boldsymbol{\beta_2}\|_{r,0}} d_{k'}^{-1/q}(\sum_{j=1}^{d_{k'}}|\beta_{2_j}[k']|^q)^{1/q}$ subject to $\sum_{k'=1}^{\|\boldsymbol{\beta_2}\|_{r,0}}\sum_{j=1}^{d_{k'}}\beta_{2_j}^2[k']{=}1$, separately. In order to solve two mentioned problems, first we need to form the Lagrangian function. Then, for the first problem we have:

$$\mathcal{L}\left(\boldsymbol{\beta_1},\lambda\right)=\sum_{k=1}^{\|\boldsymbol{\beta_1}\|_{r,0}} d_k^{\frac{1}{p'}}\left(\sum_{j=1}^{d_k}\left|\beta_{1_j}\left[k\right]\right|^p\right)^{\frac{1}{p}}+\lambda\left(1-\sum_{k=1}^{\|\boldsymbol{\beta_1}\|_{r,0}}\sum_{j=1}^{d_k}\beta_{1_j}^2\left[k\right]\right),$$

then we need to compute its critical point. Considering that $(|f|)'{=}f'\,f/|f|$, where $f'$ is derivative of $f$ with respect to $x$, i.e. $d\,f(x)/d\,x$, we have:

$$\frac{\partial\mathcal{L}}{\partial\beta_{1_j}\left[k\right]}=d_k^{\frac{1}{p'}}\,\beta_{1_j}\left[k\right]\left|\beta_{1_j}\left[k\right]\right|^{p-2}\left(\sum_{j=1}^{d_k}\left|\beta_{1_j}\left[k\right]\right|^p\right)^{\frac{1}{p}-1}-2\lambda\beta_{1_j}\left[k\right]=0$$

$$\Rightarrow\left|\beta_{1_j}\left[k\right]\right|\in\left\{0,\left(\frac{2\lambda}{d_k^{\frac{1}{p'}}\,\|\boldsymbol{\beta_1}\left[k\right]\|_p^{1-p}}\right)^{\frac{1}{p-2}}\right\}.$$

From the above last equality it can be derived that all the absolute value of the coefficients in a block $k$ has the same value. On the other hand, all the identical elements cannot be zero, because it leads to $\|\boldsymbol{\beta_1}[k]\|_{r,0}{=}0$, but from the last line of (A.10), only non-zero blocks are selected for optimisation. Therefore, it reduces to:

$$\left|\beta_{1_j}\left[k\right]\right|=\frac{1}{2\lambda}.$$

Next, applying the unit-energy constraint of the coefficients, we get:

$$1=\sum_{k=1}^{\|\boldsymbol{\beta_1}\|_{r,0}}\sum_{j=1}^{d_k}\beta_{1_j}^2\left[k\right]=\frac{d_k\,\|\boldsymbol{\beta_1}\|_{r,0}}{4\lambda^2}$$

$$\Rightarrow\lambda=\frac{d_k^{\frac{1}{2}}\,\|\boldsymbol{\beta_1}\|_{r,0}^{\frac{1}{2}}}{2}\Rightarrow\left|\beta_{1_j}\left[k\right]\right|=\left(d_k\,\|\boldsymbol{\beta_1}\|_{r,0}\right)^{\frac{-1}{2}}.$$

Therefore, $\sum_{k=1}^{\|\boldsymbol{\beta_1}\|_{r,0}} d_k^{1/p'}(\sum_{j=1}^{d_k}|\beta_{1_j}[k]|^p)^{1/p}$ is upper-bounded by $\|\boldsymbol{\beta_1}\|_{r,0}^{-1/2}\sum_{k=1}^{\|\boldsymbol{\beta_1}\|_{r,0}} d_k^{1/2}$, which is again upper-bounded by $\|\boldsymbol{\beta_1}\|_{r,0}^{1/2}d_{max}^{1/2}$. Similarly, for the second problem it can be proved that $\sum_{k'=1}^{\|\boldsymbol{\beta_2}\|_{r,0}} d_{k'}^{-1/q}(\sum_{j=1}^{d_{k'}}|\beta_{2_j}[k']|^q)^{1/q}$ is upper-bounded by $\|\boldsymbol{\beta_2}\|_{r,0}^{-1/2}\sum_{k'=1}^{\|\boldsymbol{\beta_2}\|_{r,0}} d_{k'}^{-1/2}$, which is upper-bounded by $\|\boldsymbol{\beta_2}\|_{r,0}^{1/2}d_{min}^{-1/2}$. Substituting the recently mentioned upper-bounds into (A.10), we get:

$$\begin{aligned}\forall(q,p)\in\mathbb{R}^2_{>1},\forall r\in\mathbb{R}_{\geq 0},\qquad 1&\leq d_{min}^{-\frac{1}{2}}d_{max}^{\frac{3}{2}}\overline{M}_{q,p'}\left(\boldsymbol{\Phi_1},\boldsymbol{\Phi_2}\right)\left(\|\boldsymbol{\beta_1}\|_{r,0}\,\|\boldsymbol{\beta_2}\|_{r,0}\right)^{\frac{1}{2}}\\&\leq d_{min}^{-\frac{1}{2}}d_{max}^{\frac{3}{2}}\overline{M}_{q,p'}\left(\boldsymbol{\Phi_1},\boldsymbol{\Phi_2}\right)\frac{\|\boldsymbol{\beta_1}\|_{r,0}+\|\boldsymbol{\beta_2}\|_{r,0}}{2}.\end{aligned}$$

The proof is completed in the above equation by replacing $p'$ by $p/(p-1)$, according to the condition of the Hölder's inequality [GL13], and using the inequality of arithmetic-geometric means, i.e., $\sqrt{ab} \leq (a+b)/2$.

Now, consider the case where $q=p=1$. Then, the optimisation problem in (A.11) becomes:

$$\forall r \in \mathbb{R}_{\geq 0}, \qquad \max_{k,k'} \sum_{k=1}^{\|\boldsymbol{\beta_1}\|_{r,0}} \sum_{j=1}^{d_k} \left|\beta_{1_j}[k]\right| \sum_{k'=1}^{\|\boldsymbol{\beta_2}\|_{r,0}} d_{k'}^{-1} \sum_{j=1}^{d_{k'}} \left|\beta_{2_j}\left[k'\right]\right|$$
$$s.t. \quad \sum_{k=1}^{\|\boldsymbol{\beta_1}\|_{r,0}} \sum_{j=1}^{d_k} \beta_{1_j}^2[k] = \sum_{k'=1}^{\|\boldsymbol{\beta_2}\|_{r,0}} \sum_{j=1}^{d_{k'}} \beta_{2_j}^2\left[k'\right] = 1,$$

which is equivalent to

$$\forall r \in \mathbb{R}_{\geq 0}, \qquad \max_{k,k'} \sum_{k=1}^{\|\boldsymbol{\beta_1}\|_{r,0}} \sum_{j=1}^{d_k} \beta_{1_j}[k] \sum_{k'=1}^{\|\boldsymbol{\beta_2}\|_{r,0}} d_{k'}^{-1} \sum_{j=1}^{d_{k'}} \beta_{2_j}\left[k'\right]$$
$$s.t. \quad \sum_{k=1}^{\|\boldsymbol{\beta_1}\|_{r,0}} \sum_{j=1}^{d_k} \beta_{1_j}^2[k] = \sum_{k'=1}^{\|\boldsymbol{\beta_2}\|_{r,0}} \sum_{j=1}^{d_{k'}} \beta_{2_j}^2\left[k'\right] = 1$$
$$and \quad \beta_{1_j}[k], \beta_{2_j}\left[k'\right] > 0.$$

The above optimisation problem is separable, so we need to maximise $\sum_{k=1}^{\|\boldsymbol{\beta_1}\|_{r,0}} \sum_{j=1}^{d_k} \beta_{1_j}[k]$ subject to $\sum_{k=1}^{\|\boldsymbol{\beta_1}\|_{r,0}} \sum_{j=1}^{d_k} \beta_{1_j}^2[k]=1$, $\beta_{1_j}[k]>0$ and $\sum_{k'=1}^{\|\boldsymbol{\beta_2}\|_{r,0}} d_{k'}^{-1} \sum_{j=1}^{d_{k'}} \beta_{2_j}[k']$ subject to $\sum_{k'=1}^{\|\boldsymbol{\beta_2}\|_{r,0}} \sum_{j=1}^{d_{k'}} \beta_{2_j}^2[k']=1$, $\beta_{2_j}[k']>0$, separately. But, in order to solve these problems (following the related proof in [EB02], page 3), let us consider the following Lagrangian function, in which the positivity constraint is not enforced explicitly:

$$\mathcal{L}(\boldsymbol{\beta_1}, \lambda) = \sum_{k=1}^{\|\boldsymbol{\beta_1}\|_{r,0}} \sum_{j=1}^{d_k} \beta_{1_j}[k] + \lambda \left(1 - \sum_{k=1}^{\|\boldsymbol{\beta_1}\|_{r,0}} \sum_{j=1}^{d_k} \beta_{1_j}^2[k]\right)$$
$$\frac{\partial \mathcal{L}}{\partial \beta_{1_j}[k]} = 1 - 2\lambda\beta_{1_j}[k] = 0 \Rightarrow \beta_{1_j}[k] = \frac{1}{2\lambda}.$$

Next, applying the unit-energy constraint of the coefficients, we get:

$$1 = \sum_{k=1}^{\|\boldsymbol{\beta_1}\|_{r,0}} \sum_{j=1}^{d_k} \beta_{1_j}^2[k] = \frac{d_k \|\boldsymbol{\beta_1}\|_{r,0}}{4\lambda^2}$$
$$\Rightarrow \lambda = \frac{d_k^{\frac{1}{2}} \|\boldsymbol{\beta_1}\|_{r,0}^{\frac{1}{2}}}{2} \Rightarrow \beta_{1_j}[k] = \left(d_k \|\boldsymbol{\beta_1}\|_{r,0}\right)^{\frac{-1}{2}}.$$

It can be seen that, although the constraint of $\beta_{1_j}[k]>0$ is not explicitly included in the Lagrangian function, but from the final closed form solution this constraint can be verified. Therefore, $\sum_{k=1}^{\|\boldsymbol{\beta_1}\|_{r,0}} \sum_{j=1}^{d_k} \beta_{1_j}[k]$ is upper-bounded by $\|\boldsymbol{\beta_1}\|_{r,0}^{-1/2} \sum_{k=1}^{\|\boldsymbol{\beta_1}\|_{r,0}} d_k^{1/2}$, which is again upper-bounded by $\|\boldsymbol{\beta_1}\|_{r,0}^{1/2} d_{max}^{1/2}$. Similarly, for the second problem it can be proved

that $\sum_{k'=1}^{\|\boldsymbol{\beta_2}\|_{r,0}} d_{k'}^{-1} \sum_{j=1}^{d_{k'}} \beta_{2_j}[k']$ is upper-bounded by $\|\boldsymbol{\beta_2}\|_{r,0}^{-1/2} \sum_{k'=1}^{\|\boldsymbol{\beta_2}\|_{r,0}} d_{k'}^{-1/2}$, which is again upper-bounded by $\|\boldsymbol{\beta_2}\|_{r,0}^{1/2} d_{min}^{-1/2}$. Substituting the recently mentioned upper-bounds into (A.10), we get the same results as for the previous case of $\forall(q,p)\in\mathbb{R}^2_{>1}$. Hence, the Lemma holds true for $\forall(q,p)\in\mathbb{R}^2_{\geq 1}$. □

## A.9 Proof of Property 2.17 (Basic Block-MCC$_{2,2}$ lower-bound, page 72)

*Proof.* From Property 2.6 (Block-MCC$_{q,p}$ for intra-block orthonormality, page 58), in the special case of two matrices and parameter changing of $p{\rightarrow}p/(p{-}1)$, we have:

$$\forall\,(q,p)\in\mathbb{R}^2_{>0},\qquad \overline{M}_{q,\frac{p}{p-1}}\left(\boldsymbol{\Phi_1},\boldsymbol{\Phi_2}\right)=\max_{k,k'}\frac{d_k^{-\frac{p-1}{p}}\,d_{k'}^{\frac{1}{q}}}{d_{max}}\left\|\boldsymbol{\Phi_1}^T[k]\,\boldsymbol{\Phi_2}\left[k'\right]\right\|_{q\rightarrow\frac{p}{p-1}}.$$

Then, squaring both sides and for $q{=}p{=}2$, we get:

$$\begin{aligned}\overline{M}^2_{2,2}\left(\boldsymbol{\Phi_1},\boldsymbol{\Phi_2}\right)&=\max_{k,k'}\frac{d_k^{-1}d_{k'}}{d^2_{max}}\left\|\boldsymbol{\Phi_1}^T[k]\,\boldsymbol{\Phi_2}\left[k'\right]\right\|^2_{2\rightarrow2}\\&=\max_{k,k'}\frac{d_k^{-1}d_{k'}}{d^2_{max}}\left\|\boldsymbol{\Phi_2}^T\left[k'\right]\boldsymbol{\Phi_1}[k]\,\boldsymbol{\Phi_1}^T[k]\,\boldsymbol{\Phi_2}\left[k'\right]\right\|_{2\rightarrow2}.\end{aligned}$$

The above second equality follows from a property of operator-norms, i.e., $\|\boldsymbol{A}\|^2_{2\rightarrow2}{=}\|\boldsymbol{A}^T\boldsymbol{A}\|_{2\rightarrow2}$. Summing over $k$ and $k'$, we have:

$$\begin{aligned}K^2\overline{M}^2_{2,2}\left(\boldsymbol{\Phi_1},\boldsymbol{\Phi_2}\right)&\geq\sum_{k=1}^{K}\sum_{k'=1}^{K}\frac{d_k^{-1}d_{k'}}{d^2_{max}}\left\|\boldsymbol{\Phi_2}^T\left[k'\right]\boldsymbol{\Phi_1}[k]\,\boldsymbol{\Phi_1}^T[k]\,\boldsymbol{\Phi_2}\left[k'\right]\right\|_{2\rightarrow2}\\&\geq\frac{1}{d^2_{max}}\left\|\sum_{k'=1}^{K}d_{k'}\boldsymbol{\Phi_2}^T\left[k'\right]\left(\sum_{k=1}^{K}d_k^{-1}\boldsymbol{\Phi_1}[k]\,\boldsymbol{\Phi_1}^T[k]\right)\boldsymbol{\Phi_2}\left[k'\right]\right\|_{2\rightarrow2}\\&\geq\frac{d_{min}}{d^3_{max}}\left\|\sum_{k'=1}^{K}\boldsymbol{\Phi_2}^T\left[k'\right]\left(\sum_{k=1}^{K}\boldsymbol{\Phi_1}[k]\,\boldsymbol{\Phi_1}^T[k]\right)\boldsymbol{\Phi_2}\left[k'\right]\right\|_{2\rightarrow2}\\&=\frac{d_{min}}{d^3_{max}}\left\|K\boldsymbol{I}_{d_{k'}}\right\|_{2\rightarrow2}\\&=\frac{d_{min}}{d^3_{max}}K.\end{aligned}$$

