Robust and computationally efficient estimation for high-dimensional statistical learning

Yannis Bekri

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ GRENOBLE ALPES

École doctorale : MSTII - Mathématiques, Sciences et technologies de l'information, Informatique
Spécialité : Mathématiques Appliquées
Unité de recherche : Laboratoire Jean Kuntzmann

Estimation robuste et computationnellement efficace pour l'apprentissage statistique en haute dimension

Robust and computationally efficient estimation for high-dimensional statistical learning

Présentée par :

Yannis BEKRI

Direction de thèse :

Anatoli IOUDITSKI PROFESSEUR DES UNIVERSITES, UNIVERSITE GRENOBLE ALPES	Directeur de thèse
Guillaume LECUÉ PROFESSEUR, ESSEC BUSINESS SCHOOL	Co-directeur de thèse

Rapporteurs :

JOSEPH SALMON
PROFESSEUR DES UNIVERSITES, UNIVERSITE DE MONTPELLIER
VICTOR-EMMANUEL BRUNEL
PROFESSEUR, ENSAE PARIS

Thèse soutenue publiquement le **11 juin 2024**, devant le jury composé de :

JEAN-FRANÇOIS COEURJOLLY, PROFESSEUR DES UNIVERSITES, UNIVERSITE GRENOBLE ALPES	Président
ANATOLI IOUDITSKI, PROFESSEUR DES UNIVERSITES, UNIVERSITE GRENOBLE ALPES	Directeur de thèse
JOSEPH SALMON, PROFESSEUR DES UNIVERSITES, UNIVERSITE DE MONTPELLIER	Rapporteur
VICTOR-EMMANUEL BRUNEL, PROFESSEUR, ENSAE PARIS	Rapporteur
OLEG LEPSKI, PROFESSEUR DES UNIVERSITES, AIX-MARSEILLE UNIVERSITE	Examinateur
CRISTINA BUTUCEA, PROFESSEURE, ENSAE PARIS	Examinatrice

Invités :

GUILLAUME LECUE
PROFESSEUR, ESSEC BUSINESS SCHOOL

0.1 Résumée de la thèse

Cette thèse explore différentes approches statistiques visant à estimer un signal, i.e. un vecteur, à partir d'une observation. Ces observations peuvent résulter soit d'une transformation linéaire ou non linéaire du signal, corrompue par un bruit aléatoire additif. Les estimateurs étudiés intègrent l'observation disponibles et des connaissances préalables sur le signal afin de générer des estimations. Ainsi, la qualité de l'estimation dépend fortement des informations préalables susmentionnées, telles que la fonction transformant le signal en l'observation ou des propriétés spécifiques du signal lui-même. comme le nombre de ses coefficients non nuls ou des contraintes particulières qu'il satisfait. Etant donnés que ces informations ne sont souvent pas entièrement disponibles, nous proposons des estimations robustes à la mauvaise spécification du modèle. Dans ce contexte, "robuste" signifie que nos estimateurs ne nécessitent pas la connaissance de certaines informations sur le problème et tout en fournissant des performances d'estimation proches de celles obtenues avec des estimateurs utilisant une description complète du modèle. La première partie de cette thèse se concentre sur les observations linéaires, où l'observation est

$$\omega = Ax + \xi,$$

où A est une matrice, ξ le bruit additif, et $x \in \mathcal{X}$ le signal d'intérêt. Nous examinons deux cas de modèles mal spécifiés. Le premier est la situation où la description de l'ensemble des signaux $\mathcal{X}$ dépend d'un paramètre réel inconnu δ. Le second traite de situations où la matrice d'observation A est incertaine. Dans les deux cas, nous proposons des estimateurs, linéaires ou non de l'observation, qui sont robustes face à ces mauvaises spécifications du modèle. Dans la deuxième partie de cette thèse, le signal x_* à estimer est défini comme le minimiseur de l'espérance d'une fonction aléatoire. Nous proposons une méthode stochastique de premier ordre qui étant donné des observations permettant de calculer une approximation stochastique du gradient de la fonction cible, et la connaissance préalables de certaines propriétés du problèmes, produit une estimation du signal x_*. Une version robuste, dans le sens où elle ne nécessite pas la connaissance des hyperparamètres optimaux de l'algorithme précédemment mentionné, de cette méthode est ensuite développée. Cette dernière fournit des performances presque identiques à la version non-robuste.

0.2 Abstract of the thesis

This thesis explores various statistical approaches aimed at estimating a signal, i.e., a vector, from an observation. These observations may result from either a linear or nonlinear transformation of the signal, corrupted by additive random noise. The estimators studied integrate the available observation and prior knowledge of the signal to generate estimates. Thus, the quality of the estimation strongly depends on the aforementioned prior information, such as the function transforming the signal into the observation or specific properties of the signal itself, such as the number of its non-zero coefficients or particular constraints it satisfies. Given that this information is often not entirely available, we propose robust estimates to the poor specification of the model. In this context, 'robust' means that our estimators do not require knowledge of certain information about the problem while still providing estimation performance close to those obtained with estimators using a complete description of the model. The first part of this thesis focuses on linear observations, where the observation is

$$\omega = Ax + \xi,$$

where A is a matrix, ξ the additive noise, and $x \in \mathcal{X}$ the signal of interest. We examine two cases of poorly specified models. The first is the situation where the description of the set of signals $\mathcal{X}$ depends on an unknown real parameter δ. The second deals with situations where the observation matrix A is uncertain. In both cases, we propose estimators, linear or non-linear from the observation, that are robust against these poor model specifications. In the second part of this thesis, the signal x_* to be estimated is defined as the minimizer of the expectation of a random function. We propose a first-order stochastic method that, given observations allowing for a stochastic approximation of the gradient of the target function, and prior knowledge of certain properties of the problem, produces an estimation of the signal x_*. A robust version, in the sense that it does not require knowledge of the optimal hyperparameters of the previously mentioned algorithm, of this method is then developed. This version provides performances almost identical to the non-robust version.

0.3 Notations

- Real, non-negative, and positive vector spaces of dimension n are respectively denoted by $\mathbf{R}^n$, $\mathbf{R}^n_+$, $\mathbf{R}^n_{+*}$. Spaces of matrices with dimension $m \times n$ with real, non-negative or positive coefficients are respectively denoted $\mathbf{R}^{m\times n}$, $\mathbf{R}^{m\times n}_+$, and $\mathbf{R}^{m\times n}_{+*}$.

- The cones of semi-definite and positive definite matrices of $\mathbf{R}^{n\times n}$ are denoted by $\mathbf{S}^n_+$ and $\mathbf{S}^n_{++}$. For $A, B \in \mathbf{R}^n$, notations $A \succeq B$ and $A \succ B$ mean respectively that $A - B \in \mathbf{S}^n_+$ and $A - B \in \mathbf{S}^n_{++}$.

- For $x \in \mathbf{R}^n$, we refer to p-absolute norm through notation

$$\|x\|_p := \left\{ \begin{array}{ll} \left(\sum_{i=1}^n |x_i|^p\right)^{\frac{1}{p}}, & 1 \leq p \\ \\ \max\limits_{i=1,\cdots,n} |x_i|, & p = \infty \end{array} \right\}.$$

  We also use notation $\|x\|_0 := \text{Card}\,\{i \in \{1, \cdots, n\} :\ x_i \neq 0\}$.

- For a given norm $\|\cdot\|$ on $\mathbf{R}^n$, we denote by $\|\cdot\|_*$ its dual norm:

$$\forall x \in \mathbf{R}^n, \quad \|x\|_* = \max_{z\in\mathbf{R}^n} \left\{x^T z :\ \|z\| \leq 1\right\}$$

- For $\Sigma \in \mathbf{S}^n_+$, and $z \in \mathbf{R}^n$, we use notation

$$\|z\|_\Sigma = \sqrt{z^T \Sigma z}.$$

- We define, for $q = 0$ or $q \geq 1$, notations

$$\mathcal{B}_q(r) := \left\{z \in \mathbf{R}^n :\ \|z\|_q \leq r\right\} \text{ and } \mathcal{S}_q(r) := \left\{z \in \mathbf{R}^n :\ \|z\|_q = r\right\}.$$

- For $M \in \mathbf{R}^{m\times n}$, we denote by $\mathbf{Sv}(\mathrm{M})$ the set of its singular-values, and define the Frobennius norm

$$\|M\|_{\mathbf{Fro}} := \sqrt{\mathbf{Tr}\,(M^T M)},$$

  the spectral norm

$$\|M\|_{\mathbf{Sp}} := \max_{\lambda\in\mathbf{Sv}(M)} |\lambda|,$$

  and the nuclear norm

$$\|M\|_{\mathbf{nuc}} := \max_{Q:\|Q\|_{\mathbf{Sp}}=1} \mathbf{Tr}\,(MQ) = \sum_{\lambda\in\mathbf{Sv}(M)} |\lambda|.$$

- We use notation $\mathrm{Diag}(a)$ for a diagonal matrix in $\mathbf{R}^{n\times n}$ with entries $(a_i)_{i=1}^n$. I is the identity matrix.

- Indicator function of event A is denoted

$$\mathbf{1}\{A\} := \left\{ \begin{array}{ll} 1, & \text{if } A \\ 0, & \text{if } A^c \end{array} \right\}.$$

- Positive part of a real x is $x_+ := \max\{x, 0\}$.
- Given $z \in \mathbf{R}^n$, we use notation z_s to designate the vector obtained from z by zeroing all of its coefficients except the s-largest in absolute value. For $I \subset \{1, ..., n\}$, we denote by z_I the vector obtained from z by zeroing coefficient with index outside of I.
- We use notations $f = \mathcal{O}(g)$ and $f \lesssim g$ when their exist an absolute constant C such that $f \leq Cg$.
- $a \bigvee b$ and $a \bigwedge b$ respectively denote $\max\{a, b\}$ and $\min\{a, b\}$.

Contents

Chapter 1

Introduction

1.1 Statistical Estimation Problems

1.1.1 Problem statements

In this thesis, we will consider statistical model

$$\omega = \overline{\omega}(x) + \xi, \tag{1.1}$$

with unknown *signal* $x \in \mathcal{X} \subseteq \mathbf{R}^n$ and possibly random *noise* $\xi \in \mathbf{R}^m$. The mapping $\overline{\omega} : \mathbf{R}^n \to \mathbf{R}^m$ can either be linear, i.e.

$$\omega = Ax + \xi, \tag{1.2}$$

where $A \in \mathbf{R}^{m \times n}$ is an *observation* matrix, or non-linear. In that case, we have N real observations $(\eta_i)_{i=1}^N \in \mathbf{R}$, associated with regressors $(\phi_i)_{i=1}^N \in \mathbf{R}^n$, so that each η_i is linked to x by equation

$$\eta_i = \mathfrak{r}(\phi_i^T x_*) + \sigma\xi_i, \quad \sigma > 0. \tag{1.3}$$

The *link* function $\mathfrak{r}(\cdot)$ is known and non-linear, and both *regressor* ϕ_i and noise ξ_i are random. For each problem, the aim is to build statistical procedures producing estimates of respectively $w = Bx \in \mathbf{R}^\nu$ or x_*. We focus on the now classical context of modern statistics, where dimension n, i.e. the number of possible variables of influence considered when modeling a phenomenon of interest, is very high compared to number of available observations, respectively m or N.
This constitutes an additional difficulty. For instance, even in a deterministic setting (i.e. where $\xi = 0$ almost surely), recovering x from ω as in (1.2) amounts to solving a system of m equations with n unknown variables, which is impossible if $n > m$. Since adding randomness obviously blurs information available from ω, the statistical task at hand is hopeless without additional clues as to what x can be. Thus, methods developed in this thesis, in the line of modern statistical procedures, will assume that prior knowledge on the signal to be recovered is available. This will allow to build more precise estimates of the latter when combined with available observation.
We mathematically formalize this prior knowledge by $x \in \mathcal{X}$. When concerned about model (1.2), $\mathcal{X}$ will be a bounded and closed subset of $\mathbf{R}^n$. When in context of (1.3), we will assume that most of x_*'s coefficient are zeroes, except at most s of them, with $s \leq N$. In both cases, the focus will be on methods computable in reasonable time using modern software and computers, that enjoy statistical precision almost as good as the best one could hope for.

Linear regression and Least Square Estimator: a first statistical problem

When $\mathcal{X} = \mathbf{R}^n$ and $\mathfrak{r}$ is the identity, both (1.2) and (1.3) are instantiations of the same statistical problem, the *linear regression*. When assuming that noise vector ξ is a random Gaussian vector with covariance matrix $\sigma^2\mathbf{I}$, this problem becomes *Gaussian linear regression*, stated in the following way: let $n, m \in \mathbf{N}$ be integers, given *design matrix* $A \in \mathbf{R}^{m\times n}$, assume that we observe

$$\omega = Ax + \sigma\xi, \text{ with } x \in \mathbf{R}^n, \text{ and } \xi \sim \mathcal{N}(0, \mathbf{I}). \tag{1.4}$$

Note that this also falls in (1.3). For $i \in \{1, \cdots, m\}$, let $a_i = \mathrm{Row}_i(A) \in \mathbf{R}^n$ be the i-th row of observation matrix A. Then for all $i \in [1 : m]$, assuming model (1.2) is the same as observing dataset $(\omega_i, a_i)_{i=1}^m$ with

$$\omega_i = a_i^T x + \sigma\xi_i, \text{ with } \xi_i \sim \mathcal{N}(0, 1). \tag{1.5}$$

Using these observations, statistics aims at achieving many different tasks. In this thesis, we will focus on recovering the quantity $Bx \in \mathbf{R}^\nu$, with $B \in \mathbf{R}^{\nu\times n}$. Depending on B, this covers different classical tasks of statistics. When $B = \mathbf{I}$, we are *estimating* x, whereas when $B = A$, we are solving a *regression* problem. One can also aim to accurately *predict* linear form $a_{m+1}^T x$, where $a_{m+1} \in \mathbf{R}^n$ is a new regressor. For all these tasks, one has to find how to leverage knowledge of regression matrix A and observation ω to solve them.