The above second inequality follows from the triangle inequality, whereas the third one results from considering the minimum values for variable block length coefficients in the sum operator. In fact, in this problem, having some finite integer values $d_k{\geq}1$, $\forall k$, and the corresponding orthonormal bases $\boldsymbol{\Phi}[k]$, we are utilising the lower-bound $\sum_k d_k\boldsymbol{\Phi}[k]{\geq}\min_k d_k\sum_k\boldsymbol{\Phi}[k]{=}d_{min}\sum_k\boldsymbol{\Phi}[k]$. The above first equality results from the fact that for orthonormal matrices $\boldsymbol{\Phi_1}$ and $\boldsymbol{\Phi_2}$ we have, $\sum_{k=1}^{K}\boldsymbol{\Phi_1}[k]\boldsymbol{\Phi_1}^T[k]{=}\boldsymbol{\Phi_1}\boldsymbol{\Phi_1}^T{=}\boldsymbol{I}_m$ and $\forall k'$, $\boldsymbol{\Phi_2}^T[k']\boldsymbol{\Phi_2}[k']{=}\boldsymbol{I}_{d_{k'}}$. The last equality results from the homogeneity property (Property 2.4, page 56) and unity of operator-norm of the identity matrix. The proof is then completed by taking square roots from both sides and noticing that for each matrix $\boldsymbol{\Phi_1}$ and $\boldsymbol{\Phi_2}$, we have $m{=}\sum_{k=1}^{K}d_k{\geq}\min_k d_k\sum_{k=1}^{K}1{=}K\,d_{min}$, then $K{\leq}m/d_{min}$, i.e.,:

$$\overline{M}_{2,2}\left(\boldsymbol{\Phi_1},\boldsymbol{\Phi_2}\right)\geq\sqrt{\frac{d_{min}}{K\,d^3_{max}}}\geq\sqrt{\frac{d^2_{min}}{m\,d^3_{max}}}.$$

□

## A.10 Proof of Theorem 2.3 (Block-ERC based on Block-MCC$_{q,p}$, page 73)

*Proof.* Because $\boldsymbol{x}\in Ker(\boldsymbol{\Phi})$, we have $\sum_{k'}\boldsymbol{\Phi}[k']\boldsymbol{x}[k']=\boldsymbol{0}$. Hence, for all $k$, $-\sum_{k'\neq k}\boldsymbol{\Phi}[k']\boldsymbol{x}[k']=\boldsymbol{\Phi}[k]\boldsymbol{x}[k]$. Therefore, since for all $k$ the block $\boldsymbol{\Phi}[k]$ is full column rank, we have $-\sum_{k'\neq k}\boldsymbol{\Phi}^{\dagger}[k]\boldsymbol{\Phi}[k']\boldsymbol{x}[k']=\boldsymbol{x}[k]$. Applying $\|\cdot\|_p$ to both sides and using the triangular inequality for $p\geq 1$ and $p=0$, $\sum\|\cdot\|_p\geq\|\sum\cdot\|_p$, we have $\sum_{k'\neq k}\|\boldsymbol{\Phi}^{\dagger}[k]\boldsymbol{\Phi}[k']\boldsymbol{x}[k']\|_p\geq\|\boldsymbol{x}[k]\|_p$. On the other hand, from the definition of mutual coherence constant (Definition 2.6, page 54), i.e., $\forall(q,p)\in\mathbb{R}^2_{>0}$ : $M_{q,p}(\boldsymbol{\Phi})=\max_{\substack{k,k'\neq k\\ \boldsymbol{x}[k']\neq\boldsymbol{0}}} d_k^{-1/p}d_{k'}^{1/q}d_{max}^{-1}\|\boldsymbol{\Phi}^{\dagger}[k]\boldsymbol{\Phi}[k']\boldsymbol{x}[k']\|_p/\|\boldsymbol{x}[k']\|_q$, we can see that in order to compute $M_{q,p}(\boldsymbol{\Phi})$, the value $d_k^{-1/p}d_{k'}^{1/q}d_{max}^{-1}\|\boldsymbol{\Phi}^{\dagger}[k]\boldsymbol{\Phi}[k']\boldsymbol{x}[k']\|_p/\|\boldsymbol{x}[k']\|_q$ is calculated for each $k$ and $k'\neq k$ and finally the maximum calculated value is considered as $M_{q,p}(\boldsymbol{\Phi})$. Then for any $k$ and $k'\neq k$, the $d_k^{-1/p}d_{k'}^{1/q}d_{max}^{-1}\|\boldsymbol{\Phi}^{\dagger}[k]\boldsymbol{\Phi}[k']\boldsymbol{x}[k']\|_p/\|\boldsymbol{x}[k']\|_q$ is upper-bounded by $M_{q,p}(\boldsymbol{\Phi})$, and since $d_k^{1/p}d_{k'}^{-1/q}d_{max}\|\boldsymbol{x}[k']\|_q$ is positive, we have $\|\boldsymbol{\Phi}^{\dagger}[k]\boldsymbol{\Phi}[k']\boldsymbol{x}[k']\|_p\leq d_k^{1/p}d_{k'}^{-1/q}d_{max}\|\boldsymbol{x}[k']\|_q M_{q,p}(\boldsymbol{\Phi})$. Hence, returning back to the proof, we have $\sum_{k'\neq k}d_k^{1/p}d_{k'}^{-1/q}d_{max}\|\boldsymbol{x}[k']\|_q M_{q,p}(\boldsymbol{\Phi})\geq\sum_{k'\neq k}\|\boldsymbol{\Phi}^{\dagger}[k]\boldsymbol{\Phi}[k']\boldsymbol{x}[k']\|_p\geq\|\boldsymbol{x}[k]\|_p$ or $d_{max}M_{q,p}(\boldsymbol{\Phi})\sum_{k'\neq k}d_k^{1/p}d_{k'}^{-1/q}\|\boldsymbol{x}[k']\|_q\geq\|\boldsymbol{x}[k]\|_p$. Then by rearranging the inequality we get $d_{max}M_{q,p}(\boldsymbol{\Phi})\sum_{k'\neq k}\|\boldsymbol{x}[k']\|_q/d_{k'}^{1/q}\geq\|\boldsymbol{x}[k]\|_p/d_k^{1/p}$. Adding $d_{max}M_{q,p}(\boldsymbol{\Phi})\|\boldsymbol{x}[k]\|_q/d_k^{1/q}$ to both sides, we have:

$$\forall q\in\mathbb{R}_{>0},\forall p\in\mathbb{R}_{\geq 1},\qquad d_{max}M_{q,p}\left(\boldsymbol{\Phi}\right)\sum_{k'}\frac{\|\boldsymbol{x}\left[k'\right]\|_q}{d_{k'}^{\frac{1}{q}}}\geq d_{max}M_{q,p}\left(\boldsymbol{\Phi}\right)\frac{\|\boldsymbol{x}\left[k\right]\|_q}{d_k^{\frac{1}{q}}}+\frac{\|\boldsymbol{x}\left[k\right]\|_p}{d_k^{\frac{1}{p}}},$$

which using the definition of weighted (pseudo-)mixed-norm $\ell^{\boldsymbol{w}}_{p_1,p_2}$ (Definition 2.2, page 48), with $w_k=d_k^{-1/q}$, $p_1=q$, and $p_2=1$, it is equivalent to

$$\forall q\in\mathbb{R}_{>0},\forall p\in\mathbb{R}_{\geq 1},\qquad d_{max}M_{q,p}\left(\boldsymbol{\Phi}\right)\|\boldsymbol{x}\|_{\boldsymbol{w};q,1}\geq d_{max}M_{q,p}\left(\boldsymbol{\Phi}\right)\frac{\|\boldsymbol{x}\left[k\right]\|_q}{d_k^{\frac{1}{q}}}+\frac{\|\boldsymbol{x}\left[k\right]\|_p}{d_k^{\frac{1}{p}}}.$$

Now take any vector $\boldsymbol{\beta}$ with Block-Support $S_b(\boldsymbol{\beta})$ such that $S_b(\boldsymbol{\beta_0})\subset S_b(\boldsymbol{\beta})$, then summing over blocks $k\in S_b(\boldsymbol{\beta})$, and knowing that for any constants $\alpha$ and $\forall r\geq 0$, we have $\sum_{k\in S_b(\boldsymbol{\beta})}\alpha=\alpha\|\boldsymbol{\beta}\|_{r,0}$, we get:

$$\begin{aligned}\forall q\in\mathbb{R}_{>0},\forall p\in\mathbb{R}_{\geq 1},\qquad d_{max}M_{q,p}\left(\boldsymbol{\Phi}\right)\|\boldsymbol{x}\|_{\boldsymbol{w};q,1}\|\boldsymbol{\beta}\|_{r,0}&\geq\sum_{k\in S_b(\boldsymbol{\beta})}\left(d_{max}M_{q,p}\left(\boldsymbol{\Phi}\right)\frac{\|\boldsymbol{x}\left[k\right]\|_q}{d_k^{\frac{1}{q}}}+\frac{\|\boldsymbol{x}\left[k\right]\|_p}{d_k^{\frac{1}{p}}}\right)\\&=d_{max}M_{q,p}\left(\boldsymbol{\Phi}\right)\sum_{k\in S_b(\boldsymbol{\beta})}\frac{\|\boldsymbol{x}\left[k\right]\|_q}{d_k^{\frac{1}{q}}}+\sum_{k\in S_b(\boldsymbol{\beta})}\frac{\|\boldsymbol{x}\left[k\right]\|_p}{d_k^{\frac{1}{p}}}.\end{aligned}$$

On the other hand from the proof of Property 2.3 (bounds of two (pseudo-)mixed-norms division, page 55), we have $\forall(q,p)\in\mathbb{R}^2_{>0}$ : $(\|\boldsymbol{x}[k]\|_p/d_k^{1/p})/(\|\boldsymbol{x}[k]\|_q/d_k^{1/q})\geq\min\{1,d_k^{1/q-1/p}\}$. Again using the lower-bound on the fraction of two sums of values explained in Property 2.3, we have $\forall(q,p)\in\mathbb{R}^2_{>0}$ : $(\sum_{k\in S_b(\boldsymbol{\beta})}\|\boldsymbol{x}[k]\|_p/d_k^{1/p})/(\sum_{k\in S_b(\boldsymbol{\beta})}\|\boldsymbol{x}[k]\|_q/d_k^{1/q})\geq\min_{k\in S_b(\boldsymbol{\beta})}\min\{1,d_k^{1/q-1/p}\}$, which is

a special case of Property 2.3. Hence, returning back to the proof, $\forall q{\in}\mathbb{R}_{>0}$, $\forall p{\in}\mathbb{R}_{\geq 1}$, and $\forall r{\in}\mathbb{R}_{\geq 0}$ we get:

$$\begin{aligned} d_{max}M_{q,p}\left(\boldsymbol{\Phi}\right)\left\|\boldsymbol{x}\right\|_{\boldsymbol{w};q,1}\left\|\boldsymbol{\beta}\right\|_{r,0} &\geq d_{max}M_{q,p}\left(\boldsymbol{\Phi}\right)\sum_{k\in S_b(\boldsymbol{\beta})}\frac{\left\|\boldsymbol{x}\left[k\right]\right\|_q}{d_k^{\frac{1}{q}}}+\left(\min_{k\in S_b(\boldsymbol{\beta})}\min\left\{1,d_k^{\frac{1}{q}-\frac{1}{p}}\right\}\right)\sum_{k\in S_b(\boldsymbol{\beta})}\frac{\left\|\boldsymbol{x}\left[k\right]\right\|_q}{d_k^{\frac{1}{q}}} \\ &= \left(d_{max}M_{q,p}\left(\boldsymbol{\Phi}\right)+\min_{k\in S_b(\boldsymbol{\beta})}\min\left\{1,d_k^{\frac{1}{q}-\frac{1}{p}}\right\}\right)\sum_{k\in S_b(\boldsymbol{\beta})}\frac{\left\|\boldsymbol{x}\left[k\right]\right\|_q}{d_k^{\frac{1}{q}}} \\ &\geq \left(d_{max}M_{q,p}\left(\boldsymbol{\Phi}\right)+\min_{k}\min\left\{1,d_k^{\frac{1}{q}-\frac{1}{p}}\right\}\right)\sum_{k\in S_b(\boldsymbol{\beta})}\frac{\left\|\boldsymbol{x}\left[k\right]\right\|_q}{d_k^{\frac{1}{q}}}. \end{aligned}$$

The above last line results from $\min_{k\in S} f(k){\geq}\min_k f(k)$. Then, by dividing the both sides by $\left\|\boldsymbol{x}\right\|_{\boldsymbol{w};q,1}$, $\forall q{\in}\mathbb{R}_{>0}$, $\forall p{\in}\mathbb{R}_{\geq 1}$, and $\forall r{\in}\mathbb{R}_{\geq 0}$ we have:

$$\forall\boldsymbol{x}\in Ker(\boldsymbol{\Phi}),\qquad d_{max}M_{q,p}\left(\boldsymbol{\Phi}\right)\left\|\boldsymbol{\beta}\right\|_{r,0}\geq\left(d_{max}M_{q,p}\left(\boldsymbol{\Phi}\right)+\min_{k}\min\left\{1,d_k^{\frac{1}{q}-\frac{1}{p}}\right\}\right)\frac{\displaystyle\sum_{k\in S_b(\boldsymbol{\beta})}\frac{\left\|\boldsymbol{x}\left[k\right]\right\|_q}{d_k^{\frac{1}{q}}}}{\left\|\boldsymbol{x}\right\|_{\boldsymbol{w};q,1}},$$

or

$$\forall\boldsymbol{x}\in Ker(\boldsymbol{\Phi}),\qquad d_{max}M_{q,p}\left(\boldsymbol{\Phi}\right)\left\|\boldsymbol{\beta}\right\|_{r,0}\left(d_{max}M_{q,p}\left(\boldsymbol{\Phi}\right)+\min_{k}\min\left\{1,d_k^{\frac{1}{q}-\frac{1}{p}}\right\}\right)^{-1}\geq\frac{\displaystyle\sum_{k\in S_b(\boldsymbol{\beta})}\frac{\left\|\boldsymbol{x}\left[k\right]\right\|_q}{d_k^{\frac{1}{q}}}}{\left\|\boldsymbol{x}\right\|_{\boldsymbol{w};q,1}}.$$

Since the above inequality holds true for $\forall\boldsymbol{x}{\in}Ker(\boldsymbol{\Phi})$, it also holds for the maximiser of the right-hand side (because the left-hand side does not depend on $\boldsymbol{x}$). But that maximiser gives us exactly $Q_{\boldsymbol{w};p_1,p_2}(S_b(\boldsymbol{\beta}),\boldsymbol{\Phi})$ by its definition:

$$\begin{aligned} d_{max}M_{q,p}\left(\boldsymbol{\Phi}\right)\left\|\boldsymbol{\beta}\right\|_{r,0}\left(d_{max}M_{q,p}\left(\boldsymbol{\Phi}\right)+\min_{k}\min\left\{1,d_k^{\frac{1}{q}-\frac{1}{p}}\right\}\right)^{-1} &\geq \max_{\boldsymbol{x}\in Ker(\boldsymbol{\Phi})}\frac{\displaystyle\sum_{k\in S_b(\boldsymbol{\beta})}\frac{\left\|\boldsymbol{x}\left[k\right]\right\|_q}{d_k^{\frac{1}{q}}}}{\left\|\boldsymbol{x}\right\|_{\boldsymbol{w};q,1}} \\ &= Q_{\boldsymbol{w};q,1}\left(S_b\left(\boldsymbol{\beta}\right),\boldsymbol{\Phi}\right). \end{aligned}$$

where, $\forall(q,p){\in}\mathbb{R}^2_{\geq 1}$ (constraint on $q$ was imposed by the Block-NSP condition), $\forall r{\in}\mathbb{R}_{\geq 0}$, and $w_k{=}d_k^{-1/q}$. The above last line is obtained considering the definition of $Q_{\boldsymbol{w};p_1,p_2}(S_b(\boldsymbol{\beta}),\boldsymbol{\Phi})$ in Block-NSP (Theorem 2.2, page 67). On the other hand, from the Block-NSP condition in Theorem 2.2, we have, if $Q_{\boldsymbol{w};q,1}(S_b(\boldsymbol{\beta}),\boldsymbol{\Phi}){<}1/2$ then a solution $\boldsymbol{\beta_0}$ to the problem $P_{\boldsymbol{w};q,1}$ is the unique solution whenever $S_b(\boldsymbol{\beta_0}){\subset}S_b(\boldsymbol{\beta})$. Hence, we have that if:

$$\forall(q,p)\in\mathbb{R}^2_{\geq 1},\forall r\in\mathbb{R}_{\geq 0},\qquad d_{max}M_{q,p}\left(\boldsymbol{\Phi}\right)\left\|\boldsymbol{\beta}\right\|_{r,0}\left(d_{max}M_{q,p}\left(\boldsymbol{\Phi}\right)+\min_{k}\min\left\{1,d_k^{\frac{1}{q}-\frac{1}{p}}\right\}\right)^{-1}<\frac{1}{2},$$

Then, the Block-NSP condition of Theorem 2.2 is met, hence any $\boldsymbol{\beta_0}$ solution to $\boldsymbol{y}{=}\boldsymbol{\Phi\beta}$ is unique if $S_b(\boldsymbol{\beta_0}){\subset}S_b(\boldsymbol{\beta})$. Now, by rearranging the above inequality, we have

$$\forall(q,p)\in\mathbb{R}^2_{\geq 1},\forall r\in\mathbb{R}_{\geq 0},\qquad \left\|\boldsymbol{\beta}\right\|_{r,0}<\frac{1+\left(d_{max}M_{q,p}\left(\boldsymbol{\Phi}\right)\right)^{-1}\min_{k}\min\left\{1,d_k^{\frac{1}{q}-\frac{1}{p}}\right\}}{2},$$

and since $S_b(\boldsymbol{\beta_0}) \subset S_b(\boldsymbol{\beta})$, then $\|\boldsymbol{\beta_0}\|_{r,0} \leq \|\boldsymbol{\beta}\|_{r,0}$, so

$$\forall (q,p) \in \mathbb{R}^2_{\geq 1}, \forall r \in \mathbb{R}_{\geq 0}, \qquad \|\boldsymbol{\beta_0}\|_{r,0} < \frac{1 + \left(d_{max} M_{q,p}\left(\boldsymbol{\Phi}\right)\right)^{-1} \min\limits_{k} \min\left\{1, d_k^{\frac{1}{q}-\frac{1}{p}}\right\}}{2},$$

which is exactly the condition of the theorem. Therefore, the Block-NSP theory under the mentioned condition is actually met and the unique solution of the problem $P_{\boldsymbol{w};q,1}$ is ensured. □

## A.11 Proof of Property 2.18 (Block-SL$_{q,p}$ inequalities, page 73)

*Proof.* First, considering that for all basic tractable $(q,p)$ pairs of table 2.1 (page 54), we have $q{\leq}p$, then the term $\min_k \min\{1, d_k^{1/q-1/p}\}$ in Block-SL$_{q,p}$ of Theorem 2.3 (page 73) would be equal to one. Hence, Block-SL$_{q,p}$ has directly an inverse relationship with Block-MCC$_{q,p}$ (Definition 2.6, page 54), i.e., for $1{\leq}q{\leq}p$, $Block{-}SL_{q,p}(\mathbf{\Phi}){=}(1{+}(d_{max}M_{q,p}(\mathbf{\Phi}))^{-1})/2$. Thus, according to the previously-mentioned relationship between different Block-MCC$_{q,p}$ characterisations with basic $(q,p)$ pairs according to table 2.1 (page 54) demonstrated in Property 2.5 (Block-MCC$_{q,p}$ inequalities, page 58), the Block-SL$_{q,p}$ inequalities would be in the opposite direction, which proves the inequalities of the Property. Therefore, for $q{\leq}p$ the directions in figure 2.5(a) (page 73) is in the opposite of the ones in figure 2.4 (page 59).

For the proof of general relationships (tractable and intractable) of figure 2.5(b), we have the following Block-SL$_{q,p}$ for $q{>}p$, and $d_1{=}\cdots{=}d_K{=}d$:

$$\begin{aligned}
\forall (q,p) \in \mathbb{R}^2_{\geq 1}, q > p, \qquad Block{-}SL_{q,p}(\mathbf{\Phi}) &= \frac{1 + \left(d\, M_{q,p}\left(\mathbf{\Phi}\right)\right)^{-1} d^{\frac{1}{q}-\frac{1}{p}}}{2} \\
&= \frac{1 + \left(d^{1+\frac{1}{p}-\frac{1}{q}} M_{q,p}\left(\mathbf{\Phi}\right)\right)^{-1}}{2} \\
&= \frac{1 + \left(d^{1+\frac{1}{p}-\frac{1}{q}} \max\limits_{k,k' \neq k} d^{-1-\frac{1}{p}+\frac{1}{q}} \left\| \mathbf{\Phi}^{\dagger}\left[k\right] \mathbf{\Phi}\left[k'\right] \right\|_{q\to p}\right)^{-1}}{2} \\
&= \frac{1 + \left(\max\limits_{k,k' \neq k} \left\| \mathbf{\Phi}^{\dagger}\left[k\right] \mathbf{\Phi}\left[k'\right] \right\|_{q\to p}\right)^{-1}}{2}.
\end{aligned} \tag{A.12}$$

On the other hand, for $q{=}p$, and $d_1{=}\cdots{=}d_K{=}d$, we get the following Block-SL$_{q,p}$:

$$\begin{aligned}
\forall q \in \mathbb{R}_{\geq 1}, \qquad Block{-}SL_{q,q}\left(\mathbf{\Phi}\right) &= \frac{1 + \left(d\, M_{q,q}\left(\mathbf{\Phi}\right)\right)^{-1}}{2} \\
&= \frac{1 + \left(d \max\limits_{k,k' \neq k} d^{-1} \left\| \mathbf{\Phi}^{\dagger}\left[k\right] \mathbf{\Phi}\left[k'\right] \right\|_{q\to q}\right)^{-1}}{2} \\
&= \frac{1 + \left(\max\limits_{k,k' \neq k} \left\| \mathbf{\Phi}^{\dagger}\left[k\right] \mathbf{\Phi}\left[k'\right] \right\|_{q\to q}\right)^{-1}}{2}.
\end{aligned} \tag{A.13}$$

By comparing the Block-SL$_{q,p}$ in (A.12) for $p{>}q$, with (A.13) for $q{=}p$, we see that the only difference between the equations is the $\ell_{q\to p}$ operator-norm. On the other hand, from figure 2.3, we know that for a fixed $p$ by increasing $q$, and for a fixed $q$ by decreasing $p$, the $\ell_{q\to p}$ operator-norm increases, hence, Block-SL$_{q,p}$ decreases, which results in the figure 2.5(b). □

## A.12 Proof of Property 2.19 (SL v.s. Block-SL$_{q,p}$, page 74)

*Proof.* Starting from the relation between $M(\mathbf{\Phi})$ and $M_{q,p}(\mathbf{\Phi})$ proposed in Property 2.8 (Block-MCC$_{q,p}$ bounds, page 59), the aim is to build the $d_{max}$ times of the numerator of the condition of the Block-ERC based on Block-MCC$_{q,p}$ in Theorem 2.3, i.e., $d_{max}+M_{q,p}^{-1}(\mathbf{\Phi})\min_k \min\{1, d_k^{1/q-1/p}\}$, which is comparable with the conventional sparsity level. First, we investigate the required condition on $M(\mathbf{\Phi})$ for a dictionary with full column rank blocks. Using the first part of Property 2.8 (Block-MCC$_{q,p}$ bounds, page 59) with conditions $M(\mathbf{\Phi})<d_{max}^{1/q-2}(d_{max}-1)^{-1/2}$ and $q\geq p\geq 1$, and then multiplying by the positive coefficient $\min_k \min\{1, d_k^{1/q-1/p}\}$ and summing to $d_{max}$, for $q\geq p\geq 1$ we get:

$$M_{q,p}^{-1}(\mathbf{\Phi}) \geq \frac{1-d_{max}^{2-\frac{1}{q}}(d_{max}-1)^{\frac{1}{2}}M(\mathbf{\Phi})}{d_{max}^{\frac{3}{2}-\frac{1}{p}}M(\mathbf{\Phi})},$$

$$M_{q,p}^{-1}(\mathbf{\Phi})\min_k \min\left\{1, d_k^{\frac{1}{q}-\frac{1}{p}}\right\} \geq \frac{1-d_{max}^{2-\frac{1}{q}}(d_{max}-1)^{\frac{1}{2}}M(\mathbf{\Phi})}{d_{max}^{\frac{3}{2}-\frac{1}{p}}M(\mathbf{\Phi})}\min_k \min\left\{1, d_k^{\frac{1}{q}-\frac{1}{p}}\right\},$$

$$d_{max}+M_{q,p}^{-1}(\mathbf{\Phi})\min_k \min\left\{1, d_k^{\frac{1}{q}-\frac{1}{p}}\right\} \geq d_{max}+\frac{1-d_{max}^{2-\frac{1}{q}}(d_{max}-1)^{\frac{1}{2}}M(\mathbf{\Phi})}{d_{max}^{\frac{3}{2}-\frac{1}{p}}M(\mathbf{\Phi})}\min_k \min\left\{1, d_k^{\frac{1}{q}-\frac{1}{p}}\right\}.$$

But, $\forall q\geq p$, we have $\min_k \min\{1, d_k^{1/q-1/p}\}=d_{max}^{1/q-1/p}$, which leads to the following inequality:

$$\begin{aligned} q\geq p\geq 1, \qquad d_{max}+d_{max}^{\frac{1}{q}-\frac{1}{p}}M_{q,p}^{-1}(\mathbf{\Phi}) &\geq d_{max}+\frac{1-d_{max}^{2-\frac{1}{q}}(d_{max}-1)^{\frac{1}{2}}M(\mathbf{\Phi})}{d_{max}^{\frac{3}{2}-\frac{1}{p}}M(\mathbf{\Phi})}d_{max}^{\frac{1}{q}-\frac{1}{p}} \\ &= d_{max}-d_{max}^{\frac{1}{2}}(d_{max}-1)^{\frac{1}{2}}+d_{max}^{\frac{1}{q}-\frac{3}{2}}M^{-1}(\mathbf{\Phi}). \end{aligned}$$

The above right-hand side expression under the following condition would be greater than or equal to $1+M^{-1}(\mathbf{\Phi})$:

$$\begin{aligned} &d_{max}-d_{max}^{\frac{1}{2}}(d_{max}-1)^{\frac{1}{2}}+d_{max}^{\frac{1}{q}-\frac{3}{2}}M^{-1}(\mathbf{\Phi}) \geq 1+M^{-1}(\mathbf{\Phi}), \\ &\Rightarrow M(\mathbf{\Phi}) \leq \frac{1-d_{max}^{\frac{1}{q}-\frac{3}{2}}}{(d_{max}-1)^{\frac{1}{2}}\left[(d_{max}-1)^{\frac{1}{2}}-d_{max}^{\frac{1}{2}}\right]}. \end{aligned}$$

Therefore, considering another condition of the utilised property, i.e., $M(\mathbf{\Phi})<d_{max}^{1/q-2}(d_{max}-1)^{-1/2}$ in Property 2.8 (Block-MCC$_{q,p}$ bounds, page 59), the following upper-bound condition on $M(\mathbf{\Phi})$ ensures that the proposed Block-SL$_{q,p}$ would be greater than or equal to the conventional SL, i.e., $d_{max}+M_{q,p}^{-1}(\mathbf{\Phi})\min_k \min\{1, d_k^{1/q-1/p}\}\geq 1+M^{-1}(\mathbf{\Phi})$:

$$q\geq p\geq 1, \qquad M(\mathbf{\Phi}) \leq \min\left\{\frac{d_{max}^{\frac{1}{q}-2}}{(d_{max}-1)^{\frac{1}{2}}}, \frac{1-d_{max}^{\frac{1}{q}-\frac{3}{2}}}{(d_{max}-1)^{\frac{1}{2}}\left[(d_{max}-1)^{\frac{1}{2}}-d_{max}^{\frac{1}{2}}\right]}\right\}.$$

But, the above second argument of the min operator $\forall q, d_{max}$ is negative. Therefore, although in practice the numerical simulations approve that the proposed sparsity level for general dictionaries with full column rank blocks is higher than the conventional one, in theory due to utilising the most pessimistic bounds, its supremacy cannot be proved.

Now, suppose that the dictionary $\mathbf{\Phi}$ has intra-block orthonormality, which is the case in the Property 2.19 (SL v.s. Block-SL$_{q,p}$, page 74). Using the second part of Property 2.8 (Block-MCC$_{q,p}$ bounds, page 59) as the following first line, multiplying by $\min_k \min\{1, d_k^{1/q-1/p}\}$ and summing to $d_{max}$, $\forall (q,p,q',p') \in \mathbb{R}^4_{>0}$ we get:

$$
\begin{aligned}
M_{q,p}^{-1}(\mathbf{\Phi}) &\geq \frac{d_{max}\, M^{-1}(\mathbf{\Phi})}{\max\limits_{k,k'\neq k} d_k^{\frac{1}{2}-\frac{1}{p}} d_{k'}^{\frac{1}{q}+\frac{1}{2}} \max\left\{1, d_k^{\frac{1}{p}-\frac{1}{p'}}\right\} \max\left\{1, d_{k'}^{\frac{1}{q'}-\frac{1}{q}}\right\} \max\left\{1, d_k^{\frac{1}{p'}-\frac{1}{2}}\right\} \max\left\{1, d_{k'}^{\frac{1}{2}-\frac{1}{q'}}\right\}}, \\
M_{q,p}^{-1}(\mathbf{\Phi}) \min_k \min\left\{1, d_k^{\frac{1}{q}-\frac{1}{p}}\right\} &\geq \frac{d_{max}\, M^{-1}(\mathbf{\Phi})}{\max\limits_{k,k'\neq k} d_k^{\frac{1}{2}-\frac{1}{p}} d_{k'}^{\frac{1}{q}+\frac{1}{2}} \max\left\{1, d_k^{\frac{1}{p}-\frac{1}{p'}}\right\} \max\left\{1, d_{k'}^{\frac{1}{q'}-\frac{1}{q}}\right\} \max\left\{1, d_k^{\frac{1}{p'}-\frac{1}{2}}\right\} \max\left\{1, d_{k'}^{\frac{1}{2}-\frac{1}{q'}}\right\}} \min_k \min\left\{1, d_k^{\frac{1}{q}-\frac{1}{p}}\right\}, \\
d_{max} + M_{q,p}^{-1}(\mathbf{\Phi}) \min_k \min\left\{1, d_k^{\frac{1}{q}-\frac{1}{p}}\right\} &\geq d_{max} + \frac{d_{max}\, M^{-1}(\mathbf{\Phi})}{\max\limits_{k,k'\neq k} d_k^{\frac{1}{2}-\frac{1}{p}} d_{k'}^{\frac{1}{q}+\frac{1}{2}} \max\left\{1, d_k^{\frac{1}{p}-\frac{1}{p'}}\right\} \max\left\{1, d_{k'}^{\frac{1}{q'}-\frac{1}{q}}\right\} \max\left\{1, d_k^{\frac{1}{p'}-\frac{1}{2}}\right\} \max\left\{1, d_{k'}^{\frac{1}{2}-\frac{1}{q'}}\right\}} \min_k \min\left\{1, d_k^{\frac{1}{q}-\frac{1}{p}}\right\}.
\end{aligned}
$$