1.1.2 Estimators and their error

In the context of both models (1.2) and (1.3), an *estimator* $\widehat{w}$ will designate a Borel function from $\mathbf{R}^n$ to $\mathbf{R}^\nu$ that takes as input observation ω and outputs *estimate* $\widehat{w}(\omega)$ of the object of interest:

$$\widehat{w} : \left\{ \begin{array}{ccc} \mathbf{R}^m & \to & \mathbf{R}^\nu \\ \omega & \to & \widehat{w}(\omega) \end{array} \right\}.$$

Throughout this manuscript, we will either refer to an estimator or to the estimates it produces, depending on the context. As a first example, in the context of Gaussian linear regression, the most famous statistical procedure is the *least square* estimate:

$$\widehat{x}^H_{\mathrm{LS}}(\omega) := A^+\omega \in \operatorname*{Argmin}_{z\in\mathbf{R}^n} \|\omega - Az\|_2^2, \quad A^+ = \lim_{\lambda\to 0_+} (A^TA + \lambda I)^{-1}A^T. \tag{1.6}$$

Here, the estimator is the linear mapping associated with $A^+ \in \mathbf{R}^{n\times m}$, the Moore-Penrose inverse of A, while the estimate of x is the matrix-vector product $\widehat{x}^H_{\mathrm{LS}}(\omega) \in \mathbf{R}^n$.
To assess how well a given estimator $\widehat{w}$ behaves with regard to the estimation of $w(x)$, one has to consider the *loss*

$$\mathcal{L}^{\|\cdot\|}(\widehat{w}(\omega), w(x)) = \|\widehat{w}(\omega) - w(x)\|, \tag{1.7}$$

where $\|\cdot\|$ is a norm on $\mathbf{R}^\nu$. In some applications where $\widehat{w}(\omega)$ estimates Ax, e.g. linear regression, the usual criterion is the *average squared prediction error*

$$\mathcal{L}^{\|\cdot\|_{2,m}^2}(\widehat{w}(\omega), Ax) := \|\widehat{w}(\omega) - Ax\|_{2,m}^2, \quad \|\cdot\|_{2,n} := \frac{1}{\sqrt{m}}\|\cdot\|_2. \tag{1.8}$$

In all cases, this measure of error is random by definition of observation ω. Thus, precision of a statistical procedure must be evaluated through a deterministic quantity describing that randomness. Ideally, $\mathcal{L}$ fluctuating randomly but having almost always small values would be a sign of a

good estimation of w. The first natural quantity to look at is error in expectation with respect to (w.r.t.) ξ:

$$\mathrm{Err}^{\|\cdot\|}(\widehat{w}, w(x)) := \mathbf{E}_\xi \left[\mathcal{L}^{\|\cdot\|}(\widehat{w}(\omega), w(x)) \right]. \tag{1.9}$$

Other quantities describing $\mathcal{L}$'s randomness are quantiles of it, defined for all $\epsilon \in (0,1)$ by

$$\mathrm{Err}_\epsilon^{\|\cdot\|}(\widehat{w}(\omega), w(x)) = \inf_\rho \left\{ \rho : \quad \mathrm{Prob}_\xi \left[\mathcal{L}^{\|\cdot\|}(\widehat{w}(\omega), w(x)) \leq \rho \right] \geq 1 - \epsilon \right\}. \tag{1.10}$$

Returning to the example of least square estimate and Gaussian linear regression, let $r = \mathrm{Rank}(A)$. Simple calculations (see [1] for expectation and results in chapter 2 of [2] for quantiles) yields

$$\mathrm{Err}^{\|\cdot\|_{2,m}^2}(\widehat{w}_{\mathrm{LS}}, Ax) = \frac{1}{m}\sigma^2 \mathbf{Tr}\left(AA^+\right) = \frac{r\sigma^2}{m}, \tag{1.11}$$

$$\mathrm{Err}_\epsilon^{\|\cdot\|_{2,m}^2}(\widehat{w}_{\mathrm{LS}}, Ax) \leq \frac{\sigma^2}{m}\left(r + 2\sqrt{r\ln(1/\epsilon)} + 2\ln(1/\epsilon) \right). \tag{1.12}$$

These can be extended to random vectors that almost behave like Gaussian ones, namely *Sub-Gaussian* random vectors.

Definition 1 *A random vector $\xi \in \mathbf{R}^m$ is said to be Sub-Gaussian with parameter (μ, Σ) ($\xi \sim \mathcal{SG}(\mu, \Sigma)$), $\mu \in \mathbf{R}^m$,$\Sigma \in \mathbf{S}_+^m$, if for all vectors $t \in \mathbf{R}^m$,*

$$\mathbf{E}_\xi \left[\exp\left(t^T(\xi - \mu) \right) \right] \leq \exp\left(\frac{t^T \Sigma t}{2} \right). \tag{1.13}$$

For instance, if $\xi \sim \mathcal{SG}(0, \sigma^2\mathbf{I})$, the least square estimate enjoys the same upper bound on quantiles of its $\|\cdot\|_{2,m}^2$ loss:

$$\mathrm{Err}_\epsilon^{\|\cdot\|_{2,m}^2}(\widehat{w}_{\mathrm{LS}}, Ax) \leq \frac{\sigma^2}{m}\left(r + 2\sqrt{r\ln(1/\epsilon)} + 2\ln(1/\epsilon) \right). \tag{1.14}$$

1.1.3 Comparing estimators

Work of this thesis will focus on efficient estimators w.r.t. different notions of performance, such as *minimax* optimality, *adaptivity* and low *computational complexity*.

Maximal risk and minimax optimality

For a given $x \in \mathbf{R}^n$, least square estimate's error in expectation is independent of the latter. Without assuming additional properties of x, this is the best one can do. Yet, with additional information about x, some *biased* estimator might perform better. Consider the celebrated *Ridge* estimate [3], defined for $\lambda \geq 0$ as

$$\widehat{x}_{\mathrm{Ridge}}^\lambda(\omega) := \left(A^TA + m\lambda\mathbf{I} \right)^{-1} A^T\omega = \operatorname*{argmin}_{z \in \mathbf{R}^n} \frac{1}{m}\|\omega - Az\|_2^2 + \lambda\|z\|_2^2. \tag{1.15}$$

This estimate is biased, as $\mathbf{E}_\xi\left[\widehat{w}^\lambda_{\text{Ridge}}(\omega)\right] = \left(A^TA + m\lambda\mathbf{I}\right)^{-1}A^TAx$. Thus, its error depends on x, and a good parameter choice for λ would depend on the latter. Indeed, for a given λ, an upper bound on the error is

$$\text{Err}^{\|\cdot\|_{2,m}^2}\left(A\widehat{x}^\lambda_{\text{Ridge}}, Ax\right) \leq \frac{\lambda\|x\|_2^2}{2} + \frac{\sigma^2\|A\|_{\mathbf{Fro}}^2}{2m^2\lambda}, \tag{1.16}$$

implying that choice $\lambda_* = \frac{\sigma\|A\|_{\mathbf{Fro}}}{m\|x\|_2}$ minimizes the RHS of (1.16) and leads to the bound

$$\text{Err}^{\|\cdot\|_{2,m}^2}\left(A\widehat{x}^{\lambda_*}_{\text{Ridge}}, Ax\right) \leq \frac{\sigma\|x\|_2\|A\|_{\mathbf{Fro}}}{m}. \tag{1.17}$$

Observe that the latter is smaller than (1.11) when $\|x\|_2 \leq \sigma n/\|A\|_{\mathbf{Fro}}$. Thus, for $\mathcal{X} \subset \mathbf{R}^n$, the least square might have greater error than other estimates.
A fair criterion to compare estimates is the supremum over all possible $x \in \mathcal{X}$ of either $\text{Err}^{\|\cdot\|}$ or $\text{Err}_\epsilon^{\|\cdot\|}$. This is exactly the *maximal risk*, that we define in expectation

$$\text{Risk}^{\|\cdot\|}(\widehat{w}, \mathcal{X}) := \sup_{x\in\mathcal{X}} \text{Err}^{\|\cdot\|}(\widehat{w}, Bx), \tag{1.18}$$

and in quantile, for all $\epsilon \in (0,1)$ by

$$\text{Risk}_\epsilon^{\|\cdot\|}(\widehat{w}, \mathcal{X}) := \sup_{x\in\mathcal{X}} \text{Err}^{\|\cdot\|}(\widehat{w}, Bx), \tag{1.19}$$

When looking at expectations, we will refer to it as *risk*, whereas when interested at quantiles, we will use the term ϵ-*risk*. From this notion of performance, we recall the minimal over all estimates maximal risk over signals in $\mathcal{X}$. We refer to the latter as *minimax risk*, and respectively define it for expected risk and ϵ-risk by

$$\begin{aligned}\text{RiskOpt}^{\|\cdot\|}(\mathcal{X}) &:= \inf_{\widehat{w}} \text{Risk}^{\|\cdot\|}(\widehat{w}, \mathcal{X}),\\ \text{RiskOpt}_\epsilon^{\|\cdot\|}(\mathcal{X}) &:= \inf_{\widehat{w}} \text{Risk}_\epsilon^{\|\cdot\|}(\widehat{w}, \mathcal{X}).\end{aligned} \tag{1.20}$$

Note that these quantities are theoretical ones, and eventual estimators achieving optimal risk are usually unknown. Nevertheless, it is still a relevant point of comparison for estimators, as one usually can provide lower bounds on the minimax risk using various techniques [4]. Note that the latter is a function of considered problem's parameters, e.g., dimensions n, m, ν, matrices A, B, signal set $\mathcal{X}$ and law $\mathbf{P}_\xi$ for (1.2). We refer to these as the *model* $\mathfrak{m}$, defined as

$$\mathfrak{m} := [\mathcal{X}, A, B, \|\cdot\|, \mathbf{P}_\xi] \tag{1.21}$$

when in context of (1.2) and

$$\mathfrak{m} := [\mathfrak{r}, \mathcal{X}, \|\cdot\|, \mathbf{P}_\phi, \mathbf{P}_\xi] \tag{1.22}$$

for (1.3)[1]. For a given model $\mathfrak{m}$, estimators such that

$$\text{Risk}^{\|\cdot\|}(\widehat{w}, \mathcal{X}) \leq C\mathfrak{R}(\mathfrak{m}), \quad C \geq 1, \quad \text{RiskOpt}^{\|\cdot\|}(\mathcal{X}) \geq \mathfrak{R}(\mathfrak{m})$$

are called *minimax optimal* when C is an absolute constant. We also use the terminology *nearly-minimax optimal* when, e.g., $C = C_1\ln(C_2 n^\alpha m^\beta \nu^\gamma)^\delta$, with $\alpha, \beta, \delta, C_1 \geq 0$ and $C_2 \geq 1$.

[1] As a consequence, we also use notation $\text{Risk}^{\|\cdot\|}(\widehat{w}, \mathfrak{m})$ in some parts of the thesis.

Adaptivity to parameters

In some chapters of this thesis, the studied statistical problems will be described by model $\mathfrak{m} \in \mathfrak{M}(\delta)$, $\delta \in [\underline{\delta}, \overline{\delta}] \subset \mathbf{R}_+$, with $\mathfrak{M}(\delta)$ a parameter dependent *class of problems* increasing in the latter, i.e. such that

$$\delta \leq \delta' \implies \mathfrak{M}(\delta) \subset \mathfrak{M}(\delta').$$

Chapters $2, 3, 6$ are concerned with estimates *adaptive* to δ, in the sense that they can be computed without knowledge of the latter. The problem of building estimates that adapt to unknown model parameter is extensively studied in the literature, starting with [5]. The author addresses the task of estimating a function $f \in \Sigma(\beta, L)$ corrupted by Gaussian error. Here, for $\beta, L > 0$, $\Sigma(\beta, L)$ is the class of functions that are $\lceil\beta\rceil$-fold continuously differentiable and such that

$$|f^{(\lceil\beta\rceil)}(t_1) - f^{(\lceil\beta\rceil)}(t_2)| \leq L|t_1 - t_2|^{\beta - \lceil\beta\rceil}.$$

The proposed procedure builds an estimate that adapts to unknown regularity β. Other adaptive estimation problems can be found in [6–8] for example. In our contributions, we propose adaptive estimates $\widehat{w}^{(a)}(\omega)$ that are data-dependent selection procedures of one $\widehat{w}_i(\omega)$ in the finite collection $\mathcal{W}_I := \{\widehat{w}_j(\omega), \ \ j \in [1:I]\}$, i.e.

$$\widehat{w}^{(a)}(\omega) = \widehat{w}_{\widehat{i}}(\omega), \quad \widehat{i} = \widehat{i}(\omega) \in [1:I]. \tag{1.23}$$

Each estimate $\widehat{w}_i \in \mathcal{W}_I$ is "tuned" for models in class $\mathfrak{M}(\delta_i)$, with $(\delta_i)_{i=1}^I$ an increasing *grid* of $[\delta_1, \overline{\delta}]$. By tuned, we mean that estimate $\widehat{w}_i(\omega)$ reaches the minimum of a risk upper bound we provide over the class of estimates it belongs to [2], or an upper bound of it.
Maximal risk of an adaptive estimate when assuming $\mathfrak{m} \in \mathfrak{M}(\delta)$,

$$\mathrm{Risk}^{\|\cdot\|}(\widehat{w}^{(a)}, \mathfrak{M}(\delta)) := \max_{\mathfrak{m} \in \mathfrak{M}(\delta)} \mathrm{Risk}^{\|\cdot\|}(\widehat{w}^{(a)}, \mathcal{X})$$

is then compared to either $\min_{i \in [1:I]} \{R_i : \ \ R_i \geq R(\delta)\}$ or $R(\delta)$, where

$$\mathrm{Risk}^{\|\cdot\|}(\widehat{w}_i, \mathfrak{M}(\delta_i)) \leq R_i, \quad \mathrm{Risk}^{\|\cdot\|}(\widehat{w}_\delta, \mathfrak{m}) \leq R(\delta),$$

and $\widehat{w}_\delta$ is the best estimate we are able to produce when knowing δ. Note that notation $\mathrm{Risk}^{\|\cdot\|}(\widehat{w}, \mathfrak{M}(\delta))$ denotes the maximal other models $mmd \in \mathfrak{M}(\delta)$ maximal risk of $\widehat{w}$.
In other words, a good adaptive selection procedure selects the "right" estimate $\widehat{w}_i$ in context of given $\mathfrak{m}$ without knowing δ. Among classical approach for data-driven estimate selection, are *cross-validation* (see 7.10 of [9] for example), Lepskii's adaptive procedure [5, 10, 11], and model penalization approach of [12].

[2] For instance, when considering Ridge estimates, choice of parameter $\lambda_* = \frac{\sigma\|A\|_{\mathbf{Fro}}}{m\|x\|_2}$ reaches the minimum (1.17) of upper bound (1.16). Thus estimate $\widehat{w}^{\lambda_*}_{\mathrm{Ridge}}(\omega)$ is not optimal w.r.t. maximal risk among the class of Ridge estimate

$$\left\{\widehat{w}^{\lambda}_{\mathrm{Ridge}}(\omega), \ \ \lambda \geq 0\right\},$$

but it is the best one can build using upper bound (1.16) when knowing $\|x\|_2 \leq \delta$.