The above right-hand side expression under the following condition and $\forall (q,p,q',p') \in \mathbb{R}^4_{>0}$ would be greater than or equal to $1 + M^{-1}(\mathbf{\Phi})$:

$$
\begin{aligned}
d_{max} + \frac{d_{max}\, M^{-1}(\mathbf{\Phi})}{\max\limits_{k,k'\neq k} d_k^{\frac{1}{2}-\frac{1}{p}} d_{k'}^{\frac{1}{q}+\frac{1}{2}} \max\left\{1, d_k^{\frac{1}{p}-\frac{1}{p'}}\right\} \max\left\{1, d_{k'}^{\frac{1}{q'}-\frac{1}{q}}\right\} \max\left\{1, d_k^{\frac{1}{p'}-\frac{1}{2}}\right\} \max\left\{1, d_{k'}^{\frac{1}{2}-\frac{1}{q'}}\right\}} \min_k \min\left\{1, d_k^{\frac{1}{q}-\frac{1}{p}}\right\} &\geq 1 + M^{-1}(\mathbf{\Phi}), \\
\Rightarrow M(\mathbf{\Phi}) \leq \frac{1 - \dfrac{d_{max} \min\limits_k \min\left\{1, d_k^{\frac{1}{q}-\frac{1}{p}}\right\}}{\max\limits_{k,k'\neq k} d_k^{\frac{1}{2}-\frac{1}{p}} d_{k'}^{\frac{1}{q}+\frac{1}{2}} \max\left\{1, d_k^{\frac{1}{p}-\frac{1}{p'}}\right\} \max\left\{1, d_{k'}^{\frac{1}{q'}-\frac{1}{q}}\right\} \max\left\{1, d_k^{\frac{1}{p'}-\frac{1}{2}}\right\} \max\left\{1, d_{k'}^{\frac{1}{2}-\frac{1}{q'}}\right\}}}{d_{max} - 1}.
\end{aligned}
$$

□

## A.13 Proof of Lemma 2.3 (Eldar et al.'s v.s. proposed Block-SL$_{q,p}$, page 75)

*Proof.* The proof is similar to the proof of Property 2.8 (Block-MCC$_{q,p}$ bounds). First, we investigate the required condition on $M_{Intra}^{Eldar}(\mathbf{\Phi})$ for a dictionary with full column rank blocks. Utilising the Block-MCC$_{q,p}$ proposed in Definition 2.6 (54), we have the following first line in (A.14). Using the pseudo-inverse property of full column rank matrices and submultiplicativity property of operator-norms introduced in Property 2.4 ($\ell_{q\to p}$ operator-norm properties), and for $d_1{=}\cdots{=}d_K{=}d$, $\forall(q,p){\in}\mathbb{R}^2_{>0}$ we have:

$$\begin{aligned} M_{q,p}\left(\mathbf{\Phi}\right) &= d^{\frac{1}{q}-\frac{1}{p}-1} \max_{k,k'\neq k} \left\| \mathbf{\Phi}^{\dagger}\left[k\right] \mathbf{\Phi}\left[k'\right] \right\|_{q\to p} \\ &= d^{\frac{1}{q}-\frac{1}{p}-1} \max_{k,k'\neq k} \left\| \left(\mathbf{\Phi}^{T}\left[k\right] \mathbf{\Phi}\left[k\right]\right)^{-1} \mathbf{\Phi}^{T}\left[k\right] \mathbf{\Phi}\left[k'\right] \right\|_{q\to p} \\ &\leq d^{\frac{1}{q}-\frac{1}{p}-1} \max_{k,k'\neq k} \left\| \left(\mathbf{\Phi}^{T}\left[k\right] \mathbf{\Phi}\left[k\right]\right)^{-1} \right\|_{q\to p} \left\| \mathbf{\Phi}^{T}\left[k\right] \mathbf{\Phi}\left[k'\right] \right\|_{q\to p} \max\left\{1, d^{\frac{1}{q}-\frac{1}{p}}\right\}. \end{aligned} \tag{A.14}$$

Similar to the proof of the Property 2.8 in finding upper-bound of $\|(\mathbf{\Phi}^T[k]\mathbf{\Phi}[k])^{-1}\|_{q\to p}$, by replacing $M(\mathbf{\Phi})$ by $M_{Intra}^{Eldar}(\mathbf{\Phi})$ (because $\forall i,j$, $\boldsymbol{F}_{i,j}[k]{\leq}M_{Intra}^{Eldar}(\mathbf{\Phi}){\stackrel{\text{def}}{=}}\max_{i,j\neq i}|\boldsymbol{\varphi}_i^T[k]\boldsymbol{\varphi}_j[k]|{\leq}M(\mathbf{\Phi})$), and using Property A.2 ($q{\geq}p{\geq}1$, $M_{Intra}^{Eldar}(\mathbf{\Phi}){<}d^{1/q-2}(d-1)^{-1/2}$ to meet the related condition), we have:

$$\forall k, q\geq p\geq 1, \qquad \left\| \left(\mathbf{\Phi}^{T}\left[k\right] \mathbf{\Phi}\left[k\right]\right)^{-1} \right\|_{q\to p} \leq \frac{1}{1-d^{2-\frac{1}{q}}\left(d-1\right)^{\frac{1}{2}} M_{Intra}^{Eldar}\left(\mathbf{\Phi}\right)}.$$

For $q{\geq}p{\geq}1$ (condition imposed by Property A.2), we have $\max\{1, d^{1/q-1/p}\}{=}1$ in (A.14). Therefore, by substituting the upper-bound of $\|(\mathbf{\Phi}^T[k]\mathbf{\Phi}[k])^{-1}\|_{q\to p}$ and the value of $\max\{1, d^{1/q-1/p}\}$ in (A.14), for $q{\geq}p{\geq}1$ we get:

$$\begin{aligned} M_{q,p}\left(\mathbf{\Phi}\right) &\leq \frac{d^{\frac{1}{q}-\frac{1}{p}-1}}{1-d^{2-\frac{1}{q}}\left(d-1\right)^{\frac{1}{2}} M_{Intra}^{Eldar}\left(\mathbf{\Phi}\right)} \max_{k,k'\neq k} \left\| \mathbf{\Phi}^{T}\left[k\right] \mathbf{\Phi}\left[k'\right] \right\|_{q\to p} \\ &\leq \frac{d^{\frac{1}{q}-\frac{1}{p}-1} \max\left\{1, d^{\frac{1}{p}-\frac{1}{2}}\right\} \max\left\{1, d^{\frac{1}{2}-\frac{1}{q}}\right\}}{1-d^{2-\frac{1}{q}}\left(d-1\right)^{\frac{1}{2}} M_{Intra}^{Eldar}\left(\mathbf{\Phi}\right)} \max_{k,k'\neq k} \left\| \mathbf{\Phi}^{T}\left[k\right] \mathbf{\Phi}\left[k'\right] \right\|_{2\to 2}. \end{aligned}$$

The above last inequality is achieved based on the upper-bound of $\ell_{q\to p}$ operator-norm based on $\ell_{2\to 2}$ operator-norm explained in Property A.1. Then by definition, substituting $\max_{k,k'\neq k}\|\mathbf{\Phi}^T[k]\mathbf{\Phi}[k']\|_{2\to 2}$ by $d\,M_{Inter}^{Eldar}(\mathbf{\Phi})$ in the above inequality, for $q{\geq}p{\geq}1$ we have:

$$M_{q,p}\left(\mathbf{\Phi}\right) \leq \frac{d^{\frac{1}{q}-\frac{1}{p}} M_{Inter}^{Eldar}\left(\mathbf{\Phi}\right) \max\left\{1, d^{\frac{1}{p}-\frac{1}{2}}\right\} \max\left\{1, d^{\frac{1}{2}-\frac{1}{q}}\right\}}{1-d^{2-\frac{1}{q}}\left(d-1\right)^{\frac{1}{2}} M_{Intra}^{Eldar}\left(\mathbf{\Phi}\right)},$$

or,

$$M_{q,p}^{-1}\left(\mathbf{\Phi}\right) \geq \frac{1-d^{2-\frac{1}{q}}\left(d-1\right)^{\frac{1}{2}} M_{Intra}^{Eldar}\left(\mathbf{\Phi}\right)}{d^{\frac{1}{q}-\frac{1}{p}} M_{Inter}^{Eldar}\left(\mathbf{\Phi}\right) \max\left\{1, d^{\frac{1}{p}-\frac{1}{2}}\right\} \max\left\{1, d^{\frac{1}{2}-\frac{1}{q}}\right\}}.$$

By comparing Eldar's Block-ERC, i.e., $\|\boldsymbol{\beta_0}\|_{2,0}<(1+(dM_{Inter}^{Eldar}(\boldsymbol{\Phi}))^{-1}(1-(d-1)M_{Intra}^{Eldar}(\boldsymbol{\Phi})))/2$, explained in (1.20) and the proposed equally-sized Block-ERC proposed in Theorem 2.3, i.e., $\forall(q,p)\in\mathbb{R}^2_{\geq 1}, \forall r\in\mathbb{R}_{\geq 0}\ :\ \|\boldsymbol{\beta_0}\|_{r,0}<(1+(dM_{q,p}(\boldsymbol{\Phi}))^{-1}\min\{1,d^{1/q-1/p}\})/2$, it is clear that for $q\geq p\geq 1$ (the constraint imposed up to this step) the relationship between $d^{1/q-1/p}M_{q,p}^{-1}(\boldsymbol{\Phi})$ and $(M_{Inter}^{Eldar}(\boldsymbol{\Phi}))^{-1}(1-(d-1)M_{Intra}^{Eldar}(\boldsymbol{\Phi}))$ should be investigated. By multiplying both sides of the above inequality by $d^{1/q-1/p}>0$, for $q\geq p\geq 1$ we have:

$$d^{\frac{1}{q}-\frac{1}{p}}M_{q,p}^{-1}\left(\boldsymbol{\Phi}\right)\geq\frac{1-d^{2-\frac{1}{q}}\left(d-1\right)^{\frac{1}{2}}M_{Intra}^{Eldar}\left(\boldsymbol{\Phi}\right)}{M_{Inter}^{Eldar}\left(\boldsymbol{\Phi}\right)\max\left\{1,d^{\frac{1}{p}-\frac{1}{2}}\right\}\max\left\{1,d^{\frac{1}{2}-\frac{1}{q}}\right\}}.$$

Assuming that the above right-hand side is greater than or equal to the expression that we are interested in to its relation with $d^{1/q-1/p}M_{q,p}^{-1}(\boldsymbol{\Phi})$, i.e., $(M_{Inter}^{Eldar}(\boldsymbol{\Phi}))^{-1}(1-(d-1)M_{Intra}^{Eldar}(\boldsymbol{\Phi}))$, we extract the required corresponding condition. In other words, assuming that for $q\geq p\geq 1$, we have:

$$\frac{1-d^{2-\frac{1}{q}}\left(d-1\right)^{\frac{1}{2}}M_{Intra}^{Eldar}\left(\boldsymbol{\Phi}\right)}{M_{Inter}^{Eldar}\left(\boldsymbol{\Phi}\right)\max\left\{1,d^{\frac{1}{p}-\frac{1}{2}}\right\}\max\left\{1,d^{\frac{1}{2}-\frac{1}{q}}\right\}}\geq\frac{1-\left(d-1\right)M_{Intra}^{Eldar}\left(\boldsymbol{\Phi}\right)}{M_{Inter}^{Eldar}\left(\boldsymbol{\Phi}\right)},$$

we conclude that for

$$q\geq p\geq 1,\qquad M_{Intra}^{Eldar}\left(\boldsymbol{\Phi}\right)\leq\frac{1-\max\left\{1,d^{\frac{1}{p}-\frac{1}{2}}\right\}\max\left\{1,d^{\frac{1}{2}-\frac{1}{q}}\right\}}{d^{2-\frac{1}{q}}\left(d-1\right)^{\frac{1}{2}}-\left(d-1\right)\max\left\{1,d^{\frac{1}{p}-\frac{1}{2}}\right\}\max\left\{1,d^{\frac{1}{2}-\frac{1}{q}}\right\}},$$

the assumption will hold true. On the other hand, we had another condition on the upper-bound of $M_{Intra}^{Eldar}(\boldsymbol{\Phi})$ to meet the condition of Property A.2. Therefore, the minimum of the obtained upper-bounds on the $M_{Intra}^{Eldar}(\boldsymbol{\Phi})$ ensures that $\forall q\geq p\geq 1$ the Block-ERC proposed in Theorem 2.3 improves Eldar's Block-ERC, i.e., $d^{1/q-1/p}M_{q,p}^{-1}(\boldsymbol{\Phi})\geq(M_{Inter}^{Eldar}(\boldsymbol{\Phi}))^{-1}(1-(d-1)M_{Intra}^{Eldar}(\boldsymbol{\Phi}))$:

$$\begin{aligned}M_{Intra}^{Eldar}\left(\boldsymbol{\Phi}\right)&\leq\min\left\{\frac{1}{d^{2-\frac{1}{q}}\left(d-1\right)^{\frac{1}{2}}},\frac{1-\max\left\{1,d^{\frac{1}{p}-\frac{1}{2}}\right\}\max\left\{1,d^{\frac{1}{2}-\frac{1}{q}}\right\}}{d^{2-\frac{1}{q}}\left(d-1\right)^{\frac{1}{2}}-\left(d-1\right)\max\left\{1,d^{\frac{1}{p}-\frac{1}{2}}\right\}\max\left\{1,d^{\frac{1}{2}-\frac{1}{q}}\right\}}\right\}\\&=\frac{1-\max\left\{1,d^{\frac{1}{p}-\frac{1}{2}}\right\}\max\left\{1,d^{\frac{1}{2}-\frac{1}{q}}\right\}}{d^{2-\frac{1}{q}}\left(d-1\right)^{\frac{1}{2}}-\left(d-1\right)\max\left\{1,d^{\frac{1}{p}-\frac{1}{2}}\right\}\max\left\{1,d^{\frac{1}{2}-\frac{1}{q}}\right\}}.\end{aligned}$$

The above last inequality yields from the fact that a fraction with smaller numerator and denominator is smaller. But, the resulted upper-bound is negative. Therefore, although in practice the numerical simulations approve that the proposed block-sparsity level for general dictionaries with full column rank blocks is higher than the block-sparsity level proposed by Eldar et al., in theory due to utilising the most pessimistic bounds, its supremacy cannot be proved.

Now let us consider the assumption of the lemma, i.e., intra-block orthonormality. By definition, $M_{Intra}^{Eldar}(\boldsymbol{\Phi})$ is equal to zero in this case. Therefore, according to Theorem 2.3, the

| $(q,p)$ | $(1,1)$ | $(1,2)$ | $(1,\infty)$ | $(2,2)$ | $(2,\infty)$ | $(\infty,\infty)$ |
|---|---|---|---|---|---|---|
| $\frac{M_{q,p}^{-1}(\mathbf{\Phi})\ \min\left\{1,d^{\frac{1}{q}-\frac{1}{p}}\right\}}{\left(M_{Inter}^{Eldar}(\mathbf{\Phi})\right)^{-1}} \geq$ | $d^{-\frac{1}{2}}$ | $d^{-\frac{1}{2}}$ | $d^{-1}$ | $1$ | $d^{-\frac{1}{2}}$ | $d^{-\frac{1}{2}}$ |

Table A.3: Lower-bound of $M_{q,p}^{-1}(\mathbf{\Phi})\ \min\{1,d^{1/q-1/p}\}/(M_{Inter}^{Eldar}(\mathbf{\Phi}))^{-1}$ ensuring the supremacy of the proposed sparsity level when it is greater than one, for different basic values of $(q,p)$ pairs and for a dictionary with intra-block orthonormality.

relationship between $\forall(q,p)\in\mathbb{R}^2_{\geq 1}$ : $M_{q,p}^{-1}(\mathbf{\Phi})\min\{1,d^{1/q-1/p}\}$ and $(M_{Inter}^{Eldar}(\mathbf{\Phi}))^{-1}$ should be investigated. Starting from the relationship between $M_{q,p}(\mathbf{\Phi})$ and $M_{Inter}^{Eldar}(\mathbf{\Phi})$ explained in Property 2.11, we build the required relationship $\forall(q,p)\in\mathbb{R}^2_{>0}$:

$$
\begin{aligned}
M_{q,p}^{-1}\left(\mathbf{\Phi}\right) &\geq \frac{d^{\frac{1}{p}-\frac{1}{q}}\left(M_{Inter}^{Eldar}\left(\mathbf{\Phi}\right)\right)^{-1}}{\max\left\{1,d^{\frac{1}{p}-\frac{1}{2}}\right\}\max\left\{1,d^{\frac{1}{2}-\frac{1}{q}}\right\}},\\
M_{q,p}^{-1}\left(\mathbf{\Phi}\right)\ \min\left\{1,d^{\frac{1}{q}-\frac{1}{p}}\right\} &\geq \frac{d^{\frac{1}{p}-\frac{1}{q}}\left(M_{Inter}^{Eldar}\left(\mathbf{\Phi}\right)\right)^{-1}\min\left\{1,d^{\frac{1}{q}-\frac{1}{p}}\right\}}{\max\left\{1,d^{\frac{1}{p}-\frac{1}{2}}\right\}\max\left\{1,d^{\frac{1}{2}-\frac{1}{q}}\right\}},\\
\frac{M_{q,p}^{-1}\left(\mathbf{\Phi}\right)\ \min\left\{1,d^{\frac{1}{q}-\frac{1}{p}}\right\}}{\left(M_{Inter}^{Eldar}\left(\mathbf{\Phi}\right)\right)^{-1}} &\geq \frac{\min\left\{1,d^{\frac{1}{p}-\frac{1}{q}}\right\}}{\max\left\{1,d^{\frac{1}{p}-\frac{1}{2}}\right\}\max\left\{1,d^{\frac{1}{2}-\frac{1}{q}}\right\}}.
\end{aligned}
$$

The numerator and denominator of the above left-hand side are the values that we need to compare with each other. To enhance Eldar's Block-ERC the above right-hand side should be greater than one. Those right-hand side values are shown in table A.3 for different basic values of $(q,p)$ pairs. From table A.3, it can be seen that for $q{=}p{=}2$, $M_{q,p}^{-1}(\mathbf{\Phi})\ \min\{1,d^{1/q-1/p}\}$ is greater than or equal to $(M_{Inter}^{Eldar}(\mathbf{\Phi}))^{-1}$, which is in fact just equal (not greater), because according to Property 2.11 for dictionaries with intra-block orthonormality, we have $M_{2,2}(\mathbf{\Phi}){=}M_{Inter}^{Eldar}(\mathbf{\Phi})$. Therefore, the proposed Block-ERC for $q{=}p{=}2$ as the best case equals to Eldar's Block-ERC. □

## A.14 Proof of Theorem 2.4 (Block-ERC based on Eldar's cumulative coherence, page 76)

*Proof.* The proof is similar to the proof of Theorem 3 in [EKB10a]. First, divide the whole matrix $\boldsymbol{\Phi}$ into two complementary matrices $\boldsymbol{\Phi}_{opt}$ and $\boldsymbol{\Phi}_{\overline{opt}}$. Suppose $\boldsymbol{\Phi}_{opt}$ is a full column rank $m$ by $kd$ matrix whose blocks correspond to non-zero blocks of $\boldsymbol{\beta_0}$, and let $\boldsymbol{\Phi}_{\overline{opt}}$ be its complementary matrix. From Theorem 2 in [EKB10a], a sufficient condition for block orthogonal matching pursuit and $\ell_2/\ell_1$-optimisation program algorithms to correctly recover the block $k$-sparse $\boldsymbol{\beta_0}$, is that $\rho_c(\boldsymbol{\Phi}_{opt}^{\dagger}\boldsymbol{\Phi}_{\overline{opt}})<1$, where, $\rho_c(\boldsymbol{A})=\max_j \sum_i \|\boldsymbol{A}[i,j]\|_{2\to 2}$ and $\boldsymbol{A}[i,j]$ is the $(i,j)^{th}$ $d\times d$ block of $\boldsymbol{A}$. Using Moore-Penrose pseudo-inverse property of matrices and the submultiplicativity property of $\rho_c(\cdot)$ been proved in Lemma 2 in [EKB10a], we have:

$$\begin{aligned}\rho_c\left(\boldsymbol{\Phi}_{opt}^{\dagger}\boldsymbol{\Phi}_{\overline{opt}}\right) &= \rho_c\left(\left(\boldsymbol{\Phi}_{opt}^T\boldsymbol{\Phi}_{opt}\right)^{-1}\boldsymbol{\Phi}_{opt}^T\boldsymbol{\Phi}_{\overline{opt}}\right)\\ &\le \rho_c\left(\left(\boldsymbol{\Phi}_{opt}^T\boldsymbol{\Phi}_{opt}\right)^{-1}\right)\rho_c\left(\boldsymbol{\Phi}_{opt}^T\boldsymbol{\Phi}_{\overline{opt}}\right).\end{aligned}$$

On the other hand, we have:

$$\begin{aligned}\rho_c\left(\boldsymbol{\Phi}_{opt}^T\boldsymbol{\Phi}_{\overline{opt}}\right) &= \max_j \sum_i \left\|\boldsymbol{\Phi}_{opt}^T\boldsymbol{\Phi}_{\overline{opt}}[i,j]\right\|_{2\to 2}\\ &= \max_{j\notin\Lambda} \sum_{i\in\Lambda} \left\|\boldsymbol{\Phi}^T\boldsymbol{\Phi}[i,j]\right\|_{2\to 2}\\ &= \max_{j\notin\Lambda} \sum_{i\in\Lambda} \left\|\boldsymbol{\Phi}^T[i]\,\boldsymbol{\Phi}[j]\right\|_{2\to 2}\\ &\le d\,M_{Inter}^{Eldar}(\boldsymbol{\Phi},k),\end{aligned}$$

where, $\Lambda$ is the set of indices of blocks of $\boldsymbol{\Phi}$ which are in $\boldsymbol{\Phi}_{opt}$, and by Definition 2.8 $M_{Inter}^{Eldar}(\boldsymbol{\Phi},k)\overset{\text{def}}{=}\max_{|\Lambda|=k}\max_{j\notin\Lambda}\sum_{i\in\Lambda}\|\boldsymbol{\Phi}^T[i]\boldsymbol{\Phi}[j]\|_{2\to 2}/d$. Therefore, we have:

$$\rho_c\left(\boldsymbol{\Phi}_{opt}^{\dagger}\boldsymbol{\Phi}_{\overline{opt}}\right) \le \rho_c\left(\left(\boldsymbol{\Phi}_{opt}^T\boldsymbol{\Phi}_{opt}\right)^{-1}\right) d\,M_{Inter}^{Eldar}(\boldsymbol{\Phi},k).$$

Now it remains to upper-bound $\rho_c((\boldsymbol{\Phi}_{opt}^T\boldsymbol{\Phi}_{opt})^{-1})$. To this aim, we decompose $\boldsymbol{\Phi}_{opt}^T\boldsymbol{\Phi}_{opt}$ as $\boldsymbol{\Phi}_{opt}^T\boldsymbol{\Phi}_{opt}=\boldsymbol{I}_{kd}+\boldsymbol{F}$, where, $\boldsymbol{F}$ is a $kd$ by $kd$ matrix with blocks $\boldsymbol{F}[i,j]$ of size $d\times d$, that

$$\boldsymbol{F}[i,j]=\begin{cases}\boldsymbol{\Phi}_{opt}^T[i]\,\boldsymbol{\Phi}_{opt}[j]-\boldsymbol{I}_d, & \text{if } i=j\\ \boldsymbol{\Phi}_{opt}^T[i]\,\boldsymbol{\Phi}_{opt}[j], & \text{if } i\neq j.\end{cases}$$

All the main diagonal entries of $\boldsymbol{F}$ are zero. Then, we have:

$$\begin{aligned}\rho_c(\boldsymbol{F}) &= \max_j \sum_i \|\boldsymbol{F}[i,j]\|_{2\to 2}\\ &\le \max_j \|\boldsymbol{F}[j,j]\|_{2\to 2} + \max_j \sum_{i\neq j}\|\boldsymbol{F}[i,j]\|_{2\to 2}\\ &\le (d-1)\,M_{Intra}^{Eldar}(\boldsymbol{\Phi}) + d\,M_{Inter}^{Eldar}(\boldsymbol{\Phi},k-1),\end{aligned}$$

where, the first term is obtained by applying Geršgorin's disc theorem [HR12], and using the definition of sub-coherence proposed by Eldar [EKB10a]. Precisely, from the definition of $\boldsymbol{F}[k,k]\in\mathbb{R}^{d\times d}$, we have $\forall k,\ \boldsymbol{F}_{i,j\neq i}[k,k]\leq M_{Intra}^{Eldar}(\boldsymbol{\Phi})$, while $\boldsymbol{F}_{i,i}[k,k]=0$. On the other hand, from Corollary 6.1.5 in [HR12] (Geršgorin's disc theorem), we have $\forall \boldsymbol{A}\in\mathbb{R}^{m\times m}$, $\|\boldsymbol{A}\|_{2\to 2}\leq \min\{\|\boldsymbol{A}\|_{1\to 1},\|\boldsymbol{A}\|_{\infty\to\infty}\}$. Therefore, $\max_j \|\boldsymbol{F}[j,j]\|_{2\to 2}\leq (d-1)M_{Intra}^{Eldar}(\boldsymbol{\Phi})$. The second term follows from the definition of the cumulative inter-block coherence constant, defined in Definition 2.8. Using Lemma 4 in [EKB10a] and considering the assumption of Theorem 2.4 which indicates that $\rho_c(\boldsymbol{F})<1$, we have:

$$\begin{aligned}
\rho_c\left(\left(\boldsymbol{\Phi}_{opt}^T\boldsymbol{\Phi}_{opt}\right)^{-1}\right) &= \rho_c\left(\sum_{i=0}^{\infty}(-\boldsymbol{F})^i\right)\\
&\leq \sum_{i=0}^{\infty}\left(\rho_c\left(\boldsymbol{F}\right)\right)^i\\
&= \frac{1}{1-\rho_c\left(\boldsymbol{F}\right)}\\
&\leq \frac{1}{1-(d-1)\,M_{Intra}^{Eldar}\left(\boldsymbol{\Phi}\right)-d\,M_{Inter}^{Eldar}\left(\boldsymbol{\Phi},k-1\right)},
\end{aligned}$$

where, the first inequality follows from the triangle inequality and submultiplicativity properties. Then we have the following inequality, which is a simple rearrangement of the equation in Theorem 2.4:

$$\begin{aligned}
\rho_c\left(\boldsymbol{\Phi}_{opt}^{\dagger}\boldsymbol{\Phi}_{\overline{opt}}\right) &\leq \frac{d\,M_{Inter}^{Eldar}\left(\boldsymbol{\Phi},k\right)}{1-(d-1)\,M_{Intra}^{Eldar}\left(\boldsymbol{\Phi}\right)-d\,M_{Inter}^{Eldar}\left(\boldsymbol{\Phi},k-1\right)}\\
&< 1.
\end{aligned}$$

□

# Bibliography


[AAG04] Zeynep Akalin-Acar and Nevzat G. Gençer. "An Advanced Boundary Element Method (BEM) Implementation for the Forward Problem of Electromagnetic Source Imaging." eng. In: *Physics in Medicine and Biology* 49.21 (Nov. 2004), pp. 5011–5028 (cit. on p. 10).

[ADR16] U. Ayaz, S. Dirksen, and H. Rauhut. "Uniform Recovery of Fusion Frame Structured Sparse Signals." In: *Applied and Computational Harmonic Analysis*. Sparse Representations with Applications in Imaging Science, Data Analysis, and Beyond, Part IISI: ICCHAS Outgrowth, part 2 41.2 (Sept. 2016), pp. 341–361 (cit. on pp. 37, 39, 40).

[AG03] C.E. Acar and N.G. Gencer. "Sensitivity of EEG and MEG to conductivity perturbations." In: *Engineering in Medicine and Biology Society, 2003. Proceedings of the 25th Annual International Conference of the IEEE*. Vol. 3. Sept. 2003, 2834–2837 Vol.3 (cit. on p. 13).

[Ahl+10] Seppo P. Ahlfors, Jooman Han, John W. Belliveau, and Matti S. Hämäläinen. "Sensitivity of MEG and EEG to Source Orientation." en. In: *Brain Topography* 23.3 (July 2010), pp. 227–232 (cit. on p. 13).

[AK98] Edoardo Amaldi and Viggo Kann. "On the approximability of minimizing nonzero variables or unsatisfied relations in linear systems." In: *Theoretical Computer Science* 209.1–2 (Dec. 1998), pp. 237–260 (cit. on pp. 18, 35).

[APJ16] F. Afdideh, R. Phlypo, and C. Jutten. "Recovery Guarantees for Mixed Norm $\ell_{p_1,p_2}$ Block Sparse Representations." In: *24th Eur. Signal Process. Conf.* (Aug. 2016), pp. 378–382 (cit. on p. 46).

[Aya14] Ulas Ayaz. "Sparse Recovery with Fusion Frames." PhD thesis. Rheinischen Friedrich-Wilhelms-Universität Bonn, 2014 (cit. on pp. 37, 39, 40).

[Bar+08] Richard Baraniuk, Mark Davenport, Ronald DeVore, and Michael Wakin. "A Simple Proof of the Restricted Isometry Property for Random Matrices." English. In: *Constructive Approximation* 28.3 (2008), pp. 253–263 (cit. on p. 20).

[Bar+10] R.G. Baraniuk, V. Cevher, M.F. Duarte, and C. Hegde. "Model-Based Compressive Sensing." In: *Information Theory, IEEE Transactions on* 56.4 (Apr. 2010), pp. 1982–2001 (cit. on p. 35).