Computational tractability

The last point of comparison of estimators that we consider here is the computational time required to produce them. More specifically, we are interested in the *computational complexity*, i.e. the worst case amount of numerical operations required to compute a given estimate. The latter is usually an increasing function of statistical problem's dimensions n, m, ν. As we deal with high-dimensional settings, we focus on estimators that have at the very least polynomial computational complexity in n, m, ν. For this reason, estimates considered in this thesis are computed by solving convex optimization problem.
To do so, we will often define quantities as minimizers of either *Linear* or *Conic Programs* (see chapter 1 and 2 of [13]). The considered programs are solvable either using *First Order Methods* (FOM), e.g., Mirror Descent [14], or by state-of-the-art optimization software that relies upon *Interior Point Methods* (IPM) [15], e.g., MOSEK or GUROBI, available when using the modeling system CVX [16]. *Second Order Methods* such as IPM are known to provide high accuracy solutions, where accuracy ϵ of a given solution $\overline{x}$ when minimizing function f is defined by

$$f(\overline{x}) \leq \min_x f(x) + \epsilon.$$

The computational complexity of IPM is dominated by the cost of a Newton step, scaling as $\mathcal{O}(N^3)$ for problems of size N. On the other hand, FOM iterations usually have linear or log-linear in n computational complexity, but need a lot of iterations to reach a high accuracy solution.
Since estimates we target usually minimize an upper bound on the maximal risk, a candidate minimizer $\overline{x}$ such that

$$f(\overline{x}) \leq 2\min_x f(x)$$

is accurate enough, provided that $\min_x f(x)$ is nearly-minimax or minimax. As this is not a high-accuracy solution, a FOM can produce it in a usually small number of iteration. Thus, when possible, we prefer using FOM to IPM to compute estimates.

1.2 Linear and Polyhedral estimates for linear inverse problems

1.2.1 Linear estimates

When treating problem (1.2), the historically most studied type of estimates are linear ones, both for their simplicity and efficiency.

Definition 2 *Let $H \in \mathbf{R}^{m\times\nu}$. We define a linear estimate of $w = Bx$ associated with parameter H as*

$$\widehat{w}^H_{Lin}(\omega) := H^T\omega. \tag{1.24}$$

Indeed, with no prior assumption on x and assuming we are in the case of (1.4), the minimax risk is (See [1] and references therein)

$$\mathrm{RiskOpt}^{\|\cdot\|_{2,m}^2}(\mathbf{R}^n) = \sigma^2\mathrm{Rank}(A)/m, \tag{1.25}$$

which equals the maximal risk of linear estimate $\widehat{w}^{AA^+}_{\mathrm{Lin}}(\omega)$, i.e. the least square. In that situation, this is the best one can achieve without any assumption about x. Yet, in situations where $n > m$, the latter behaves poorly when estimating x. Indeed, one has

$$\sup_{x\in\mathbf{R}^n} \mathrm{Err}^{\|\cdot\|_2^2}(\widehat{w}^{A^+}_{\mathrm{Lin}}(\omega), x) = \infty. \tag{1.26}$$

In fact, the minimax rate of estimation is also infinite in that case (see Proposition 1 of [1]). This is no surprise as for $x \in \mathbf{Ker}(A)$, observation $\omega = Ax + \xi = \xi$. If $\mathrm{Prob}_\xi[\xi \in \mathbf{Im}(A)] > 0$, accurately retrieving x becomes impossible.
This changes under additional information on signal x, of the form $x \in \mathcal{X}$, where $\mathcal{X}$ is either a bounded subset of $\mathbf{R}^n$ or unbounded subset of a subspace of $\mathbf{R}^n$.
Under such assumptions, other linear estimates can outperform the least square when looking at regression. For instance, this is so when $\|x\|_2 \leq \sigma \mathrm{Rank}(A)\, \|A\|^{-1}_{\mathbf{Fro}}$ and one considers a properly tuned Ridge estimate (cf (1.17)). Thus, as long as $\mathcal{X}$ is bounded, there are still reasons to consider linear estimates for minimax optimality. In fact, a huge body of the statistical literature focuses on them, starting from [17] (see [18] for a modern point of view), and developed in [19–22] and references therein.

Filtering estimators: first results on optimality

First results of near-optimality were derived for "simple" situations where $A = B = \mathbf{I}$. They demonstrate that the lowest possible maximal risk of a linear estimate is minimax over X, when the latter is a specific type of closed and bounded subset of $\mathbf{R}^n$. The first result of this type dates back to the seminal work of Pinsker [23].

Proposition 1.2.1 ([23], Theorem 2) *Let signal set $\mathcal{X}$ be a Sobolev Ellipsoids described as*

$$\mathcal{X} = \Big\{ z \in \mathbf{R}^n : \sum_{i=1}^{n} i^{2\alpha} z_i^2 \leq L \Big\}, \tag{1.27}$$

and observations of the form $\omega = x + \sigma\xi$ with ξ a normal centered Gaussian vector. Then one has

$$\forall n, \alpha, L, \quad \lim_{\sigma \to 0} \frac{\mathrm{Risk}^{\|\cdot\|_2^2}(\widehat{w}^*_{Lin}, \mathcal{X})}{\mathrm{RiskOpt}^{\|\cdot\|_2^2}(\mathcal{X})} = 1, \tag{1.28}$$

i.e. the best linear estimate is asymptotically the minimax one.

This result extends to the case of diagonal matrices A and B in [24]. A particular sub-class of linear estimates called *filters* has been widely used in linear inverse problems, e.g., in cases where A is a diagonal matrix. Note that one can always transform observation $\omega = Ax + \sigma\xi$, to obtain new *diagonal* observation

$$\widetilde{\omega}_i = \nu_i z_i + \sigma \widetilde{\xi}_i, \quad z := V^T x,\ \widetilde{\omega} = U^T \omega,\ \widetilde{\xi} = U^T \xi \sim \mathcal{N}(0, \mathbf{I}),$$

with U and V defined by *Singular Value Decomposition* (SVD) $U\mathrm{Diag}(\nu)V^T = A$, and non-negative ν_is sorted in decreasing order. This framework is treated in [25–27]. A filtering estimator $\widehat{x}^\lambda_f$ (in finite dimension) is then defined by

$$\widehat{x}^\lambda_f(\omega) = V\widehat{z}^\lambda_f(\omega), \ \left(\widehat{z}^f_\lambda\right)_i(\omega) = \lambda_i \widetilde{\omega}_i. \tag{1.29}$$

In the context of direct observation $\omega = x + \xi$, the filtering estimator associated with *Pinsker* weights

$$\lambda_i = \left(1 - \left(\frac{i}{\beta}\right)^\alpha\right)_+, \quad \alpha, \beta > 0 \tag{1.30}$$

is proven to be minimax optimal [23].
For more general sets (see [28]), linear estimates are also minimax optimal, in the sense that an absolute constant upper bounds the ratio of the best linear estimate's maximal risk to the minimax risk.

Definition 3 *Let $\mathcal{X} \subset \mathbf{R}^n$. $\mathcal{X}$ is*

- *orthosymmetric if for all $x \in \mathcal{X}$, any vector obtained from it by changing the sign of some of its coefficient also belongs to $\mathcal{X}$.*

- *quadratically convex if subset*

$$\left\{(z_i^2)_{i=1}^n \in \mathbf{R}^n, \quad z \in \mathcal{X}\right\} \text{ is convex.}$$

- *A weighted ℓ_p-ball if for some vector $a \in \mathbf{R}^n_+$,*

$$\mathcal{X} = \left\{ z \in \mathbf{R}^n : \quad \sum_{i=1}^n a_i |z_i|^p \leq 1 \right\}.$$

- *An hyperrectangle if for some vector $a \in \mathbf{R}^n_+$,*

$$\mathcal{X} = \left\{ z \in \mathbf{R}^n : \quad \max_i a_i |z_i| \leq 1 \right\}.$$

- *An ellipsoid if for some positive semidefinite matrix $R \in \boldsymbol{S}^n_+$,*

$$\mathcal{X} = \left\{ z \in \mathbf{R}^n : \quad z^T R z \leq 1 \right\}.$$

Ellipsoids, weighted ℓ_p-balls with $p \geq 2$ and hyperrectangles are both orthosymmetric and quadratically convex.

Proposition 1.2.2 ([28], Lemma 6 and Theorem 7) *Let $x \in \mathcal{X}$ where $\mathcal{X}$ is orthosymmetric, convex, compact, and quadratically convex, and $\omega = x + \xi$, $\xi \sim \mathcal{N}(0, \sigma^2 I)$. Then, the linear estimate with lowest maximal risk $\widehat{w}^*_{Lin}$ satisfies*

$$\mathrm{Risk}^{\|\cdot\|_2^2}(\widehat{w}^*_{Lin}, \mathcal{X}) \leq 1.25 \, \mathrm{RiskOpt}^{\|\cdot\|_2^2}(\mathcal{X}). \tag{1.31}$$

The *projective* or *truncated SVD* filtering estimator [29], defined by *projective* weights

$$\lambda_i(j) := \mathbf{1}\{i \leq j\} \sigma_i^{-1}$$

is proven to be minimax over weighted ℓ_p-ball and ellipsoids, for properly selected truncation index i_*.

Juditsky and Nemirovski: ellitopes and optimality of linear estimates

Recent work by A. Juditsky and A. Nemirovski [30, 31] proves results of near minimax optimality for linear estimates, without making specific assumptions about the matrices $A \in \mathbf{R}^{m\times n}$, $B \in \mathbf{R}^{\nu\times n}$, and over a rich class of convex sets called *Ellitopes*.[3]

Definition 4 *A subset $\mathcal{X}$ of $\mathbf{R}^n$ is a basic ellitope if it admits the following description*

$$\mathcal{X} = \left\{ z \in \mathbf{R}^n : \ \exists t \in \mathcal{T} \subset \mathbf{R}^K_+ : \ z^T R_k z \le t_k, \ 1 \le k \le K \right\}, \tag{1.32}$$

where:

- *$R_k \succeq 0$ are $n \times n$ matrices with $\sum_{k=1}^K R_k \succ 0$.*
- *$\mathcal{T}$ is a monotone, nonempty computationally tractable convex compact subset of $\mathbf{R}^K_+$ intersecting the latter's interior.*

First notice that all signal classes studied in [23, 28] are simple examples of ellitopes. Indeed, with $K = n$, $R_k = e_k e_k^T$, where e_k is the k-th element of the canonic base of $\mathbf{R}^n$, and respectively

$$\mathcal{T} = \prod_{k=1}^n [0, a_k^{-2}], \text{ or } \mathcal{T} = \left\{ t \in \mathbf{R}^n_+, \ \|\mathrm{Diag(a)}t\|_{p/2} \le 1 \right\}, \ p \ge 2,$$

X is an hyperectangle or a weighted ℓ_p body.
Authors of [30, 31] prove that the maximum of a quadratic form over ellitopes can be tightly upper-bounded by an efficiently minimizable function, using *Semidefinite* relaxation.

Proposition 1.2.3 ([32], Proposition 4.6) *Let $\Gamma \in \mathbf{S}^n_+$ and ellitope $\mathcal{X}$ as in (1.32), and consider the following two optimization problems*

$$\begin{array}{rcl} \mathrm{Opt}_*(\Gamma) & := & \max\limits_{x\in\mathcal{X}} x^T\Gamma x, \\ \mathrm{Opt}(\Gamma) & := & \min\limits_{\lambda\ge 0} \left\{ \phi_{\mathcal{T}}(\lambda) : \ \ \sum_k \lambda_k R_k \succeq \Gamma \right\}, \ \textit{with } \phi_{\mathcal{T}}(\lambda) := \max\limits_{t\in\mathcal{T}} t^T\lambda. \end{array}$$

The second problem is efficiently computable, and is a tight upper bound on $\mathrm{Opt}_*(\Gamma)$, in the sense that

$$\mathrm{Opt}_*(\Gamma) \le \mathrm{Opt}(\Gamma) \le 3\ln(\sqrt{3}K)\mathrm{Opt}_*(\Gamma). \tag{1.33}$$

Building on this property, Theorem 2.1 in [30] states that the best linear estimate $\widehat{w}^{H_*}(\omega)$ one can build using the outlined minimization procedure is nearly-minimax optimal:

Proposition 1.2.4 ([32], Proposition 4.4 and 4.5) *Let $\mathcal{X}$ be an ellitope as in (1.32). Assuming $\xi \sim \mathbf{P}_\xi = \mathcal{N}(0, \sigma^2 \boldsymbol{I})$, let H_* be defined as a solution to the convex optimization problem*

$$\mathrm{Opt} = \min_{H,\lambda} \left\{ \phi_{\mathcal{T}}(\lambda) + \sigma^2 \boldsymbol{Tr}(\Theta) : \ \left[\begin{array}{c|c} \sum_k \lambda_k R_k & B^T - A^T H \\ \hline B - H^T A & I \end{array} \right] \succeq 0, \ \Theta \succeq H^T H, \ \lambda \ge 0 \right\}. \tag{1.34}$$

[3] These results are actually extended in [31] for signal sets described as *spectratopes*, a class of convex sets including ellitopes.

In terms of maximal risk

$$\mathrm{Risk}^{\|\cdot\|_2}(\widehat{w}^{H_*}_{Lin},\mathcal{X}) := \sup_{x\in\mathcal{X}}\left\{\mathbf{E}_\xi\left[\left\|\widehat{w}^{H_*}_{Lin}(\omega)-Bx\right\|_2^2\right]^{1/2}\right\},$$

linear estimate $\widehat{w}^{H_*}_{Lin}$ *is nearly-minimax in the sense that*

$$\mathrm{Risk}^{\|\cdot\|_2}(\widehat{w}^{H_*}_{Lin},\mathcal{X}) \le \sqrt{\mathrm{Opt}} \le 192\sqrt{5\ln(2)\ln(32K)}\,\mathrm{RiskOpt}^{\|\cdot\|_2}(\mathcal{X}), \tag{1.35}$$

where $\mathrm{RiskOpt}^{\|\cdot\|_2}(\mathcal{X})$ *is the minimax risk associated with* $\mathrm{Risk}^{\|\cdot\|_2}(\cdot,\mathcal{X})$.

These two results are extended to the case of more general recovery norm $\|\cdot\|$, which is said to be *co-ellitopic*.