[BD08] Thomas Blumensath and Mike E. Davies. "Iterative Thresholding for Sparse Approximations." en. In: *Journal of Fourier Analysis and Applications* 14.5-6 (Sept. 2008), pp. 629–654 (cit. on p. 19).

[BD09a] T. Blumensath and M.E. Davies. "Sampling Theorems for Signals From the Union of Finite-Dimensional Linear Subspaces." In: *Information Theory, IEEE Transactions on* 55.4 (Apr. 2009), pp. 1872–1882 (cit. on p. 32).

[BD09b] Thomas Blumensath and Mike E. Davies. "Iterative hard thresholding for compressed sensing." In: *Applied and Computational Harmonic Analysis* 27.3 (2009), pp. 265 –274 (cit. on p. 19).

[BDE09] A.M. Bruckstein, D.L. Donoho, and M. Elad. "From Sparse Solutions of Systems of Equations to Sparse Modeling of Signals and Images." In: *SIAM Rev.* 51.1 (2009), pp. 34–81 (cit. on pp. 21, 29, 65, 66, 74).

[Ber99] Dimitri P. Bertsekas. *Nonlinear Programming*. Anglais. 2nd ed. Athena Scientific, Sept. 1999 (cit. on p. 19).

[BF10] E. van den Berg and M.P. Friedlander. "Theoretical and Empirical Results for Recovery From Multiple Measurements." In: *Information Theory, IEEE Transactions on* 56.5 (May 2010), pp. 2516–2527 (cit. on p. 32).

[BHE11] Z. Ben-Haim and Y.C. Eldar. "Near-Oracle Performance of Greedy Block-Sparse Estimation Techniques From Noisy Measurements." In: *Selected Topics in Signal Processing, IEEE Journal of* 5.5 (Sept. 2011), pp. 1032–1047 (cit. on pp. 35, 38, 40).

[BKR11] P. Boufounos, G. Kutyniok, and H. Rauhut. "Sparse Recovery From Combined Fusion Frame Measurements." In: *IEEE Transactions on Information Theory* 57.6 (June 2011), pp. 3864–3876 (cit. on pp. 37, 39, 40, 67, 81).

[BM99] A.P. Berg and W.B. Mikhael. "A survey of mixed transform techniques for speech and image coding." In: *Circuits and Systems, 1999. ISCAS '99. Proceedings of the 1999 IEEE International Symposium on.* Vol. 4. July 1999, pp. 106–109 (cit. on p. 15).

[BN03] Dimitri Bertsekas and Angelia Nedic. *Convex Analysis and Optimization.* English. Belmont, Mass: Athena Scientific, Apr. 2003 (cit. on p. 19).

[BV04] Stephen Boyd and Lieven Vandenberghe. *Convex Optimization.* Cambridge, U.K.: Cambridge Univ. Press, 2004 (cit. on p. 19).

[Can08] Emmanuel J. Candès. "The restricted isometry property and its implications for compressed sensing." In: *Comptes Rendus Mathematique* 346.9-10 (2008), pp. 589–592 (cit. on p. 20).

[CD02] Emmanuel J. Candès and David L. Donoho. "New Tight Frames of Curvelets and Optimal Representations of Objects with C2 Singularities." In: *Comm. Pure Appl. Math.* 57 (Nov. 2002), pp. 219–266 (cit. on p. 15).

[CDD09] A. Cohen, W. Dahmen, and R. Devore. "Compressed sensing and best k -term approximation." In: *Journal of The American Mathematical Society* 22.1 (July 2009), pp. 211–231 (cit. on pp. 20, 27, 67, 81).

[CDS01] Scott Shaobing Chen, David L. Donoho, and Michael A. Saunders. "Atomic Decomposition by Basis Pursuit." In: *SIAM J. Sci. Comput.* 43.1 (2001), pp. 129–159 (cit. on pp. 15, 16, 19, 40).

[CH03] D. Cohen and E. Halgren. "Magnetoencephalography (Neuromagnetism)." In: *Encyclopedia of Neuroscience* (2003) (cit. on p. 8).

[CH06] Jie Chen and X. Huo. "Theoretical Results on Sparse Representations of Multiple-Measurement Vectors." In: *Signal Processing, IEEE Transactions on* 54.12 (Dec. 2006), pp. 4634–4643 (cit. on p. 32).

[Cha07] R. Chartrand. "Exact Reconstruction of Sparse Signals via Nonconvex Minimization." In: *IEEE Signal Process. Letters* 14.10 (Oct. 2007), pp. 707–710 (cit. on pp. 18, 25).

[CK04] P. G. Casazza and G. Kutyniok. "Frames of subspaces." In: *Wavelets, Frames, and Operator Theory, ser. Contemp. Math. . Providence, RI: Amer. Math. Soc.* 345 (2004), pp. 87–113 (cit. on p. 32).

[CM06] G. Cormode and S. Muthukrishnan. "Combinatorial Algorithms for Compressed Sensing." In: *2006 40th Annual Conference on Information Sciences and Systems.* Mar. 2006, pp. 198–201 (cit. on p. 19).

[Cot+05] S.F. Cotter, B.D. Rao, K. Engan, and K. Kreutz-Delgado. "Sparse solutions to linear inverse problems with multiple measurement vectors." In: *Signal Processing, IEEE Transactions on* 53.7 (July 2005), pp. 2477–2488 (cit. on p. 32).

[CR02] S.F. Cotter and B.D. Rao. "Sparse channel estimation via matching pursuit with application to equalization." In: *Communications, IEEE Transactions on* 50.3 (Mar. 2002), pp. 374–377 (cit. on pp. 15, 32).

[CRT06] Emmanuel Candès, Justin Romberg, and Terence Tao. "Stable signal recovery from incomplete and inaccurate measurements." In: *Communications on Pure and Applied Mathematics* 59.8 (Aug. 2006), pp. 1207–1223 (cit. on p. 17).

[CT05] E.J. Candès and T. Tao. "Decoding by linear programming." In: *IEEE Trans. Inf. Theory* 51.12 (Dec. 2005), pp. 4203–4215 (cit. on pp. 16–20, 25).

[CT06] E.J. Candès and T. Tao. "Near-Optimal Signal Recovery From Random Projections: Universal Encoding Strategies?" In: *IEEE Trans. Inf. Theory* 52.12 (Dec. 2006), pp. 5406–5425 (cit. on p. 20).

[CT07] Emmanuel Candès and Terence Tao. "The Dantzig selector: statistical estimation when p is much larger than n." In: *The Annals of Statistics* 35.6 (2007), pp. 2313–2351 (cit. on pp. 15, 17, 19).

[CWX10] T.T. Cai, Lie Wang, and Guangwu Xu. "Shifting Inequality and Recovery of Sparse Signals." In: *Signal Processing, IEEE Transactions on* 58.3 (Mar. 2010), pp. 1300–1308 (cit. on pp. 19, 25).

[CXZ09] T.T. Cai, Guangwu Xu, and Jun Zhang. "On Recovery of Sparse Signals Via $\ell_1$ Minimization." In: *Information Theory, IEEE Transactions on* 55.7 (July 2009), pp. 3388–3397 (cit. on p. 19).

[Dau06] L. Daudet. "Sparse and Structured Decompositions of Signals with the Molecular Matching Pursuit." In: *IEEE Transactions on Audio, Speech, and Language Processing* 14.5 (Sept. 2006), pp. 1808–1816 (cit. on p. 32).

[DCL97] V.E. DeBrunner, Lixiang Chen, and Hong-Jian Li. "Lapped multiple bases algorithms for still image compression without blocking effect." In: *Image Processing, IEEE Transactions on* 6.9 (Sept. 1997), pp. 1316–1321 (cit. on p. 15).

[DDM08] Ingrid Daubechies, Michel Defrise, and Christine De Mol. "An Iterative Thresholding Algorithm For Linear Inverse Problems With A Sparsity Constraint." In: *ArXiv Mathematics e-prints* (Feb. 2008) (cit. on p. 19).

[DE03a] D.L. Donoho and M. Elad. "Maximal Sparsity Representation via $l_1$ Minimization." In: *Proceedings of the National Academy of Sciences of the United States of America* 100.50 (2003), pp. 2197–2202 (cit. on pp. 16, 19, 21–23, 26, 29, 65, 66).

[DE03b] D.L. Donoho and M. Elad. "Optimally sparse representation in general (nonorthogonal) dictionaries via $\ell^1$ minimization." In: *in Proc. Natl. Acad. Sci.* 100.5 (Mar. 2003), pp. 2197–2202 (cit. on pp. 16, 19–23, 26, 27, 29, 31, 46, 52, 65–67, 74, 81, 83, 117).

[DET06] D.L. Donoho, M. Elad, and V.N. Temlyakov. "Stable recovery of sparse overcomplete representations in the presence of noise." In: *Information Theory, IEEE Transactions on* 52.1 (Jan. 2006), pp. 6–18 (cit. on pp. 16–19, 25, 30, 74).

[DG08] Mike Davies and Remi Gribonval. *Restricted Isometry Constants where p sparse recovery can fail for* $0 < p \leq 1$. Tech. rep. July 2008 (cit. on pp. 17, 18, 20).

[DG09a] M. E. Davies and R. Gribonval. "lp minimization and sparse approximation failure for compressible signals." In: *SAMPTA* (2009) (cit. on pp. 20, 27).

[DG09b] Mike Davies and Remi Gribonval. "The Restricted Isometry Property and $\ell^p$ sparse recovery failure." In: *SPARS09* (Apr. 2009) (cit. on pp. 18, 20).

[DH01] D.L. Donoho and X. Huo. "Uncertainty principles and ideal atomic decomposition." In: *IEEE Trans. Inf. Theory* 47.7 (Nov. 2001), pp. 2845–2862 (cit. on pp. 16, 19, 20, 22, 25–29, 31, 46, 61, 65–67, 69, 74, 76, 81).

[DMA97] G. Davis, S. Mallat, and M. Avellaneda. "Adaptive greedy approximations." English. In: *Constructive Approximation* 13.1 (1997), pp. 57–98 (cit. on p. 19).

[Don04a] David L. Donoho. *For most large underdetermined systems of equations, the minimal* $\ell^1$*-norm near-solution approximates the sparsest near-solution.* Tech. rep. Comm. Pure Appl. Math, Aug. 2004 (cit. on p. 17).

[Don04b] David L. Donoho. "For Most Large Underdetermined Systems of Linear Equations the Minimal $\ell^1$-norm Solution is also the Sparsest Solution." In: *Comm. Pure Appl. Math* 59 (2004), pp. 797–829 (cit. on p. 17).

[Don05] David L. Donoho. "High-Dimensional Centrally Symmetric Polytopes with Neighborliness Proportional to Dimension." en. In: *Discrete and Computational Geometry* 35.4 (Dec. 2005), pp. 617–652 (cit. on p. 20).

[Dos05] C. Dossal. "Estimation de fonctions géométriques et déconvolution." PhD thesis. 2005 (cit. on pp. 24, 31).

[DS89] D.L. Donoho and P.B. Stark. "Uncertainty Principles and Signal Recovery." In: *SIAM J. Appl. Math.* 49.3 (1989), pp. 906–931 (cit. on pp. 26, 28, 65, 66, 69).

[DT05] David L. Donoho and Jared Tanner. "Sparse nonnegative solution of underdetermined linear equations by linear programming." In: *Proceedings of The National Academy of Sciences*. Vol. 102. 2005, pp. 9446–9451 (cit. on p. 20).

[DT06] D. L. Donoho and J. Tanner. "Thresholds for the Recovery of Sparse Solutions via L1 Minimization." In: *2006 40th Annual Conference on Information Sciences and Systems*. Mar. 2006, pp. 202–206 (cit. on p. 20).

[EB01] M. Elad and A.M. Bruckstein. "On sparse signal representations." In: *Image Processing, 2001. Proceedings. 2001 International Conference on* 1 (2001), pp. 3–6 (cit. on pp. 19, 20, 22, 26–29, 46, 65–67, 69, 81, 140).

[EB02] M. Elad and A.M. Bruckstein. "A generalized uncertainty principle and sparse representation in pairs of bases." In: *IEEE Trans. Inf. Theory* 48.9 (Sept. 2002), pp. 2558–2567 (cit. on pp. 22, 26–28, 46, 65–67, 69, 140–142).

[EB09] Y.C. Eldar and H. Bolcskei. "Block-sparsity: Coherence and efficient recovery." In: *Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on*. Apr. 2009, pp. 2885–2888 (cit. on pp. 35, 36, 38–40, 46, 61, 63, 69, 76, 81, 140).

[EKB10a] Y. Eldar, P. Kuppinger, and H. Bölcskei. "Block-sparse signals: Uncertainty relations and efficient recovery." In: *IEEE Trans. Signal Process.* 58.6 (June 2010), pp. 3042–3054 (cit. on pp. 32, 35, 36, 38–40, 46, 61, 63, 69, 76, 81, 140, 154, 155).

[EKB10b] Y.C. Eldar, P. Kuppinger, and H. Bölcskei. "Compressed Sensing of Block-Sparse Signals: Uncertainty Relations and Efficient Recovery." In: *IEEE Transactions on Signal Processing* 58.6 (June 2010), pp. 3042–3054. arXiv: `0906.3173` (cit. on pp. 35, 36, 39, 40, 46, 61, 63, 69, 81, 140).

[Eks11] E.M. Eksioglu. "A clustering based framework for dictionary block structure identification." In: *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on*. May 2011, pp. 4044–4047 (cit. on p. 85).

[Ela10] Michael Elad. *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*. Ed. by Springer. {Springer}, 2010 (cit. on p. 121).

[Eld09] Y.C. Eldar. "Compressed Sensing of Analog Signals in Shift-Invariant Spaces." In: *Signal Processing, IEEE Transactions on* 57.8 (Aug. 2009), pp. 2986–2997 (cit. on p. 32).

[Elh12] Ehsan Elhamifar. "Sparse modeling for high-dimensional multi-manifold data analysis." PhD thesis. United States – Maryland: The Johns Hopkins University, 2012 (cit. on p. 35).

[EM09] Y.C. Eldar and M. Mishali. "Robust Recovery of Signals From a Structured Union of Subspaces." In: *IEEE Trans. Inf. Theory* 55.11 (Nov. 2009), pp. 5302–5316 (cit. on pp. 35, 40).

[ER10] Y.C. Eldar and H. Rauhut. "Average Case Analysis of Multichannel Sparse Recovery Using Convex Relaxation." In: *Information Theory, IEEE Transactions on* 56.1 (Jan. 2010), pp. 505–519 (cit. on p. 32).

[ES05] Stephen Erickson and Chiara Sabatti. "Empirical Bayes Estimation of a Sparse Vector of Gene Expression Changes." eng. In: *Statistical Applications in Genetics and Molecular Biology* 4.1 (2005), p. 22 (cit. on pp. 15, 32).

[ESV12] E. Elhamifar, G. Sapiro, and R. Vidal. "See all by looking at a few: Sparse modeling for finding representative objects." In: *2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. June 2012, pp. 1600–1607 (cit. on p. 32).

[EV09] E. Elhamifar and R. Vidal. "Sparse subspace clustering." In: *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*. June 2009, pp. 2790–2797 (cit. on p. 32).

[EV10] E. Elhamifar and R. Vidal. "Clustering disjoint subspaces via sparse representation." In: *2010 IEEE International Conference on Acoustics, Speech and Signal Processing*. Mar. 2010, pp. 1926–1929 (cit. on p. 32).

[EV11] E. Elhamifar and R. Vidal. "Robust classification using structured sparse representation." In: *2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. June 2011, pp. 1873–1879 (cit. on pp. 32, 35, 37, 41).

[EV12] E. Elhamifar and R. Vidal. "Block-Sparse Recovery via Convex Optimization." In: *IEEE Transactions on Signal Processing* 60.8 (2012), pp. 4094–4107 (cit. on pp. 32, 35, 37, 38, 41).

[EV13] E. Elhamifar and R. Vidal. "Sparse Subspace Clustering: Algorithm, Theory, and Applications." In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 35.11 (Nov. 2013), pp. 2765–2781 (cit. on p. 32).

[FL09] Simon Foucart and Ming-Jun Lai. "Sparsest solutions of underdetermined linear systems via $\ell_q$-minimization for $0 < q \leq 1$." In: *App. Comput. Harmonic Anal.* 26.3 (2009), pp. 395–407 (cit. on pp. 17, 18, 20).

[FN03a] A. Feuer and A. Nemirovski. "On sparse representation in pairs of bases." In: *IEEE Transactions on Information Theory* 49.6 (June 2003), pp. 1579–1581 (cit. on pp. 20, 27, 29, 67, 81).

[FN03b] M. A. T. Figueiredo and R. D. Nowak. "An EM Algorithm for Wavelet-Based Image Restoration." In: *IEEE Transactions on Image Processing* 12.8 (Aug. 2003), pp. 906–916 (cit. on p. 19).

[FR08] M. Fornasier and H. Rauhut. "Recovery Algorithms for Vector-Valued Data with Joint Sparsity Constraints." In: *SIAM Journal on Numerical Analysis* 46.2 (Jan. 2008), pp. 577–613 (cit. on p. 32).

[FR13] Simon Foucart and Holger Rauhut. *A Mathematical Introduction to Compressive Sensing*. Birkhäuser, 2013 (cit. on pp. 20, 22–24, 27, 29–31).

[Fuc04] J.-J. Fuchs. "On sparse representations in arbitrary redundant bases." In: *Information Theory, IEEE Transactions on* 50.6 (June 2004), pp. 1341–1344 (cit. on p. 20).

[GA04] Nevzat G. Gençer and Can E. Acar. "Sensitivity of EEG and MEG Measurements to Tissue Conductivity." en. In: *Physics in Medicine and Biology* 49.5 (2004), p. 701 (cit. on p. 13).

[GC08a] J. F. Gemmeke and B. Cranen. "Noise robust digit recognition using sparse representations." In: *ISCA ITRW* (2008) (cit. on p. 32).

[GC08b] J. F. Gemmeke and B. Cranen. "Using sparse representations for missing data imputation in noise robust speech recognition." In: *Signal Processing Conference, 2008 16th European*. Aug. 2008, pp. 1–5 (cit. on p. 32).

[GE10] K. Gedalyahu and Y.C. Eldar. "Time-Delay Estimation From Low-Rate Samples: A Union of Subspaces Approach." In: *Signal Processing, IEEE Transactions on* 58.6 (June 2010), pp. 3017–3031 (cit. on p. 32).

[GGR95] I. F. Gorodnitsky, J. S. George, and B. D. Rao. "Neuromagnetic Source Imaging with FOCUSS: A Recursive Weighted Minimum Norm Algorithm." eng. In: *Electroencephalogr. Clin. Neurophysiol.* 95.4 (Oct. 1995), pp. 231–251 (cit. on p. 32).

[GH04] Sara A. Van De Geer and Hans C. Van Houwelingen. "High-dimensional data: $p \gg n$ in mathematical statistics and bio-medical applications." In: *Bernoulli* 10.6 (2004), pp. 939–942 (cit. on p. 15).

[Gil+06] A. C. Gilbert, M. J. Strauss, J. A. Tropp, and R. Vershynin. "Algorithmic linear dimension reduction in the $\ell_1$ norm for sparse vectors." In: *in Proc. 44th Annu. Allerton Conf. Communication, Control, Computing* (2006) (cit. on p. 19).

[Gil+07] A. C. Gilbert, M. J. Strauss, J. A. Tropp, and R. Vershynin. "One Sketch for All: Fast Algorithms for Compressed Sensing." In: *Proceedings of the Thirty-ninth Annual ACM Symposium on Theory of Computing*. STOC '07. New York, NY, USA: ACM, 2007, pp. 237–246 (cit. on p. 19).

[GL13] Gene H. Golub and Charles F. van Loan. *Matrix Computations*. 4th ed. The Johns Hopkins Univ. Press, 2013 (cit. on pp. 18, 54, 55, 57, 120, 121, 123, 124, 140, 142).

[GMW90] Philip E. Gill, Walter Murray, and Margaret H. Wright. *Numerical Linear Algebra and Optimization*. English. Vol. 1. Redwood City, Calif.: Perseus Books, Sept. 1990 (cit. on p. 19).

[GN03a] R. Gribonval and M. Nielsen. "Sparse decompositions in "incoherent" dictionaries." In: *in Proc. IEEE Int. Conf. Image Process.* 1 (Sept. 2003), pp. I–33–I–36 (cit. on pp. 15, 19, 26, 27, 29, 46, 65–67, 74).

[GN03b] R. Gribonval and M. Nielsen. "Sparse representations in unions of bases." In: *IEEE Trans. Inf. Theory* 49.12 (Dec. 2003), pp. 3320–3325 (cit. on pp. 15, 16, 18, 20, 21, 26, 27, 29, 46, 52, 65–67, 81).

[GN04] Rémi Gribonval and Morten Nielsen. "On the Strong Uniqueness of Highly Sparse Representations from Redundant Dictionaries." In: *Int. Conf. Ind. Compon. Anal. Blind Signal Sep.* 3195 (2004), pp. 201–208 (cit. on pp. 18, 27, 67).

[GN07] R. Gribonval and M. Nielsen. "Highly sparse representations from dictionaries are unique and independent of the sparseness measure." In: *Appl. Comput. Harmonic Anal.* 22.3 (2007), pp. 335–355 (cit. on pp. 18, 20, 27, 30, 67, 74).

[GR97] I.F. Gorodnitsky and B.D. Rao. "Sparse signal reconstruction from limited data using FOCUSS: a re-weighted minimum norm algorithm." In: *Signal Processing, IEEE Transactions on* 45.3 (Mar. 1997), pp. 600–616 (cit. on pp. 21, 26, 65, 66).

[GZM09] A. Ganesh, Z. Zhou, and Y. Ma. "Separation of a subspace-sparse signal: Algorithms and conditions." In: *Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on.* Apr. 2009, pp. 3141–3144 (cit. on pp. 35–37, 39, 41, 76).

[Had02] Jacques Hadamard. "Sur les problèmes aux dérivés partielles et leur signification physique." In: *Princeton Univ.* 13 (1902), pp. 49–52 (cit. on p. 15).

[Hau+14] J. Haueisen, S. Lau, L. Flemming, H. Sonntag, B. Maess, and D. Güllmar. "Influence of Volume Conductor Modeling on Source Reconstruction in Magnetoencephalography and Electroencephalography." In: *General Assembly and Scientific Symposium (URSI GASS), 2014 XXXIth URSI.* Aug. 2014, pp. 1–2 (cit. on p. 10).

[Hos+78] R. S. Hosek, A. Sances, R. W. Jodat, and S. J. Larson. "The Contributions of Intracerebral Currents to the EEG and Evoked Potentials." In: *IEEE Transactions on Biomedical Engineering* BME-25.5 (Sept. 1978), pp. 405–413 (cit. on p. 10).

[HP17] Riitta Hari and Aina Puce. *MEG-EEG Primer.* 2017 (cit. on p. 13).

[HR09] N. Hurley and Scott Rickard. "Comparing Measures of Sparsity." In: *IEEE Trans. Inf. Theory* 55.10 (Oct. 2009), pp. 4723–4741 (cit. on p. 18).

[HR12] Johnson C.R. Horn R.A. *Matrix Analysis.* 2ed. New York, NY: Cambridge Press, 2012 (cit. on pp. 56, 132, 155).

[HSP06] R. W. Heath, T. Strohmer, and A. J. Paulraj. "On Quasi-Orthogonal Signatures for CDMA Systems." In: *IEEE Transactions on Information Theory* 52.3 (Mar. 2006), pp. 1217–1226 (cit. on p. 22).

[Huo99] Xiaoming Huo. "Sparse Image Representation Via Combined Transforms." PhD thesis. The Department Of Statistics Of Stanford University, Aug. 1999 (cit. on p. 15).

[KKB07] Kwangmoo Koh, Seung-Jean Kim, and S. Boyd. "An Efficient Method for Large-Scale l1-Regularized Convex Loss Minimization." In: *Information Theory and Applications Workshop, 2007.* Jan. 2007, pp. 223–230 (cit. on p. 19).

[KOP13] N. Keriven, K. O'Hanlon, and M. D. Plumbley. "Structured Sparsity Using Backwards Elimination for Automatic Music Transcription." In: *2013 IEEE International Workshop on Machine Learning for Signal Processing (MLSP).* Sept. 2013, pp. 1–6 (cit. on p. 32).

[KR12] H. Kwon and B.D. Rao. "On the benefits of the block-sparsity structure in sparse signal recovery." In: *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on.* Mar. 2012, pp. 3685–3688 (cit. on p. 32).

[Krs+05] S. Krstulovic, R. Gribonval, P. Leveau, and L. Daudet. "A Comparison of Two Extensions of the Matching Pursuit Algorithm for the Harmonic Decomposition of Sounds." In: *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, 2005.* Oct. 2005, pp. 259–262 (cit. on p. 32).

[Lan67] H.J. Landau. "Necessary density conditions for sampling and interpolation of certain entire functions." English. In: *Acta Mathematica* 117.1 (1967), pp. 37–52 (cit. on p. 32).

[LBW11] Xiaolei Lv, Guoan Bi, and C. Wan. "The Group Lasso for Stable Recovery of Block-Sparse Signal Representations." In: *Signal Processing, IEEE Transactions on* 59.4 (Apr. 2011), pp. 1371–1382 (cit. on p. 35).

[LD08] Y.M. Lu and M.N. Do. "Sampling Signals from a Union of Subspaces." In: *Signal Processing Magazine, IEEE* 25.2 (Mar. 2008), pp. 41–47 (cit. on p. 32).

[LDP07] Michael Lustig, David Donoho, and John M. Pauly. "Sparse MRI: The application of compressed sensing for rapid MR imaging." In: *Magnetic Resonance in Medicine* 53.6 (2007), pp. 1182–95 (cit. on p. 15).

[LL11] Ming-Jun Lai and Yang Liu. "The null space property for sparse recovery from multiple measurement vectors." In: *Applied and Computational Harmonic Analysis* 30.3 (2011), pp. 402–406 (cit. on p. 32).

[LWY13] Qing Ling, Zaiwen Wen, and Wotao Yin. "Decentralized Jointly Sparse Optimization by Reweighted $\ell_q$ Minimization." In: *Signal Processing, IEEE Transactions on* 61.5 (Mar. 2013), pp. 1165–1170 (cit. on p. 32).

[Mal08] Stéphane Mallat. *A Wavelet Tour of Signal Processing*. 2008 (cit. on p. 19).

[Mal09] A. Maleki. "Coherence Analysis of Iterative Thresholding Algorithms." In: *47th Annual Allerton Conference on Communication, Control, and Computing, 2009. Allerton 2009*. Sept. 2009, pp. 236–243 (cit. on pp. 19, 30).

[Mal11] Jaakko Malmivuo. "Comparison of the Properties of EEG and MEG in Detecting the Electric Activity of the Brain." en. In: *Brain Topography* 25.1 (Sept. 2011), pp. 1–19 (cit. on p. 13).

[MCW05] D. Malioutov, M. Cetin, and A.S. Willsky. "A sparse signal reconstruction perspective for source localization with sensor arrays." In: *Signal Processing, IEEE Transactions on* 53.8 (Aug. 2005), pp. 3010–3022 (cit. on p. 32).

[MD10] A. Maleki and D. L. Donoho. "Optimally Tuned Iterative Reconstruction Algorithms for Compressed Sensing." In: *IEEE Journal of Selected Topics in Signal Processing* 4.2 (Apr. 2010), pp. 330–341 (cit. on p. 19).

[ME08] M. Mishali and Y.C. Eldar. "Reduce and Boost: Recovering Arbitrary Sets of Jointly Sparse Vectors." In: *Signal Processing, IEEE Transactions on* 56.10 (Oct. 2008), pp. 4692–4702 (cit. on p. 32).

[ME09] M. Mishali and Y.C. Eldar. "Blind Multiband Signal Reconstruction: Compressed Sensing for Analog Signals." In: *Signal Processing, IEEE Transactions on* 57.3 (Mar. 2009), pp. 993–1009 (cit. on p. 32).

[ME10] M. Mishali and Y.C. Eldar. "From Theory to Practice: Sub-Nyquist Sampling of Sparse Wideband Analog Signals." In: *IEEE J. Sel. Topics Signal Process.* 4.2 (Apr. 2010), pp. 375–391 (cit. on p. 32).

[Mis+11] M. Mishali, Y. C. Eldar, O. Dounaevsky, and E. Shoshan. "Xampling: Analog to Digital at Sub-Nyquist Rates." In: *IET Circuits, Devices Systems* 5.1 (Jan. 2011), pp. 8–20 (cit. on p. 32).

[Mol+08] A. Molins, S. M. Stufflebeam, E. N. Brown, and M. S. Hämäläinen. "Quantification of the benefit from integrating MEG and EEG data in minimum l2-norm estimation." In: *NeuroImage* 42.3 (2008), pp. 1069–1077 (cit. on p. 103).

[MY09] Nicolai Meinshausen and Bin Yu. "Lasso-type recovery of sparse representations for high-dimensional data." In: *The Annals of Statistics* 37.1 (2009), pp. 246–270 (cit. on p. 19).

[MZ06] Dmitri Model and Michael Zibulevsky. "Signal Reconstruction in Sensor Arrays Using Sparse Representations." In: *Signal Processing*. Sparse Approximations in Signal and Image ProcessingSparse Approximations in Signal and Image Processing 86.3 (Mar. 2006), pp. 624–638 (cit. on p. 32).

[MZ93] S.G. Mallat and Z. Zhang. "Matching pursuits with time-frequency dictionaries." In: *Signal Processing, IEEE Transactions on* 41.12 (Dec. 1993), pp. 3397–3415 (cit. on pp. 15, 19, 28).

[Nat95] B. K. Natarajan. "Sparse Approximate Solutions To Linear Systems." In: *SIAM J. Comput.* 24.2 (Apr. 1995), pp. 227–234 (cit. on pp. 18, 35).

[OF97] Bruno A. Olshausen and David J. Field. "Sparse coding with an overcomplete basis set: A strategy employed by V1?" In: *Vision Research* 37.23 (1997), pp. 3311–3325 (cit. on p. 15).

[Oka93] Y. Okada. "Empirical Bases for Constraints in Current-Imaging Algorithms." eng. In: *Brain Topography* 5.4 (1993), pp. 373–377 (cit. on p. 8).

[Oos+11] Robert Oostenveld, Pascal Fries, Eric Maris, and Jan-Mathijs Schoffelen. *FieldTrip: Open Source Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data.* en. `https://www.hindawi.com/journals/cin/2011/156869/`. Research article. 2011 (cit. on p. 98).