Definition 5 *A norm* $\|\cdot\|$ *on* $\mathbf{R}^\nu$ *is said to be co-ellitopic if its associated dual norm* $\|\cdot\|_*$ *is such that the associated ball of radius* 1, $B_{\|\cdot\|_*}(1)$, *is a basic ellitope:*

$$\mathcal{B}_{\|\cdot\|_*} := \{z\in\mathbf{R}^\nu : \|z\|_*\le 1\} = \left\{z\in\mathbf{R}^\nu :\ \exists t\in\mathcal{W}\subset\mathbf{R}^L_+ :\ z^TW_lz\le t_l,\ 1\le l\le L\right\}. \tag{1.36}$$

Typical co-ellitopic norms are p-absolute norm for $1\le p\le 2$, $\|\cdot\|_\Sigma$ for $\Sigma\in\mathbf{S}^n_+$ and the discrete Wasserstein ℓ_1 distance (see exercise 4.66 of [32]). With $\|\cdot\|$ satisfying this requirement, the following proposition extends the near-optimality of linear estimates:

Proposition 1.2.5 ([32], Proposition 4.16) [4] *Let* $\mathcal{X}$ *be an ellitope described as in* (1.32) *and* $\|\cdot\|$ *a co-ellitopic norm as in* (1.36). *Assuming* $\xi\sim\mathbf{P}_\xi=\mathcal{N}(0,\sigma^2\boldsymbol{I})$, *let* H_* *be defined as a solution to the convex optimization problem*

$$\mathrm{Opt} = \min_{H,\Theta,\lambda,\mu}\left\{\phi_{\mathcal{T}}(\lambda)+\phi_{\mathcal{W}}(\mu)+\sigma^2\boldsymbol{Tr}(\Theta):\ \begin{array}{c}\left[\begin{array}{c|c}\sum_k\lambda_kR_k & B^T-A^TH\\ \hline B-H^TA & \sum_l\mu_lW_l\end{array}\right]\succeq 0,\\ \Theta\succeq H^TH,\ \lambda\ge 0,\mu\ge 0\end{array}\right\}. \tag{1.37}$$

Linear estimate $\widehat{w}^{H_*}_{Lin}$ *is nearly-minimax in the sense that*

$$\mathrm{Risk}^{\|\cdot\|}(\widehat{w}^{H_*},\mathcal{X}) \le \mathrm{Opt} \le 192\sqrt{\ln(32K)\ln(32L)}\mathrm{RiskOpt}^{\|\cdot\|}(\mathcal{X}). \tag{1.38}$$

Notice that for $\mathcal{X}$ which is an ℓ_q-body with $q\ge 2$ and $\|\cdot\| = \|\cdot\|_p$ with $1\le p\le 2$, ellitopic dimensions K and L are respectively equal to n and ν. This shows that sub-optimality of linear estimates depends only weakly on dimensions n,ν.

1.2.2 Nonlinear estimation and polyhedral estimates

Non-optimality of linear estimates

Linear estimates are a simple yet powerful tool. However, for some sets, they can be heavily sub-optimal. Consider the task of estimating x from direct observation

$$\omega = x+\sigma\xi,\quad \xi\sim\mathcal{N}(0,\mathbf{I}),\quad x\in\mathcal{B}_1(1).$$

[4] Note that this result is originally stated for Spectratopic set $\mathcal{X}$ and $\mathcal{B}_{\|\cdot\|_*}$, but we provide the restricted ellitope version to keep light notations

It can be easily shown that the best linear estimate is $\widehat{w}^h(\omega) := h\omega$, with $h = (1+n\sigma^2)^{-1}$, and has associated risk

$$\mathrm{Risk}^{\|\cdot\|_2^2}(\widehat{w}^h, \mathcal{B}_1(1)) = \frac{n\sigma^2}{1+n\sigma^2}. \tag{1.39}$$

This is exactly the same risk as for $x \in \mathcal{B}_2(1)$, yet, in this situation, it is no longer minimax. Indeed, a simple non linear estimator can have a much smaller risk. Let us denote

$$\widehat{w}_\infty(\omega) := \underset{z\in\mathcal{B}_1(1)}{\mathrm{argmin}} \ \|z-\omega\|_\infty . \tag{1.40}$$

As demonstrated in [33] that when $\sigma \le 2n/\sqrt{e}$, the risk of $\widehat{w}_\infty$ is

$$\mathrm{Risk}^{\|\cdot\|_2^2}(\widehat{w}_\infty, \mathcal{X}) \le 8\sqrt{2\ln(2n/\sigma)}\sigma. \tag{1.41}$$

For instance, when $\sigma = \frac{1}{2\sqrt{n}}$, the risk of the best linear estimate equals $1/5$, while $\widehat{w}_\infty$'s is less than $16\sqrt{3\ln(en)/n}$, implying suboptimality of the linear estimate in that situation.

Early results of nearly-minimax estimation with nonlinear estimators

In the case of direct observation $\omega = x + \sigma\xi$, one can consider thresholding estimators of the form

$$\forall i \in [1:n], \ \ \widehat{x}_i^t(\omega) = \omega_i t_i(\omega_i),$$

with $t_i : \mathbf{R} \to [0,1]$ a nonlinear threshold rule. Two common choices for t are hard-thresholds and soft-thresholds [6, 8, 34], respectively defined for $\lambda > 0$ by

$$t_h^\lambda(y) := y\mathbf{1}\{|y| > \lambda\}$$

and

$$t_s^\lambda(y) := (|y|-\lambda)_+ \frac{y}{|y|}.$$

In [8], authors show that these two type of estimators are nearly-minimax optimal over Besov spaces for Gaussian ξ [5]. They also prove that linear estimates are not nearly-minimax over certain types of Besov spaces.

Polyhedral Estimate, a nearly minimax estimate

Authors in [33] extend operational and almost "assumption-free" theory of linear estimation developed in [30, 31] to nonlinear estimation. They do so by building on "simple" estimate (1.40) to propose a novel analysis and computation procedure of the *Polyhedral estimate*, stemming from the early ideas of [35, 36]. What motivates these estimates is the fact that, for a given $h \in \mathbf{R}^m$, linear form $h^T\omega$ accurately estimate h^TAx. More precisely, given $h \in \mathbf{R}^m$, when $\xi \sim \mathcal{SG}(0, \sigma^2 I)$, one has

$$\mathrm{Prob}_\xi\left[|h^T\omega - h^TAx| \le \sigma\|h\|_2\sqrt{2\ln(2/\epsilon)}\right] \ge 1-\epsilon. \tag{1.42}$$

[5] See [8] for a proper definition of Besov spaces

Thus, selecting N columns h_i such that $\|h_i\|_2 = \left(\sigma\sqrt{2\ln(2N/\epsilon)}\right)^{-1}$, one can build the closed, convex subset of $\mathbf{R}^n$

$$\mathcal{Z} := \left\{z \in \mathbf{R}^n : \quad \forall i \le N, \quad |h_i^T[\omega - Az]| \le 1\right\}.$$

Note that by construction, $x \in \mathcal{Z}$ with probability greater than $1-\epsilon$. The key idea is then to select appropriate columns h_i so that any point $\widehat{x} \in \mathcal{Z} \cap \mathcal{X}$ results in a good estimate of Bx when using $\widehat{w} := B\widehat{x}$. This is the argument that motivates *polyhedral* estimate

$$\widehat{w}^H_{\rm Poly}(\omega) := B\widehat{x}^H_{\rm Poly}(\omega), \quad \widehat{x}^H_{\rm Poly}(\omega) \in \operatorname*{Argmin}_{z\in\mathcal{X}} \left\|H^T[\omega - Az]\right\|_\infty. \tag{1.43}$$

Authors in [33] provide two operational approaches for computing "presumably good" contrast parameter H_* and tight upper bound on its associated ϵ-risk, both via convex optimization. In some special cases, it is also shown to be nearly-minimax. Specifically, first technique of construction selects $N = \nu$ columns by solving the ν optimization problems

$$\mathrm{Opt}_i = \min_g \left\{\max_{z\in\mathcal{X}} \left(\left[\mathrm{Row}_i(B) - A^T g\right]^T x\right) + \sigma\sqrt{2\ln(2\nu/\epsilon)}\,\|g\|_2\right\} \tag{1.44}$$

and setting $\mathrm{Col}_i(H) = (\sigma\sqrt{2\ln(2\nu/\epsilon)}\,\|g_i\|_2)^{-1} g_i$, with g_i being a solution to Opt_i. This construction is shown to produce nearly-optimal polyhedral estimates for $\mathcal{X}$ as a weighted ℓ_q-ball, A and B diagonal and $\|\cdot\| = \|\cdot\|_r$ with $r \ge q$.
On the other hand, for any matrices A, B, ellitope $\mathcal{X}$ and co-ellitopic $\|\cdot\|$, authors propose the following construction, which is also nearly-minimax.

Proposition 1.2.6 ([32], Proposition 5.8 and 5.10) *Let X be an ellitope as in (1.32) and let $\|\cdot\|$ be a co-ellitopic norm (1.36). Consider the convex optimization problem*

$$\mathrm{Opt} = \min_{\mu,\lambda,\Theta} \left\{ \Phi_{\mathcal{T}}(\lambda) + \Phi_{\mathcal{W}}(\mu) + 2\sigma^2\ln(\tfrac{2m}{\epsilon})\,\boldsymbol{Tr}(\Theta) : \begin{array}{l} \left[\begin{array}{c|c} \sum_l \mu_l W_l & B/2 \\ \hline B^T/2 & \sum_k \lambda_k R_k + A^T\Theta A \end{array}\right] \succeq 0, \\ \Theta \succeq 0, \lambda \ge 0, \mu \ge 0. \end{array} \right\} \tag{1.45}$$

Let $(\Theta_*, \lambda_*, \mu_*)$ be an optimal solution to the aforementioned problem. Selecting

$$H_* = \frac{1}{\sigma\sqrt{2\ln(2m/\epsilon)}}\Gamma \in \mathbf{R}^{m\times m},$$

with $\Gamma^T S\Gamma$ being the SVD of Θ_*, yields a polyhedral estimate such that

$$\mathrm{Risk}^{\|\cdot\|}(\widehat{w}^{H_*}_{Poly}, \mathcal{X}) \le \mathrm{Opt}. \tag{1.46}$$

Moreover, for $\epsilon \in (0, \frac{1}{8}]$, this estimate is nearly minimax, in the sense that

$$\mathrm{Risk}^{\|\cdot\|}_\epsilon(\widehat{w}^H_{Poly}, \mathcal{X}) \le 48\sqrt{2\ln(2m/\epsilon)\ln(32K)\ln(32L)}\mathrm{RiskOpt}^{\|\cdot\|}_\epsilon(\mathcal{X}). \tag{1.47}$$

1.2.3 Contributions to robust linear and nonlinear estimation for linear inverse problems

In Chapter 2 and 3, we discuss adaptive linear and polyhedral estimates when observing noisy linear transformations of a signal known to belong to an ellitope, under various forms of uncertainty in the model $\mathfrak{m}$. These results are thus in direct continuation of [30–33], and based on working paper

Adaptive counterpart of linear inverse problems.
The first contribution of chapter 2 is a routine to produce "presumably good" linear estimates w.r.t. ϵ-risk. We do so by selecting associated matrix parameter H as a solution of a convex minimization problem.
We then proceed to build adaptive linear estimates under the "general" model of uncertainty where

$$\mathfrak{m}(\delta) = \{\mathcal{Z}(\delta), A, B, \|\cdot\|, \mathbf{P}_\xi\}, \quad \delta \in [0, \overline{\delta}], \quad \overline{\delta} \in \mathbf{R}_{+*} \tag{1.48}$$

i.e., set $\mathcal{Z}$ such that signal $x \in \mathcal{Z}$ is uncertain. We thus assume that $\mathcal{Z}$ belongs to a parametric family of sets $(\mathcal{Z}(\delta))_{\delta \geq 0}$ increasing [6] with parameter δ.
Given an integer I, we build a finite family of sets

$$\mathcal{Z}(\delta_0) \subset \mathcal{Z}(\delta_1) \subset \cdots \subset \mathcal{Z}(\delta_I)$$

and associated provably best linear and polyhedral estimates for each of them. We then show how Lepski's adaptive procedure [5, 10, 11] allows one to select a linear estimate, resulting in nearly-optimal performances. We also propose an adaptive polyhedral estimate built using contrast parameters associated to specific δ_i, also leading to nearly-optimal ϵ-risk. Finally, specific results are provided for the special case where $\mathcal{Z}(\delta) = \mathcal{X} + \mathcal{U}(\delta)$, with, for $\delta \geq 0$, $\mathcal{X}$ and $\mathcal{U}(\delta)$ being ellitopes, and recovery norm being either co-ellitopic or $\|\cdot\|_2$. In the second case, we provide an additional aggregation routine involving estimation of quadratic forms.
In chapter 3, two specific settings falling into the aforementioned framework are then presented as applications.

- The first, rather general, is adaptivity to unknown parameter δ when $\mathcal{Z}(\delta) = \mathcal{X} + \delta\mathcal{U}$, where $\mathcal{X}$ and $\mathcal{U}$ are ellitopes.
- The second setting deals with robustness to s-sparse data-corruption. Here, we assume to be given the observation

  $$\omega = Ax + u + \xi, \quad x \in \mathcal{X}, \ \|u\|_0 \leq s, \tag{1.49}$$

  where $\mathcal{X} \in \mathbf{R}^n$ is an ellitope, and s is unknown. Deterministic vector u is an unknown "nuisance" that has to be considered *adversarial*, i.e. that maximizes our error of estimation when targeting $w = Bx$. As a consequence, we consider the following maximal risk

  $$\mathrm{Risk}^{\|\cdot\|}(\widehat{w}, \mathcal{X} \times \mathcal{B}_0(s)) := \max_{x \in \mathcal{X}} \max_{u \in \mathcal{B}_0(s)} \mathbf{E}_\xi \left[\|\widehat{w}(Ax + u + \xi) - Bx\|\right]. \tag{1.50}$$

Under the first item's setting , we propose comprehensive versions of the linear and polyhedral estimates, i.e. estimates with guaranteed upper bound r on the ϵ-risk when $\delta = 0$, and such that

$$\mathrm{Risk}_\epsilon[\widehat{w}|\mathcal{Z}(\delta)] \leq r + \delta t.$$

We focus on efficient routines allowing for computation of the minimal t such that the latter holds when in context of $\mathfrak{m}(\delta)$.