[OWK97] Yoshio C. Okada, Jie Wu, and Shinichi Kyuhou. "Genesis of MEG Signals in a Mammalian CNS Structure." In: *Electroencephalography and Clinical Neurophysiology* 103.4 (Oct. 1997), pp. 474–485 (cit. on p. 8).

[Par+08] F. Parvaresh, H. Vikalo, S. Misra, and B. Hassibi. "Recovering Sparse Signals Using Sparse Measurement Matrices in Compressed DNA Microarrays." In: *IEEE J. Sel. Topics Signal Process.* 2.3 (June 2008), pp. 275–285 (cit. on p. 32).

[PH67] Robert Plonsey and Dennis B. Heppner. "Considerations of Quasi-Stationarity in Electrophysiological Systems." en. In: *The bulletin of mathematical biophysics* 29.4 (Dec. 1967), pp. 657–664 (cit. on p. 10).

[PL15] Y. Peng and B. L. Lu. "Robust group sparse representation via half-quadratic optimization for face recognition." In: *2015 IEEE 28th Canadian Conference on Electrical and Computer Engineering (CCECE)*. May 2015, pp. 146–151 (cit. on p. 32).

[PRK93] Y.C. Pati, R. Rezaiifar, and P.S. Krishnaprasad. "Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition." In: *Signals, Systems and Computers, 1993. 1993 Conference Record of The Twenty-Seventh Asilomar Conference on*. Nov. 1993, 40–44 vol.1 (cit. on p. 19).

[PV07] L.. Peotta and P. Vandergheynst. "Matching Pursuit With Block Incoherent Dictionaries." In: *Signal Processing, IEEE Transactions on* 55.9 (Sept. 2007), pp. 4549–4557 (cit. on p. 61).

[RRN12] Nikhil Rao, Benjamin Recht, and Robert Nowak. "Universal Measurement Bounds for Structured Sparse Signal Recovery." In: *Proceedings of the 15th International Conference on Artificial Intelligence and Statistics (AISTATS)* (2012), pp. 942–950 (cit. on p. 32).

[Sar87] J. Sarvas. "Basic Mathematical and Electromagnetic Concepts of the Biomagnetic Inverse Problem." en. In: *Physics in Medicine and Biology* 32.1 (1987), p. 11 (cit. on p. 11).

[SCD02] J.-L. Starck, E.J. Candes, and D.L. Donoho. "The curvelet transform for image denoising." In: *Image Processing, IEEE Transactions on* 11.6 (June 2002), pp. 670–684 (cit. on p. 15).

[SED03] J.-L. Starck, M. Elad, and D. L. Donoho. "Image Decomposition: Separation of Texture from Piecewise Smooth Content." In: *SPIE Meeting* (Aug. 2003) (cit. on p. 15).

[SJ03] Thomas Strohmer and Robert W. Heath Jr. "Grassmannian frames with applications to coding and communication." In: *Applied and Computational Harmonic Analysis* 14.3 (2003), pp. 257–275 (cit. on p. 22).

[SPH09] M. Stojnic, F. Parvaresh, and B. Hassibi. "On the Reconstruction of Block-Sparse Signals With an Optimal Number of Measurements." In: *IEEE Trans. Signal Process.* 57.8 (Aug. 2009), pp. 3075–3085 (cit. on pp. 35, 38, 39, 67, 81).

[Spm] *SPM - Statistical Parametric Mapping.* `http://www.fil.ion.ucl.ac.uk/spm/` (cit. on p. 98).

[SV08] K. Schnass and P. Vandergheynst. "Dictionary Preconditioning for Greedy Algorithms." In: *IEEE Transactions on Signal Processing* 56.5 (May 2008), pp. 1994–2002 (cit. on p. 23).

[SXH08] M. Stojnic, W. Xu, and B. Hassibi. "Compressed sensing - probabilistic analysis of a null-space characterization." In: *IEEE Int. Conf. Acoust., Speech, Signal Process.* (Mar. 2008), pp. 3377–3380 (cit. on pp. 20, 27, 67, 81).

[Tib94] Robert Tibshirani. "Regression Shrinkage and Selection Via the Lasso." In: *Journal of the Royal Statistical Society, Series B* 58 (1994), pp. 267–288 (cit. on p. 19).

[TP13] Andreas M. Tillmann and Marc E. Pfetsch. "The Computational Complexity of Spark, RIP, and NSP." In: *Signal Processing with Adaptive Sparse Structured Representations (SPARS'13 )* (2013) (cit. on pp. 21, 26, 28).

[TP14] A.M. Tillmann and M.E. Pfetsch. "The Computational Complexity of the Restricted Isometry Property, the Nullspace Property, and Related Concepts in Compressed Sensing." In: *Information Theory, IEEE Transactions on* 60.2 (Feb. 2014), pp. 1248–1259 (cit. on p. 28).

[Tro04a] J.A. Tropp. "Greed is good: algorithmic results for sparse approximation." In: *IEEE Trans. Inf. Theory* 50.10 (Oct. 2004), pp. 2231–2242 (cit. on pp. 20, 23, 24, 29–31, 63, 74).

[Tro04b] Joel Aaron Tropp. "Topics in sparse approximation." eng. Computational and Applied Mathematics. Univ. Texas at Austin, Aug. 2004 (cit. on pp. 1, 32, 54).

[Tro06] J.A. Tropp. "Just relax: convex programming methods for identifying sparse signals in noise." In: *Information Theory, IEEE Transactions on* 52.3 (Mar. 2006), pp. 1030–1051 (cit. on pp. 19, 20).

[Vel+16] J. Velmurugan, Sanjib Sinha, Madhu Nagappa, N. Mariyappa, P. S. Bindu, G. S. Ravi, Nandita Hazra, K. Thennarasu, V. Ravi, A. B. Taly, and P. Satishchandra. "Combined MEG–EEG Source Localisation in Patients with Sub-Acute Sclerosing Pan-Encephalitis." en. In: *Neurological Sciences* 37.8 (Apr. 2016), pp. 1221–1231 (cit. on p. 7).

[VPF15] Samuel Vaiter, Gabriel Peyré, and Jalal Fadili. "Low Complexity Regularization of Linear Inverse Problems." en. In: *Sampling Theory, a Renaissance*. Ed. by Götz E. Pfander. Applied and Numerical Harmonic Analysis. Springer International Publishing, 2015, pp. 103–153 (cit. on p. 25).

[VS92] A. M. Vershik and P. V. Sporyshev. "Asymptotic behavior of the number of faces of random polyhedra and the neighborliness problem." In: *Sel Math Soviet* 11.2 (Jan. 1992), pp. 181–201 (cit. on p. 20).

[Wal14] Éric Walter. *Numerical Methods and Optimization - A Consumer Guide*. Springer, 2014 (cit. on pp. 17, 18).

[Wel74] L. Welch. "Lower Bounds on the Maximum Cross Correlation of Signals (Corresp.)" In: *IEEE Transactions on Information Theory* 20.3 (May 1974), pp. 397–399 (cit. on p. 22).

[Wri+09] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma. "Robust Face Recognition via Sparse Representation." In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 31.2 (Feb. 2009), pp. 210–227 (cit. on p. 32).

[Wyn79] A.D. Wyner. "An analog scrambling scheme which does not expand bandwidth, Part II: Continuous time." In: *IEEE Trans. Inf. Theory* 25.4 (July 1979), pp. 415–425 (cit. on p. 19).

[XH07] W. Xu and B. Hassibi. "Efficient Compressive Sensing with Deterministic Guarantees Using Expander Graphs." In: *IEEE Information Theory Workshop, 2007. ITW '07.* Sept. 2007, pp. 414–419 (cit. on p. 19).

[YL06] Ming Yuan and Yi Lin. "Model selection and estimation in regression with grouped variables." In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 68.1 (Feb. 2006), pp. 49–67 (cit. on p. 35).

[YNH91] Y. Yan, P. L. Nunez, and R. T. Hart. "Finite-Element Model of the Human Head: Scalp Potentials Due to Dipole Sources." en. In: *Medical and Biological Engineering and Computing* 29.5 (Sept. 1991), pp. 475–481 (cit. on p. 10).

[Zel+08a] A. C. Zelinski, L. L. Wald, K. Setsompop, V. K. Goyal, and E. Adalsteinsson. "Sparsity-Enforced Slice-Selective MRI RF Excitation Pulse Design." In: *IEEE Transactions on Medical Imaging* 27.9 (Sept. 2008), pp. 1213–1229 (cit. on p. 32).

[Zel+08b] A. C. Zelinski, V. K. Goyal, E. Adalsteinsson, and L. L. Wald. "Sparsity in MRI RF Excitation Pulse Design." In: *42nd Annual Conference on Information Sciences and Systems, 2008. CISS 2008*. Mar. 2008, pp. 252–257 (cit. on p. 32).

[Zha+15] Junxi Zhao, Rongfang Song, Jie Zhao, and Wei-Ping Zhu. "New Conditions for Uniformly Recovering Sparse Signals via Orthogonal Matching Pursuit." In: *Signal Processing* 106 (Jan. 2015), pp. 106–113 (cit. on pp. 24, 31).

[Zha05] Y. Zhang. *A Simple Proof for Recoverability of $\ell_1$-Minimization: Go Over or Under?* Tech. rep. Depart. Comput. Appl. Math., Rice Univ., Houston, 2005 (cit. on pp. 20, 27, 67, 81).

[Zia+10] A. Ziaei, A. Pezeshki, S. Bahmanpour, and M.R. Azimi-Sadjadi. "Compressed sensing of different size block-sparse signals: Efficient recovery." In: *Signals, Systems and Computers (ASILOMAR), 2010 Conference Record of the Forty Fourth Asilomar Conference on*. Nov. 2010, pp. 818–821 (cit. on pp. 37, 38, 40).

[ZMRE12] L. Zelnik-Manor, K. Rosenblum, and Y.C. Eldar. "Dictionary Optimization for Block-Sparse Representations." In: *Signal Processing, IEEE Transactions on* 60.5 (May 2012), pp. 2386–2395 (cit. on p. 38).

[ZRY09] Peng Zhao, Guilherme Rocha, and Bin Yu. "The Composite Absolute Penalties Family for Grouped and Hierarchical Variable Selection." In: *Ann. Stat.* 37.6 (2009), pp. 3468–3497 (cit. on p. 50).

[pah05] pahio. *Summed Numerator and Summed Denominator*. 2005. URL: `https://planetmath.org/SummedNumeratorAndSummedDenominator` (cit. on p. 55).

**Résumé** — De nombreux phénomènes naturels sont trop complexes pour être pleinement reconnus par un seul instrument de mesure ou par une seule modalité. Par conséquent, le domaine de recherche de la multi-modalité a émergé pour mieux identifier les caractéristiques riches du phénomène naturel de la multi-propriété naturelle, en analysant conjointement les données collectées à partir d'uniques modalités, qui sont en quelque sorte complémentaires. Dans notre étude, le phénomène d'intérêt multi-propriétés est l'activité du cerveau humain et nous nous intéressons à mieux la localiser au moyen de ses propriétés électromagnétiques, mesurables de manière non invasive. En neurophysiologie, l'électroencéphalographie (EEG) et la magnétoencéphalographie (MEG) constituent un moyen courant de mesurer les propriétés électriques et magnétiques de l'activité cérébrale. Notre application dans le monde réel, à savoir le problème de reconstruction de source EEG / MEG, est un problème fondamental en neurosciences, allant des sciences cognitives à la neuropathologie en passant par la planification chirurgicale. Considérant que le problème de reconstruction de source EEG / MEG peut être reformulé en un système d'équations linéaires sous-déterminé, la solution (l'activité estimée de la source cérébrale) doit être suffisamment parcimonieuse pour pouvoir être récupérée de manière unique. La quantité de parcimonie est déterminée par les conditions dites de récupération. Cependant, dans les problèmes de grande dimension, les conditions de récupération conventionnelles sont extrêmement strictes. En regroupant les colonnes cohérentes d'un dictionnaire, on pourrait obtenir une structure plus incohérente. Cette stratégie a été proposée en tant que cadre d'identification de structure de bloc, ce qui aboutit à la segmentation automatique de l'espace source du cerveau, sans utiliser aucune information sur l'activité des sources du cerveau et les signaux EEG / MEG. En dépit du dictionnaire structuré en blocs moins cohérent qui en a résulté, la condition de récupération conventionnelle n'est plus en mesure de calculer la caractérisation de la cohérence. Afin de relever le défi mentionné, le cadre général des conditions de récupération exactes par bloc-parcimonie, comprenant trois conditions théoriques et une condition dépendante de l'algorithme, a été proposé. Enfin, nous avons étudié la multi-modalité EEG et MEG et montré qu'en combinant les deux modalités, des régions cérébrales plus raffinées sont apparues.

**Mots clés :** (q,p)-constante cohérence mutuelle par bloc, propriété null space par bloc, bloc-spark, conditions de reconstruction exacte par bloc-parcimonieux, Principe d'incertitude par bloc-parcimonieux, bloc-parcimonieux, identification de la structure de bloc, segmentation spatiale de l'activité cérébrale, (q,p)-constante cohérence mutuelle cumulée par bloc, multimodalité en EEG et MEG.

**Abstract** — Many natural phenomena are too complex to be fully recognised by only a single measurement instrument or mono-modality. Therefore, the research domain of multi-modality has emerged to better identify the rich characteristics of the natural multi-property phenomenon, through jointly analysing the data collected from mono-modalities, which are somehow complementary. In our study, the multi-property phenomenon of interest is the human brain activity and we are interested in better localising it by means of its electromagnetic properties which are measurable non-invasively. In neurophysiology, a common way to measure the electric and magnetic properties of the brain activity is ElectroEncephaloGraphy (EEG), and MagnetoEncephaloGraphy (MEG), respectively. Our real-world application, i.e., EEG/MEG source reconstruction problem, is a fundamental problem in neuroscience ranging from cognitive science to neuropathology to surgical planning. Considering that the EEG/MEG source reconstruction problem can be reformulated as an underdetermined system of linear equations, the solution (estimated brain source activity) should be sufficiently sparse in order to be recovered uniquely. The amount of sparsity is determined by the so-called recovery conditions. However, in high-dimensional problems, the conventional recovery conditions are extremely strict. By regrouping the coherent columns of a dictionary, the more incoherent structure could be achieved. This strategy was proposed as a block structure identification framework, which results in the automatic segmentation of the brain source space, without using any information about the brain sources activity and EEG/MEG signals. Despite the resulted less coherent block-structured dictionary, the conventional recovery condition is no longer capable of computing the coherence characterisation. To address the mentioned challenge, the general framework of block-sparse exact recovery conditions including three theoretical and one algorithmic-dependent conditions was proposed. Finally, we investigated the EEG and MEG multi-modality and demonstrated that by combining the two modalities, more refined brain regions appeared.

**Keywords:** $(q, p)$-Block Mutual Coherence Constant (Block-MCC$_{q,p}$), Block Null Space Property (Block-NSP), Block-Spark, Block-sparse Exact Recovery Conditions (Block-ERC), block-sparse uncertainty principle, Block-sparsity, block structure identification, brain source space segmentation, cumulative Block-MCC$_{q,p}$, EEG and MEG multi-modality.

Grenoble Images Parole Signal Automatique laboratoire (GIPSA-Lab), 11 rue des mathématiques, Domaine Universitaire, BP 46
38402 Saint Martin d'Hères cedex