[6] w.r.t. the inclusion relationship for sets, i.e.

$$\delta_1 \leq \delta_2 \implies \mathcal{Z}(\delta_1) \subset \mathcal{Z}(\delta_2).$$

Finally, we provide numerical experiments that illustrate results about adaptive linear and polyhedral estimates when concerned with the uncertainty $\mathcal{X} + \delta\mathcal{U}$.

Chapter 4, based on the paper *First order algorithms for computing linear and polyhedral estimates* [37] submitted to *Open Journal of Mathematical Optimization* (OJMO), deals with efficient computation of linear and polyhedral estimates.
When considering ellitopic $\mathcal{X}$ and co-ellitopic recovery norm $\|\cdot\|$, design of these estimates relies upon building a good contrast parameter H. This in turn involves solving a semidefinite optimization problems with dense matrix variables, e.g., as in (1.45). Using *Interior Point Semidefinite Programming* to do so severely limits how big dimensions n, m, ν can be. Indeed, formulation of said problems involves matrix variables of the size $\mathcal{O}\left(((n+\nu)\bigvee m)^2\right)$, entailing computational complexity of the order $\mathcal{O}\left(((n+\nu)\bigvee m)^6\right)$.
As a remedy, we propose the following three contributions.

- We first provide a reformulation of the optimization problem (1.45). The latter replaces previously mentioned matrix variable by vector ones of size $\mathcal{O}(K+L)$, and allows for the use of first order methods.

- We show how to augment feasible solution (λ, μ, Θ) of (1.45), with associated value $\mathfrak{p}[\Theta]$, to variables $(2\lambda, \mu, \Theta, H)$ feasible for (1.37) with associated value $\mathfrak{p}[\Theta]$. A minimal solution to (1.45) can then be converted to an almost minimal solution of (1.37), in the sense that value of the converted solution is within a factor

$$\max\left\{2; \sqrt{1+\ln(2m)}\right\}$$

  of problem (1.37)'s minimum.

- We finally develop and analyse a novel bundle-like algorithm, *Composite Truncated Level* (CTL), that can treat the aforementioned reformulation of (1.45). The computational cost of one step of CTL is dominated by the one involved in inversion of square matrices of dimension ν and n and/or computation of SVDs of square matrices of size m. We support our analysis with numerical experiments.

Chapter 5 is based on papers *Robust signal recovery under uncertain-but-bounded perturbations in observation matrix* [38], *Estimation from indirect observations under stochastic uncertainty in observation matrix* [39] submitted to "Journal of Optimization Theory and Application". Here we consider the task of building linear and polyhedral estimates which are robust w.r.t. observation matrix uncertainty. More precisely, we assume that we are in the situation where

$$\mathfrak{m}(\eta) = \{\mathcal{X}, A[\eta], B, \|\cdot\|, \mathbf{P}_\xi\}, \quad A[\eta] := A + \sum_{\alpha=1}^{q} \eta_\alpha A_\alpha, \tag{1.51}$$

with given matrices A and $(A_\alpha)_{\alpha=1}^q$. We then build almost best linear and polyhedral estimates when considering the maximal ϵ-risk over all $x \in \mathcal{X}$ and all possible observation matrices. We do so by defining their associated contrast matrices H_* as minimizers of tight upper bounds of the ϵ-risk of respectively linear or polyhedral estimate with contrast H. We consider separately matrix uncertainty of two types.

- First, we assume that, η is a centered random vector, and we build linear and polyhedral estimates which are robust to such random uncertainty. Performance is then measured through

$$\mathrm{Risk}_\epsilon^{\|\cdot\|}(\widehat{w},\mathcal{X}) := \sup_{x\in\mathcal{X}} \inf_{\rho>0} \left\{\rho : \ \mathrm{Prob}_{\xi,\eta}\left[\|\widehat{w}-Bx\| > \rho\right] \le \epsilon\right\}.$$

  This is done for sub-Gaussian η, and for η with bounded second moment when repeated observations are available.

- Next, we consider the case of η which is deterministic but unknown, and belongs to a bounded convex compact subset $Q \subset \mathbf{R}^q$. The considered risk is

$$\mathrm{Risk}_\epsilon^{\|\cdot\|}(\widehat{w},\mathcal{X}) := \sup_{x\in\mathcal{X},\eta\in Q} \inf_{\rho>0} \left\{\rho : \ \mathrm{Prob}_{\xi}\left[\|\widehat{w}-Bx\| > \rho\right] \le \epsilon\right\}.$$

1.3 Sparse recovery

1.3.1 Minimax results on sparse recovery

We now consider the situation of (1.2) with $m < n$ and x such that it is s-*sparse*, i.e. $\|x\|_0 = s$. If $m > s$, we retrieve in a sense the classical setting of estimation from linear observation. With access to oracle

$$I := \{i \in [1:n] : \ \ x_i \neq 0\}, \quad |I| \le s, \tag{1.52}$$

one can reformulate problem (1.2) as

$$\omega_I = A_I x_I + \xi_I, \tag{1.53}$$

with $A_I \in \mathbf{R}^{m\times s}$ is obtained from A by concatenating columns with indexes in I, and $x_I \in \mathbf{R}^s$ is the vector obtained from x when removing zero coefficients. The corresponding least square estimate enjoys average Euclidean error of regression

$$\sigma^2 \frac{\mathrm{Rank}(A_I)}{m} \le \frac{\sigma^2 s}{m}. \tag{1.54}$$

Yet, this is an idealistic situation, as there would be no sense in assuming that $x \in \mathbf{R}^n$ if one actually knows I. Therefore, the challenge lies in accurately guessing the support of x in order to estimate it.
The first approach is to constrain least square to s-sparse vectors, resulting in estimate

$$\widehat{x}^s_{c,0}(\omega) \in \underset{z\in\mathcal{B}_0(s)}{\mathrm{Argmin}} \ \|\omega - Az\|_2^2, \tag{1.55}$$

or its penalized counterpart

$$\widehat{x}^\lambda_{p,0}(\omega) \in \underset{z\in\mathbf{R}^n}{\mathrm{Argmin}} \ \|\omega - Az\|_2^2 + \lambda\|z\|_0. \tag{1.56}$$

When prediction error is concerned, these estimates enjoy the following upper bound (see chapter 2.4 in [40]) on their maximal risk under the assumption of s-sparsity:

$$\begin{aligned} \mathrm{Risk}^{\|\cdot\|_{2,m}^2}(A\widehat{x}^s_{c,0}, \mathcal{B}_0(s)) \le & \quad 32\tfrac{s\sigma^2}{m}\ln(en/s), \\ \mathrm{Risk}^{\|\cdot\|_{2,m}^2}(A\widehat{x}^{\lambda_0}_{p,0}, \mathcal{B}_0(s)) \le & \quad 16\tfrac{\sigma^2}{m}(s\ln(e^2n)+1), \qquad \lambda_0 = \tfrac{8\sigma^2}{m}\ln(2n). \end{aligned} \tag{1.57}$$

The extra factor $\ln(en/s)$ compared to (1.54) seems to be optimal in the minimax sense in some situations, as pointed out by results in [41] stating that

$$\mathrm{RiskOpt}^{\|\cdot\|_2^2}\left(\mathcal{B}_0(s)\right) \gtrsim \frac{n\sigma^2 s}{\|A\|_{\mathbf{Fro}}^2}\left(\ln(n/s) \bigvee 1\right). \tag{1.58}$$

As a consequence, (1.55) and (1.56) are minimax optimal in the direct observation model (i.e. (1.2) with $A = \mathbf{I}$ and $n = m$). The main drawback of the estimates (1.55) and (1.56) is that the underlying minimization problems quickly become intractable as n grows. This limitation arises because the $\|\cdot\|_0$ norm is not convex, and has motivated similar approaches that replace it with $\|\cdot\|_1$.

1.3.2 Lasso, Dantzig estimates and design conditions

This choice is justified as the latter is the convex envelope of the former on $\mathcal{B}_\infty(1)$. This motivates the celebrated *Lasso* ([42])

$$\widehat{x}_{\mathrm{L}}^{\lambda} \in \underset{z}{\mathrm{Argmin}} \left\{ \frac{1}{2m}\|\omega - Az\|_2^2 + \lambda\|z\|_1 \right\}, \tag{1.59}$$

and the *Dantzig selector* introduced in [43],

$$\widehat{x}_{\mathrm{D}}^{\mu} \in \underset{z}{\mathrm{Argmin}} \left\{ \|z\|_1 : \ \frac{1}{m}\left\|A^T[\omega - Az]\right\|_\infty \leq \mu \right\}. \tag{1.60}$$

For both estimates, the appropriate choice of parameter λ or μ is crucial to ensure good performance. In practical applications, this is usually done using cross-validation. In situations where $\xi \sim \mathbf{S}\mathcal{N}(0, \sigma^2 I)$ and A satisfies some conditions, theoretical analysis advocates that they be of order $\sigma\sqrt{2\ln(n)/m}$.

The null space property

Historically, the problem of sparse recovery was first addressed with noiseless observations [44]. In this case, it is not trivial that

$$\forall x \in \mathcal{B}_0(s), \quad x = \underset{z}{\mathrm{argmin}} \ \{\|z\|_1 : \ \ Az = Ax\} \tag{1.61}$$

holds without the design matrix A satisfying some property involving the level of sparsity s. The necessary and sufficient condition on A for (1.61) to hold is the celebrated *Null-Space Property* (NSP) [45].

Definition 6 *Let $0 < \kappa < \frac{1}{2}$. A matrix A is said to possess the Null-Space property NSP(s,κ) if the following holds true:*

$$\forall z, \ Az = 0 \implies \|z_s\|_1 \leq \kappa\|z\|_1. \tag{1.62}$$

Using the following equivalent property, it is possible to derive a risk upper bound for the Lasso and Dantzig selector.

Proposition 1.3.1 ([32], Section 1.2.2.1) *Let $0 < \kappa < \frac{1}{2}$. The NSP(s,κ) for a given matrix A is equivalent to existence of $C_{\|\cdot\|}$ such that the following holds:*

$$\forall z, \ \ \|z_s\|_1 < \kappa \|z\|_1 + C_{\|\cdot\|}(s,\kappa) \|Az\| . \tag{1.63}$$

For $\|\cdot\| = \|\cdot\|_\infty$,

$$C_{\|\cdot\|_\infty}(s,\kappa) \leq \frac{\kappa n + s}{\rho(A)}, \quad \rho(A) := \min \ \ \{\rho : \rho \in \boldsymbol{Sp}(A), \ \ \rho > 0\}$$

Assuming that A^TA satisfies the $\mathbf{NSP}(s,\kappa)$ with $C_{\|\cdot\|_\infty}(s,\kappa)$, that $\xi \sim \mathbf{S}\mathcal{G}(0,\sigma^2 I)$ and that $\max_i \|\mathrm{Col}_i(A)\|_2 \leq \sqrt{m}$, choice $\mu = \sigma\sqrt{\frac{2}{m}\ln\left(\frac{2n}{\epsilon}\right)}$ yields the following upper bound (see Chapter 1 of [32])

$$\mathrm{Risk}_\epsilon^{\|\cdot\|_1}(\widehat{x}_{\mathrm{D}}^{\mu}, \mathcal{B}_0(s)) \leq \frac{4m\mu C_{\|\cdot\|_\infty}(s,\kappa)}{1-2\kappa} \leq \frac{4(\kappa n + s)\sigma\sqrt{\frac{2}{m}\ln\left(\frac{2n}{\epsilon}\right)}}{(1-2\kappa)\rho(A^TA/m)}. \tag{1.64}$$

Provided that

$$\frac{mC_{\|\cdot\|_\infty}(s,\kappa)}{1-2\kappa} \lesssim \mathcal{O}\left(\frac{s}{\sqrt{\rho(A^TA)}}\right),$$

first inequality of (1.64) implies the rate

$$\mathrm{Risk}_\epsilon^{\|\cdot\|_1}(\widehat{x}_{\mathrm{D}}^{\mu}, \mathcal{B}_0(s)) \lesssim \mathcal{O}\left(\sigma s\sqrt{\frac{\ln\left(\frac{2n}{\epsilon}\right)}{m\rho(A^TA)}}\right). \tag{1.65}$$

Notice that (1.57), and $\widehat{x}_{c,0}^{s} - x \in \mathcal{B}_0(2s)$ implies that

$$\mathrm{Risk}^{\|\cdot\|_1}(\widehat{x}_{c,0}^{s}, \mathcal{B}_0(s)) \lesssim \sigma s\sqrt{\frac{\ln(n)}{m\rho(A^TA)}}.$$

Thus, A^TA satisfying the NSP with good constants provides a rate for ℓ_1 sparse recovery that is essentially as good as ℓ_0 minimization routines.
Nevertheless, checking whether a given matrix A^TA satisfies the NSP for s-sparse signals and computing $s, \kappa, C_{\|\cdot\|_\infty}$ is an algorithmic task that suffers from computational complexity of $\mathcal{O}(n^s)$. Second, it is not known how to compute matrix A that verifies the NSP(s,κ).

The Restricted Isometry Property

The celebrated work of Candes and Tao [46] introduces the following condition on the design matrix A:

Definition 7 *A satisfies the Restricted Isometry Property RIP(s,δ), if, for a given sparsity level s, it exists $\delta \in (0,1)$ such that*

$$\forall x \in \mathbf{R}^n, \ \ \|x\|_0 \leq s, \implies (1-\delta)\|x\|_2^2 \leq \|Ax\|_2^2 \leq (1+\delta)\|x\|_2^2 . \tag{1.66}$$

The first important fact about RIP is that it implies NSP with constants that ensure good recovery properties of ℓ_1 methods.

Proposition 1.3.2 ([47], Lemma 2.2 and Theorem 1.2) *Let matrix A satisfy the RIP(s, δ_{2s}) condition with $\delta_{2s} < \sqrt{2} - 1$. Then A satisfies the NSP(s, κ) with*

$$\kappa = \frac{\sqrt{2}\delta_{2s}}{1 + (\sqrt{2} - 1)\delta_{2s}}. \tag{1.67}$$

Moreover, under sub-Gaussian noise, when defining the Dantzig selector as

$$\widehat{x}_D \in \underset{z}{Argmin} \left\{ \|z\|_1 : \ \|Az - \omega\|_2 \leq \sigma_*(\sqrt{m} + \sqrt{2\ln(1/\epsilon)}) \right\},$$

the following holds with probability greater than $1 - \epsilon$ for every s-sparse vector x:

$$\|\widehat{x}_D - x\|_2 \leq \frac{2\sqrt{1 + \delta_{2s}}}{1 + (\sqrt{2} - 1)\delta_{2s}} \sigma_*(\sqrt{m} + \sqrt{2\ln(1/\epsilon)}). \tag{1.68}$$

Furthermore, provided m is sufficiently large, an observation matrix with coefficients drawn at random is very likely to satisfy RIP. The good news being that how large m has to be is actually almost proportional to level of sparsity s, as stated by next proposition.[7]

Proposition 1.3.3 *Let $A \in \mathbf{R}^{m \times n}$ be a random matrix with entries $a_{i,j} \sim \mathcal{N}(0, 1/m)$ or $\text{Prob}[\sqrt{m}a_{i,j} = 1] = \text{Prob}[\sqrt{m}a_{i,j} = -1] = 1/2$. Then A satisfies RIP(s, δ_s) with probability at least $1 - \epsilon$ as long as*

$$m \geq \frac{48}{\delta^2} \left[s \ln(9en/s) + \ln(2/\epsilon) \right]. \tag{1.69}$$

Thus, design matrix A satisfies NSP(s, κ), $\kappa < 1/2$, when it verifies RIP$(2s, 0.4)$. Taking $m \geq 600s\lceil \ln(13en/s) \rceil + 1799$ is sufficient for both this to be true and

$$\|\widehat{x}_D - x\|_2 \leq 50\sigma \left(\sqrt{s \ln(36n/s)} + 1 \right)$$

to hold simultaneously, with probability greater than 0.99. Notice that before $n = 9429$, their is no such m even for $s = 1$[8].
Second, as with the NSP, one cannot check whether a given matrix satisfies RIP in less than $\mathcal{O}(n^s)$ operations.

The Restricted Eigenvalue Condition

Under a second condition with essentially the same advantages and weaknesses as RIP, [48] provides upper bounds on the ϵ-risk of the Lasso and Dantzig selector.

Definition 8 *A satisfies the Restricted Eigenvalue condition RE(s, κ, δ), if, for a given sparsity level s, it exists $\delta > 0$ and $\kappa > 0$ such that*

$$\forall I \subset [1 : n], \ |I| \leq s, \quad \|x_{I^c}\|_1 \leq \delta \|x_I\|_1 \implies \|Ax\|_2^2 \geq \kappa \|x_I\|_2^2. \tag{1.70}$$

[7] Similar results are also available when A is created by selecting randomly m rows of an orthonormal matrix, with m sufficiently large w.r.t. s,.

[8] It is important to mention that despite this, in practice, ℓ_1 minimization methods seem to work efficiently even in small dimensions with random design and very few measurements.

This condition is also shown to hold with overwhelming probability for the same type of random matrices as for RIP, as long as $m \gtrsim s \ln(en/s)$ (see [49]). Similar to RIP, checking whether a matrix satisfies the RE condition is also of computational complexity $\mathcal{O}(n^s)$.
Under this condition, [50][9] proposes the following upper bound on the Lasso's ϵ-risk with a parameter choice independent of ϵ.

Proposition 1.3.4 ([50], Proposition 3.2) *Let A satisfy RE(s, κ, δ) and such that all diagonal elements of $\frac{1}{m}A^TA$ are smaller than 1, and $\lambda = \frac{2\sigma\delta}{\delta-1}\sqrt{\frac{2\ln(n)}{m}}$, $\delta \in (0,1)$. Then,*

$$\forall \epsilon \in (0,1), \quad \mathrm{Risk}_\epsilon^{\|\cdot\|_2^2}(A\widehat{x}_L^\lambda, \mathcal{B}_{\|\cdot\|_0}(s)) \leq \frac{\sigma^2}{m}\left(\sqrt{s}\left(\frac{\delta\sqrt{2\ln(n)}}{\kappa(\delta-1)}+1\right)+2.8+\sqrt{2\ln(1/\epsilon)}\right)^2. \quad (1.71)$$

Regarding the Dantzig selector, [48] proposes the following.

Proposition 1.3.5 ([48], Theorem 7.1) *Let A satisfy RE$(s, \kappa, 1)$ and such that all diagonal elements of $\frac{1}{m}A^TA$ are smaller than 1, and $\mu = \sigma\sqrt{2\ln(n/\epsilon)/m}$, $\epsilon \in (0,1)$. Then,*

$$\begin{aligned} \mathrm{Risk}_\epsilon^{\|\cdot\|_2^2}(A\widehat{x}_D^\mu, \mathcal{B}_{\|\cdot\|_0}(s)) &\leq \frac{32\sigma^2 s}{m\kappa^2}\ln\left(\frac{n}{\epsilon}\right), \\ \mathrm{Risk}_\epsilon^{\|\cdot\|_1}(\widehat{x}_D^\mu, \mathcal{B}_{\|\cdot\|_0}(s)) &\leq \frac{8\sigma s}{\kappa^2}\sqrt{\frac{2\ln(n/\epsilon)}{m}}. \end{aligned} \quad (1.72)$$

Note that [48] also proposes similar results for the Lasso with $\lambda = \mathcal{O}\left(\sigma\sqrt{ln(n/\epsilon)/m}\right)$[10].

Computational complexity of computing Lasso and Dantzig selector

For both the Dantzig selector and the Lasso, generating the associated estimate has computational complexity that scales poorly with increasing parameter n. Solving (1.60) relies upon IPM, which scales to at least $\mathcal{O}(n^3)$ by using at least one Newton step. Among the popular methods to solve (1.59) are Iterative Shrinkage-Thresholding Algorithms (ISTA) [51] and Fast ISTA [52]. Both aim to minimizing the function

$$F_\lambda(z) := \|Az-\omega\|_2^2 + \lambda\|z\|_1, \quad z \in \mathbf{R}^n,$$

and use first order algorithms involving the computation of a subgradient of $F_\lambda(z)$. Thus each step of either ISTA or FISTA scales as $\mathcal{O}(nm)$. Authors in [52] proposes the two following result on ISTA and FISTA.

Proposition 1.3.6 ([52], Theorem 3.1) *Let $(x_k)_{k\geq 0}$ be the sequence of iterate produced by ISTA with starting point x_0, and x_* a minimizer of F_λ. Then*

$$\forall k \geq 1, \quad F_\lambda(x_k) - F_\lambda(x_*) \leq \left\|A^TA\right\|_{\boldsymbol{Fro}} \frac{\|x_*-x_0\|_2^2}{k}. \quad (1.73)$$

Proposition 1.3.7 ([52], Theorem 4.4) *Let $(x_k)_{k\geq 0}$ be the sequence of iterate produced by FISTA with starting point x_0, and x_* a minimizer of F_λ. Then*

$$\forall k \geq 1, \quad F_\lambda(x_k) - F_\lambda(x_*) \leq \left\|A^TA\right\|_{\boldsymbol{Fro}} \frac{4\|x_*-x_0\|_2^2}{(k+1)^2}. \quad (1.74)$$

[9] Authors provide a restricted eigenvalue condition stated with $\|x\|_2$ instead of $\|x_I\|_2$ in the denominator of (1.70), leading to constant smaller by at most a factor $\sqrt{1+\delta^{-1}}$.

[10] The two differences are higher numerical constants and the fact that design matrix A must satisfy the condition RE(s,κ,3).

In practice, the first order algorithm used to solve the Lasso is Coordinate Gradient Descent (CGD) [53] with associated computational cost $\mathcal{O}(mn)$ for one iteration, and the following convergence result.

Proposition 1.3.8 ([54], Corollary 3.8) *Let $(x_k)_{k\geq 0}$ be the sequence of iterates produced by CGD with starting point x_0, and x_* a minimizer of F_λ. Then*

$$\forall k \geq 1, \quad F_\lambda(x_k) - F_\lambda(x_*) \lesssim \max_i \frac{n \left\| Row_i(A) \right\|_2^2}{m} \frac{\|x_* - x_0\|_2^2}{k + 1/n}. \tag{1.75}$$

1.3.3 Stochastic approximation, multistage procedures and sparse recovery in very high dimensions

We now turn to problem (1.3). We are given random samples $(\phi_i, \eta_i)_{i=1}^m \in \mathbf{R}^n \times \mathbf{R}$, linked by the following equation

$$\eta_i = \mathfrak{r}(\phi^T x_*) + \sigma \xi_i,$$

where $(\phi_i, \xi_i)_{i=1}^m$ are independent copies of r.v. $(\phi, \xi) \in \mathbf{R}^n \times \mathbf{R}$. Under this random design, consider the following *Stochastic Optimization* problem:

$$\min_{x \in \mathcal{X}} \left\{ g(x) := \mathbf{E}\left[G(x, (\phi, \eta)) \right] \right\}, \quad G(x, (\phi, \eta)) := \mathfrak{s}(\phi^T x) - \phi^T x \eta, \tag{1.76}$$

with $\mathfrak{s}'(t) = \mathfrak{r}(t)$. When $\mathbf{E}[\phi\xi] = 0$, x_* is a critical point of g, since

$$\nabla g(x) = \mathbf{E}\left[\phi [\mathfrak{r}(\phi^T x) - \mathfrak{r}(\phi^T x_*)] \right]. \tag{1.77}$$

Thus, when g has a unique minimizer, retrieving x_* and minimizing g are equivalent problems. To do the latter, one can minimize the *sample average approximation* (SAA) of g

$$\widehat{g}_N(x) := \frac{1}{N} \sum_{i=1}^N G(x, (\phi_i, \eta_i)), \tag{1.78}$$

as it is an unbiased estimate of $g(x)$, i.e.,

$$\mathbf{E}[\widehat{g}_N(x)] = g(x) \quad \forall x \in \mathcal{X}. \tag{1.79}$$

One can also compute unbiased *Stochastic Approximation* (SA) [55]

$$\nabla G(x, (\phi_i, \eta_i)) := \phi_i \left[\mathfrak{r}(\phi_i^T x) - \eta_i \right] \tag{1.80}$$

of $\nabla g(x)$, and use *Stochastic First Order Methods* (SFOM) to minimize g.
In the context of linear regression, i.e. $\mathfrak{r}(t) = t$, observe that

$$g(x) = \mathbf{E}\left[\frac{1}{2} (\phi^T [x_* - x] + \sigma\xi)^2 \right] - \mathbf{E}\left[\tfrac{1}{2} \eta^2 \right],$$

thus minimizing g using SAA is the same as solving the least square problem

$$\min_{x \in \mathcal{X}} \left\{ L_N(x) := \frac{1}{2N} \sum_{i=1}^N (\eta_i - \phi_i^T x)^2 \right\}.$$

When assuming $\|x_*\|_0 = s$ with $s << n$, one can thus treat this task by solving the Lasso or Dantzig selector. Considering the Lasso problem and assuming $\xi \sim \mathcal{N}(0,1)$, one can use CGD to compute an estimate $\widehat{x}_N^\lambda$ such that

$$L_N(\widehat{x}_N^\lambda) + \lambda \left\|\widehat{x}_N^\lambda\right\|_1 \leq \min_{x\in\mathcal{X}} \{L_N(x) + \lambda \|x\|_1\} + \tfrac{1}{2}\sigma^2 \tag{1.81}$$

in

$$\mathcal{O}\left(n^2\left(\left\|x^\lambda - x_0\right\|_2^2 \max_i \|\phi_i\|_2^2 \sigma^{-2} \bigvee 1\right)\right) \tag{1.82}$$

operations, with x^λ a minimizer of the Lasso problem.
On the other hand, SFOM using SA $\nabla G(x,(\phi_i,\eta_i)) := \phi_i(\phi_i^T x - \eta_i)$ of g's gradient at x usually have sub-iterations scaling linearly with dimension n. Thus, one can compute an estimate in $\mathcal{O}(nN)$ operations with this approach, which is computationally cheaper.

Stochastic Mirror Descent

A popular first order algorithm to solve (1.76) is the celebrated *Stochastic Mirror Descent* (SMD) [56]. The latter is the repeatition of the following iteration

$$x_{i+1} = \mathrm{Prox}_{\mathcal{X}}^D\left(x_i, \gamma_{i+1} G(x_i, w_{i+1})\right), \quad \mathrm{Prox}_{\mathcal{X}}^D(v,y) := \operatorname*{argmin}_{z\in X} \left\{y^T z + D(v,z)\right\}, \tag{1.83}$$

with metric D the *proximal function* and $\mathrm{Prox}_{\mathcal{X}}^D$ its associated *proximal operator*.
When measuring error of estimation using a norm $\|\cdot\|$, one usually requires proximal function D to be such that

$$\forall z, v \in \mathcal{X}, \quad D(z,v) \geq \frac{\alpha}{2}\|z - v\|^2.$$

A popular choice of metric fitting this requirement is the *Bregman Divergence* [57] generated by a function α-strongly convex w.r.t. norm $\|\cdot\|$.

Definition 9 *A convex continuously differentiable function $d : \mathcal{X} \to \mathbf{R}$ is said to be α-strongly convex w.r.t. norm $\|\cdot\|$ if*

$$\forall z, v \in \mathcal{X}, \quad V_d(v,z) := d(v) - d(z) - \nabla d(z)^T[v - z] \geq \frac{\alpha}{2}\|v - z\|^2. \tag{1.84}$$

Definition 10 *Let distance-generating function (d.-g.f.) $d : \mathcal{X} \to \mathbf{R}$ be a continuously differentiable convex function that is strongly convex w.r.t a norm $\|\cdot\|$. The Bregman divergence $V_d : \mathcal{X} \times \mathcal{X} \to \mathbf{R}_+$ associated with d is defined by*

$$\forall (v,z) \in \mathcal{X},\ V_d(v,z) = d(z) - d(v) - \nabla d(v)^T[z - v]. \tag{1.85}$$

Among the most widely used Bregman divergences for SMD, are the following ones:

- the squared euclidian distance:

$$\forall (v,z) \in \mathbf{R}^n,\ V_d(v,z) = \|v - z\|_2^2, \quad d(\cdot) = \|\cdot\|_2^2.$$

- the Kullback-Liebler divergence, defined over the simplex $\Delta_1 := \left\{z \in \mathbf{R}^n_+ : \sum_{i=1}^n z_i = 1\right\}$ and generated by the negative entropy $h(v) = \sum_i^n v_i \ln(v_i)$:

$$\forall (v,z) \in \Delta_1,\ V_h(v,z) := \sum_{i=1}^n v_i \ln(v_i/z_i).$$

- the divergences defined when using $d(\cdot) := \|\cdot\|_p^2$ or $d(\cdot) = \|\cdot\|_p^p$, with p close to one, e.g., $p = 1 + \frac{1}{\ln(n)}$.

Upper bounds on the expected inaccuracy of an estimate produced by a properly tuned SMD are available for general functions g

$$\forall x \in \mathcal{X},\ \ g(x) := \mathbf{E}_w\left[G(x,w)\right] \tag{1.86}$$

being *Lipschitz* or *Smooth*, with $G(\cdot, w)$ being convex for every realization w.

Definition 11 *A continuously differentiable function $d : \mathcal{X} \to \mathbf{R}$ is L-Lipschitz w.r.t. norm $\|\cdot\|$ if*

$$\forall z \in \mathcal{X}, \quad \|\nabla d(x)\|_* \leq L. \tag{1.87}$$

Definition 12 *A continuously differentiable function $d : \mathcal{X} \to \mathbf{R}$ is β-smooth if its gradient ∇d is β-Lipschitz w.r.t. norm $\|\cdot\|$:*

$$\forall z, v \in \mathcal{X}, \quad \|\nabla d(z) - \nabla d(v)\|_* \leq \beta \|z - v\|. \tag{1.88}$$

The two result we are about to present assume that available stochastic subgradient $\nabla G(x,w)$ is such that its expectation is in the sub-differential of g at x, i.e.

$$\forall x \in \mathcal{X}, \quad \nabla g(x) := \mathbf{E}_w\left[\nabla G(x,w)\right] \in \partial g(x), \tag{1.89}$$

and that it has bounded second moment:

$$\forall x \in \mathcal{X}, \quad \mathbf{E}\left[\|\nabla G(x,w) - \nabla g(x)\|_*^2\right] \leq \varkappa,\ \ \varkappa > 0. \tag{1.90}$$

They both use in (1.83) the Bregman divergence associated with d.-g.f. d that is 1-strongly convex w.r.t. $\|\cdot\|$ and associated with constant

$$\Theta := \max_{x,y \in \mathcal{X}} V_d(x,y). \tag{1.91}$$

Proposition 1.3.9 ([58], Theorem 4.1) *Assume that (1.90) holds and g is L-Lipschitz. Let sequence of points $(x_1, ..., x_N)$ generated by iteration (1.83) with proximal function $D = V_d$, and define*

$$\widehat{x}_t = \sum_{\tau=1}^t \frac{x_\tau}{t}. \tag{1.92}$$

For $t \geq 1$ and constant step-size strategy $\gamma_i = \sqrt{\frac{\Theta}{k(L^2+\varkappa)}}$, one has that

$$\mathbf{E}\left[g(\widehat{x}_N) - g(x_*)\right] \leq 2\sqrt{\frac{\Theta(L^2+\varkappa)}{N}}. \tag{1.93}$$

Proposition 1.3.10 ([58], Theorem 4.2) *Assume that (1.90) holds and g is β-smooth. Let sequence of points $(x_1, ..., x_{N+1})$ generated by iteration (1.83) with proximal function $D = V_d$, and define*

$$\widehat{x}_{t+1}^{av} = \sum_{\tau=2}^{t+1} \frac{x_\tau}{t}. \tag{1.94}$$

For $t \geq 1$ and constant step-size strategy $\gamma_i = \min\left\{\frac{1}{2\beta}, \sqrt{\frac{\Theta}{k\varkappa}}\right\}$, one has that

$$\mathbf{E}\left[g(\widehat{x}_{N+1}^{av}) - g(x_*)\right] \leq \frac{2\beta\Theta}{N} + 2\sqrt{\frac{\Theta\varkappa}{N}}. \tag{1.95}$$

Stochastic Dual Averaging

Another FOM is the Dual Averaging (DA) algorithm [59, 60]. In [60], authors propose the following setting. Let d be $\mu(d)$-strongly convex w.r.t. $\|\cdot\|$, and such that

$$\forall x \in \mathcal{X}, \;\; d(x) \leq C(d)\|x\|^2 \text{ and } d(x) \geq d(0) = 0. \tag{1.96}$$

For a given *proximal center* $\overline{x} \in \mathcal{X}$ such that $\|x_* - \overline{x}\| \leq R$, define the *proximal operator*

$$\mathrm{Prox}_{z,R,\beta}(v) := \arg\max_y \left\{v^T[y - z] - \beta d_{z,R}(y) : \;\; y \in \mathcal{X}, \;\; \|y - z\| \leq R\right\}, \; \beta > 0, \tag{1.97}$$

with d.-g.f. $d_{z,R}(y)$ defined by

$$\forall y \in \mathcal{X}, \quad d_{z,R}(y) := d\left(\frac{y - z}{R}\right). \tag{1.98}$$

For seqences of positive reals $\lambda_1, ..., \lambda_i$ and $\beta_1 \leq ... \leq \beta_{i+1}$, and filtration $(w_1, ..., w_i)$, the $i+1$-th iteration of DA produces estimate $\widehat{x}_{i+1}$ through

$$\widehat{x}_{i+1} = \mathrm{Prox}_{\overline{x},R,\beta_{i+1}}\left(-\sum_{t=1}^{i} \lambda_t \nabla G(\widehat{x}_t, w_t)\right). \tag{1.99}$$

Authors in [60] propose a result on both the convergence of the expected inaccuracy and the expected error of estimation, for g being *uniformly convex* and Lipschitz.

Definition 13 *For $p \geq 2$, a convex continuously differentiable function $d : \mathcal{X} \to \mathbf{R}$ is said to be (μ, p)-uniformly convex w.r.t. norm $\|\cdot\|$ if*

$$\forall z, v \in \mathcal{X}, \quad V_d(v, z) \geq \frac{\mu}{p}\|v - z\|^p. \tag{1.100}$$

Proposition 1.3.11 ([60], Corollary 5) *Assume that g is $(\mu(g), p)$-strongly convex and L-Lipschitz w.r.t. $\|\cdot\|$, and that stochastic subgradient satisfy (1.90). Let $\lambda_k = 1$ and $\beta_k = \gamma\sqrt{N+1}$, with*

$$\gamma = \sqrt{\frac{L^2 + \varkappa}{2C(d)\mu(d)}}. \tag{1.101}$$

Then the approximate solution $\widehat{x}_N(\overline{x}, R)$ produced by the DA algorithm after N iterations is such that

$$\mathbf{E}\left[g(\widehat{x}_N(\overline{x}, R))\right] - g(x_*) \leq R\sqrt{\frac{2C(d)(L^2+\varkappa)}{\mu(d)(N+1)}} \tag{1.102}$$

$$\mathbf{E}\left[\|\widehat{x}_N(\overline{x}, R) - w_*\|^p\right] \leq \frac{R}{\mu(g)}\sqrt{\frac{2C(d)(L^2+\varkappa)}{\mu(d)(N+1)}}. \tag{1.103}$$

Multistage procedures

Rates (1.93), (1.95),(1.102) and (1.103) can be improved using multistage procedures [60, 61] when assuming objective to be strongly or uniformly convex w.r.t. $\|\cdot\|$. The latter are algorithms restarting instances of FOM algorithms, e.g., SMD or DA, called *stage*. At a given stage, say, the i-th, we are given a starting point $\overline{x}_i \in \mathcal{X}$ such that

$$\|\overline{x}_i - x_*\| \leq R_i. \tag{1.104}$$

The stage then runs a FOM for m_i iterations using d.-g.f. $d_{\overline{x}_i, R_i}$ defined as in (1.98), with initial point

$$x_0^{(i)} \in \mathcal{X} \cap \{z \in \mathbf{R}^n,\ \|z - \overline{x}_i\| \leq R_i\}.$$

At the end of the stage, the procedure sets $\overline{x}_{i+1} := \widehat{x}_{m_i}^{(i)}$, with guarantee

$$\|\overline{x}_{i+1} - x_*\| \leq R_{i+1}$$

holding either in expectation or with high probability when subgradients are stochastic.

Mulstistage routines applied to Sparse recovery

Observe that for functions such as (1.76), one has [11]

$$\varkappa \geq \sigma^2 \mathbf{E}\left[\|\phi\|_*^2 \xi^2\right]. \tag{1.105}$$

If $\|\cdot\| = \|\cdot\|_2$, i.e., the setting is Euclidean, one can consider $d(\cdot) = \|\cdot\|_2^2$. In this case, the corresponding SMD algorithm is the *Projected Stochastic Gradient Descent*. The latter may suffer from poor convergence of the expected inaccuracy for very high dimensions. For example, in linear regression, where $\phi \sim \mathcal{N}(0, I)$ and $\xi \sim \mathcal{N}(0, 1)$ are independent, $\varkappa \geq \sigma^2 n$. Thus, unless a specific assumption is made on regressor ϕ, such as sparsity (see [62]), one cannot avoid a term larger than the dimension n in (1.95). Thus, for very large n, SFOM must be considered in a non-Euclidean setting.

For example, considering $\|\cdot\| = \|\cdot\|_1$, (1.105) gives $\varkappa \gtrsim \sigma^2 \ln(n)$ for $\phi \sim \mathcal{N}(0, I)$ and $\xi \sim \mathcal{N}(0, I)$. For such setting, one can use multistage procedures running instances of the SMD algorithm in each stage. Assuming that the objective is $\mu(g)$-strongly convex and $\beta(g)$-smooth w.r.t. $\|\cdot\|_1$, (1.95) yields

$$\mathbf{E}[\|\widehat{x}_N - x_*\|_1^2] \lesssim \frac{\beta(g)\Theta}{\mu(g)N} + \sqrt{\frac{\Theta\varkappa}{\mu(g)^2 N}}.$$

[11] Simply observe the decomposition

$$\mathbf{E}\left[\left\|\phi(\phi^T x - \eta) - \mathbf{E}\left[\phi\phi^T[x - x_*]\right]\right\|_*^2\right] = \mathbf{E}\left[\left\|\phi\phi^T[x - x_*] - \mathbf{E}\left[\phi\phi^T[x - x_*]\right] - \sigma\phi\xi\right\|_*^2\right]$$

taken at $x = x_*$.

In cases such as linear regression, this approach suffers from the fact that the ratio $\frac{\beta}{\mu}$ is lower bounded by dimension n (see [60], p.2 for proof of this claim). Assuming that $\Theta \leq C(d) \|x_0 - x_*\|_1^2$, N should be larger than $nC(d)$ to reduce initial error. Thus a multistage procedure cannot converge in the setting of sparse recovery, i.e., $N \leq n$, when assuming strong convexity w.r.t. $\|\cdot\|_1$. This motivates the need for more restrictive assumptions.
For example, the authors in [63] study a multistage procedure called *Regularization Annealed epoch Dual AveRaging* (RADAR) under the *Local Strong Convexity* (LSC) condition.

Definition 14 *Function $g : \mathcal{X} \to \mathbf{R}$ is $\gamma(R)$-locally strongly convex (LSC) w.r.t. $\|\cdot\|_2^2$ if for all z, v such that $\|z\|_1 \leq R$ and $\|v\|_1 \leq R$,*

$$V_g(v, z) \geq \frac{\gamma}{2} \|z - v\|_2^2 . \tag{1.106}$$

At the i-th stage of RADAR, an instance of DA with $d = \frac{1}{2(p-1)} \|\cdot\|_p^2$ and $p = \frac{2\ln(n)}{2\ln(n)-1}$ is ran for m_i iteration to solve intermediary problem

$$\min_{z \in \mathcal{X}} \left\{ g(z) + \lambda_i \|z\|_1 : \ \|z - \overline{x}_i\|_p \leq R_i \right\}. \tag{1.107}$$

The main difference with (1.99) is the use of stochastic subgradient $\nabla G(x, w) + \lambda_i \nu(x)$ instead of $\nabla G(x, w)$, with $\nu(x) \in \partial \|x\|_1$.

Proposition 1.3.12 ([63], Theorem 1) *Assume that g is $\gamma(R)$-LSC and that it satisfies the following assumptions:*

- *g is $G(R)$-Lipschitz for all z, v such that $\|z - x_*\|_1 \leq R$ and $\|v - x_*\|_1 \leq R$:*

$$|g(z) - g(v)| \leq G(R) \|z - v\|_1 . \tag{1.108}$$

- *For all z such that $\|z\|_1 \leq R$, gradients are centered and sub-Gaussian, i.e. for some $\sigma^2(R)$,*

$$\mathbf{E}_w \left[\exp\left(\|\nabla G(z, w) - \nabla g(z)\|_\infty^2 / \sigma^2(R) \right) \right] \leq \exp(1). \tag{1.109}$$

RADAR produces after N oracle calls an estimate $\widehat{x}_N$ such that with probability greater than $1 - 6\exp(-t/12)$,

$$\|\widehat{x}_N - x_*\|_2^2 \lesssim (\sigma^2(R) + G^2(R)) \frac{s(\ln(n) + t)}{\gamma^2 N}. \tag{1.110}$$

Authors illustrate result (1.110) in the case of sparse least square regression. For the latter, assuming that regressors are uniformly bounded, i.e.

$$\|\phi\|_\infty \leq B,$$

and that noise $\xi \sim \mathcal{N}(0, \sigma^2)$, they show that

$$\sigma^2(R) \lesssim B^4 R^2 + B^2 \sigma^2, \tag{1.111}$$

$G^2(R) \leq \|\Sigma\|_{\mathbf{Fro}} R^2$, and $\gamma = \rho_{min}(\Sigma)$, for $\Sigma = \mathbf{E}\left[\phi\phi^T\right]$. Thus, in that setting, (1.110) reads

$$\|\widehat{x}_N - x_*\|_2^2 \lesssim \left((B^4 + \|\Sigma\|_{\mathbf{Fro}}) R^2 + B^2 \sigma^2 \right) \frac{s(\ln(n) + t)}{\rho_{min}(\Sigma)^2 N}. \tag{1.112}$$

This result is in contrast with others in the sparse recovery literature has it makes assumption on Σ instead of the empirical covariance matrix

$$\Sigma_N := \frac{1}{N}\sum_{i=1}^{N} \phi_i \phi_i^T. \tag{1.113}$$

Yet, compared to results such as (1.72) or (1.71), the obtained rate contains an additional term proportional to R^2 and decreasing sub-linearly with the sample size N. Additionally, the length of stages in the RADAR is such that $m_i \asymp s^2$, which limits the range of sparsity one can consider with $N \leq n$.
Recently, [64] proposed another multistage procedure for minimizing the L-smooth objective g (w.r.t. $\|\cdot\|_1$). The underlying FOM is the SMD algorithm with d.-g.f.

$$d_{\overline{x}}(x) = \frac{e}{2}\ln(n) n^{q(p)} \|v - x_0\|_p^2, \quad p = 1 + \ln(n)^{-1},\ q(p) = \frac{(p-1)(2-p)}{p},$$

where $\overline{x}$ is the current proximal center of that stage. Their analysis relies upon a noise decomposition more suited to smooth functions such as (1.76). Namely, they assume that for all $x \in \mathcal{X}$, stochastic gradient $\nabla G(x,w)$ is such that $\mathbf{E}_w[\nabla G(x,w)] = \nabla g(x)$ and that for some $\varkappa, \varkappa', \nu$ such that $1 \leq \varkappa, \varkappa' < \infty$ and $L \leq \nu$,

$$\mathbf{E}_w\left[\|\nabla G(x,w) - \nabla g(x)\|_\infty^2\right] \leq \varkappa\nu V_g(x, x_*) + \varkappa'\sigma_*^2, \tag{1.114}$$

where

$$\sigma_*^2 := \mathbf{E}_w\left[\|\nabla G(x_*,w) - \nabla g(x_*)\|_\infty^2\right].$$

They also require their objective to satisfy these additional assumptions.

- for all $z \in \mathcal{X}$, g the satisfies *quadratic growth condition* w.r.t. Euclidian norm, i.e.

$$g(z) - g_* \geq \tfrac{1}{2}\kappa \|z - x_*\|_2^2. \tag{1.115}$$

- Given a vector $z \in \mathcal{X}$, one can efficiently solve the following problem

$$\mathrm{sparse}_s(z) := \operatorname*{Argmin}_{v \in \mathcal{X}} \{\|z - v\|_2 : \ \|v\|_0 = s\}. \tag{1.116}$$

Proposition 1.3.13 ([64], Proposition 2.1) *Let $(x_1, ..., x_m)$ the sequence of iterates generated by the aforementioned SMD algorithm with constant step-size $2\varkappa\nu$ and initiall starting point x_0, with $\|x_* - x_0\|_1 \leq R$. The average of iterates $\widehat{x}_m(x_0) := \frac{1}{m}\sum_{i=1}^m x_i$ is such that*

$$\mathbf{E}[g(\widehat{x}_m)] - g_* \leq \frac{4e^2\nu\varkappa R^2 \ln(n)}{m} + \frac{\varkappa'\sigma_*^2}{\varkappa\nu}. \tag{1.117}$$

A direct consequence of (1.117) and (1.115) is that defining $y_m = \mathrm{sparse}_s(\widehat{x}_m)$, one has that

$$\mathbf{E}[\|\widehat{y}_m - x_*\|_1] \leq \frac{4s}{\kappa}\left(\frac{4e^2\nu\varkappa s \ln(n) R^2}{m} + \frac{\varkappa' s\sigma_*^2}{\varkappa\nu}\right) \tag{1.118}$$

The multistage procedure based on (1.118), called **SMD-SR**, is a two-phase algorithm. The notion of phases is related to the number m of iterations per stage. In the first phase, $m = m_0$ is chosen

so that the first term of the RHS in (1.118) is less than $\frac{R^2}{2}$. These stages with constant number of iterations are repeated until the first term is of the order of the second. When this occurs, the algorithm enters the second "asymptotic" phase, where the number of iterations per stage is multiplied by two after the completion of a stage (i.e. $m_k \asymp m_0 \times 2^k$ at the k-th asymptotic stage.). This procedure produces an estimate with the following guarantiees:

Theorem 1.3.1 ([64], Theorem 2.1) *Let $\widehat{x}_N$ be the output of the last stage one can possibly do using a total number of iteration N of the previously outlined procedure, and $\widehat{y}_N$ its sparsification. The latter enjoy the following guarantees*

$$\mathbf{E}[g(\widehat{x}_N)] - g_* \leq \frac{\kappa R^2}{s} \exp\left\{-\frac{cN\kappa}{\ln(n)\varkappa s\nu}\right\} + C\frac{\varkappa'\sigma_*^2 s\ln(n)}{\kappa N}, \tag{1.119}$$

$$\mathbf{E}\left[\|\widehat{y}_N - x_*\|_1\right] \leq C_1 R \exp\left\{-\frac{cN\kappa}{\ln(n)\varkappa s\nu}\right\} + C_2\frac{\sigma_* s\sqrt{\varkappa'\ln(n)}}{\kappa\sqrt{N}}. \tag{1.120}$$

Authors illustrate their result in the case of (1.76) where τ is the identity, ϕ is such that

$$\|\phi\|_\infty \leq B, \;\; \Sigma \succeq \kappa_\Sigma I, \;\; \|\Sigma\|_\infty \leq \upsilon,$$

with $B, \kappa_\Sigma > 0$ and υ known, and ξ such that

$$\mathbf{E}[\xi] = 0 \text{ and } \mathbf{E}[\xi^2] \leq 1.$$

In this special case, their assumptions are satisfied for $\varkappa'\sigma_*^2 \asymp B^2\sigma^2$, $\kappa = \kappa_\Sigma$ and $\varkappa\nu \asymp (B + \sqrt{\upsilon})^2$, yielding error of estimation (1.119)

$$\mathbf{E}\left[\|\widehat{y}_N - x_*\|_1\right] \lesssim R\exp\left\{-\frac{cN\kappa_\Sigma}{s\ln(n)(B+\sqrt{\upsilon})}\right\} + \frac{B\sigma s}{\kappa_\Sigma}\sqrt{\frac{\ln(n)}{N}}. \tag{1.121}$$

In contrast to (1.112), (1.121) has a dependency on prior R decreasing exponentially fast, with additional term proportional to $\sigma s\sqrt{\ln(n)/N}$, in the same faschion as (1.72). Same as in [63], assumptions are made on the distributions of regressors and Σ instead of the empirical covariance matrix.

1.3.4 Contribution on SA applied to sparse recovery

Chapter 6 is based on paper *Stochastic Mirror Descent for Large-Scale Sparse Recovery* presented at AIStat 2023. We propose a novel two-phase multistage algorithm to minimize a smooth objective g with sub-Gaussian stochastic gradients. This multistage procedure is based on the *Composite Stochastic Mirror Descent* (CSMD), which uses a composite proximal operator instead of (1.83), defined as

$$\mathrm{Prox}^d_{h,x_0}(\zeta, x) := \operatorname*{argmin}_{z\in X_R(x_0)} \left\{\zeta^T x + h(z) + V_{x_0}(x,z)\right\}, \; X_R(x_0) := \{z \in \mathcal{X} : \|z - x_0\| \leq R\} \tag{1.122}$$

The penalty $h(\cdot)$ is chosen as $\kappa_k\|\cdot\|_1$ in the k-th stage, in order to minimize the following sub-problem

$$\min_{x\in X_{R_k}(\overline{x}_k)} \left\{F_{\kappa_k}(x) := g(x) + \kappa_k\|x\|_1\right\}, \tag{1.123}$$

where $\overline{x}_k$ is the current proximal center and R_k is an upper bound on $\|\overline{x}_k - x_*\|_1$ that holds with large probability.
Our analysis of the ϵ-error of the output produced by the CSMD-*Sparse Recovery* (CSMD-SR) is based on two main component. First, we propose a novel analysis of the CSMD algorithm using a noise decomposition such as (1.114), along with large deviations inequalities for super-martingales. Second, we link a high-probability upper bound on the inaccuracy to the error of estimation when g satisfies the *Reduced Strong Convexity* (RSC) assumption:

$$\forall x \in X, \quad \|x - x_*\|_1 \leq \delta \left[\rho s \kappa + \frac{(F_\kappa(x) - F_\kappa(x_*))_+}{\kappa} \right]. \tag{1.124}$$

In the context of sparse generalized linear regression, the latter holds when the objective g has quadratic growth as in (1.115), and the covariance matrix Σ satisfies the $Q_{\lambda,\psi}$ condition:

$$\forall z \in \mathbf{R}^n, \ \ \forall I \subset \{1, ..., n\}, \ |I| \leq s, \quad \|z_I\|_1 \leq \sqrt{\frac{s}{\lambda}} \|z\|_\Sigma + \frac{1}{2}(1 - \psi) \|z\|_1 . \tag{1.125}$$

We consider two variants of the CSMD-SR algorithm. Both have the same preliminary phase with constant stage-length m. The first version uses a stage-length proportional to $\mathcal{O}(4^k m)$ for the k-th asymptotic stage. The second, called the "mini-batch" version, has constant stage-length m in the asymptotic phase. It computes at each step i of the k-th asymptotic stage the mini-batch

$$H_k(\widehat{x}_i^{(k)}) = \frac{1}{L_k} \sum_{l=1}^{L_k} G(\widehat{x}_i^{(k)}, w_{k,l}), \quad L_k = \mathcal{O}(\ln(n) 4^k), \tag{1.126}$$

where $\widehat{x}_i^{(k)}$ is the current output at step i. $\widehat{x}_{i+1}^{(k)}$ is then computed using $H_k(\widehat{x}_i^{(k)})$ instead of $\nabla G(\widehat{x}_i^{(k)}, w_{k,i})$ in (1.83).
For both algorithms, we provide a high probability upper bound on the error of estimation in $\|\cdot\|_1$ of the output produced by the last possible stage when using N samples. Similar to result in [64], these upper bounds exhibit an exponentially fast convergence to the noise regime, which is of the order of

$$\mathcal{O}\left(\delta^2 \rho \sigma_* s \sqrt{\frac{\ln(n)}{N}} \right)$$

for the first version of CSMD-SR, and

$$\mathcal{O}\left(\delta^2 \rho \sigma_* s \frac{\ln(n)}{\sqrt{N}} \right)$$

for the mini-batch version. Finally, we support our analysis with experiments.

Chapter 6 provides two extensions to these results. The first one proposes an estimate that adapts to parameter ρs in (1.124), as it does not require knowledge of ρs to be computed. It uses Lepski's procedure [5] to select one estimate in a finite collection. Each estimate in this collection is the output of an instance of the CSMD-SR algorithm's mini-batch version with different stage-lengths m_i tuned for parameter $\rho_i s_i$. We show that this procedure selects with high probability an estimate such that its error is within a constant multiplicative factor of the best estimate in

said collection. If the stage-lengths of each of these algorithm is properly chosen, this remains true when comparing our selected estimate's error to

$$\left\|\widehat{x}(\rho s) - x_*\right\|_1,$$

where $\widehat{x}(\rho s)$ is the output of a CSMD-SR algorithm that knows the smallest parameter ρs such that the objective g satisfies the RSC(ρs). In the second part of the chapter, we describe choices of hyperparameters of the CSMD-SR algorithm adapted to assumption that the objective satisfies the *reduced uniform convexity* (RUC) instead of the RSC. The former is a "uniform convexity" counterpart of the latter. For some critical radius R_{RUC} and $q \in [1, 2)$, we assume that it holds that

$$\forall x \in \mathcal{X}, \ \ \|x - x_*\|_1 \leq R_{\mathrm{RUC}} \implies \|x - x_*\|_1 \leq \frac{\delta}{\kappa}\left(\rho s^{\frac{q}{2}}\kappa^q + (F_\kappa(x) - F_\kappa(x_*))_+\right). \tag{1.127}$$

Under this assumption, we obtain an upper bound on the estimation error of the CSMD-SR algorithm for its mini-batch variant with asymptotic term scaling as

$$\mathcal{O}\left(\rho\left(\frac{(\delta\sqrt{s})^p\sigma_*\ln(n)}{\sqrt{N}}\right)^{\frac{1}{(p-1)}}\right).$$

Chapter 2

Robust Counterparts of Linear Inverse Problems

Abstract

We consider an uncertain linear inverse problem as follows. Given observation $\omega = Az + \xi$ where $A \in \mathbf{R}^{m\times n}$ and $\xi \in \mathbf{R}^m$ is observation noise, we want to recover a linear transformation of unknown signal z, known to belong to a convex set $\mathcal{Z} \subset \mathbf{R}^n$. As opposed to "usual" settings of such problem, we allow feasible set $\mathcal{Z}$ to be uncertain. For instance, the latter can be the direct sum $\mathcal{X} + \delta\mathcal{U}$, where $\mathcal{X}$ and $\mathcal{U}$ are known and δ is unknown. In a series of problem settings, under various assumptions on the nature of problem uncertainty, we discuss the properties of two types of estimates—linear estimates and polyhedral estimates (A particular class of non-linear estimates introduced in [32, 33]). We show that in the situation where the signal set is an uncertain ellitope (essentially, a symmetric convex set delimited by quadratic surfaces), nearly minimax optimal (up to a moderate suboptimality factor) estimates can be constructed by means of efficient convex optimization routine.

2.1 Building robust estimates

Given an observation

$$\omega = Az + \xi \in \mathbf{R}^m \tag{2.1}$$

where $A \in \mathbf{R}^{m\times n}$ and random noise $\xi \sim P_z$, our objective is to build an estimate $\widehat{w}(\cdot)$ of the linear image $w = Bz \in \mathbf{R}^\nu$ of unknown signal $z \in \mathcal{Z} \subset \mathbf{R}^n$. In what follows, we assume that the collection

$$\mathfrak{m} = [A, B, \mathcal{Z}, (P_z, z \in \mathcal{Z})] \tag{2.2}$$

of the problem data is uncertain and belongs to a known family $\mathfrak{M} = \{\mathfrak{m}(\delta),\ \delta \in \mathbf{R}_+\}$, where δ is an unknown parameter; we refer to $\mathfrak{m}(\delta)$ as *uncertain observation model*.

More precisely, A, B, and $P_z = P$ are given (certain), while $z = [x; u]$ [1], with $x \in \mathcal{X} \subset \mathbf{R}^{n_x}$ and $u \in \mathcal{U}(\delta) \subset \mathbf{R}^{n_u}$, so that $z \in \mathcal{Z}(\delta) = \mathcal{X} \times \mathcal{U}(\delta) \subset \mathbf{R}^n$. Signal set $\mathcal{X}$ is a known symmetric w.r.t. the origin, convex and compact set, while the "uncertainty set" $\mathcal{U}(\delta)$ belongs to a known family of sets with same properties as $\mathcal{X}$, parameterised with unknown $\delta \geq 0$. As a consequence, the uncertainty models we consider are of the form

$$\mathfrak{m}(\delta) = [A, B, \mathcal{X} \times \mathcal{U}(\delta), P]. \tag{2.3}$$

Building on the foundations laid in [30, 31, 33], we do not make any assumptions on matrices A and B. As a consequence, our setting accommodates with a wide range of linear transformations commonly used in nonparametric statistics, e.g., FFT transforms, k-th order derivation, and wavelet basis expansions. Moreover, the formulations of observation (2.1) and target $w = Bz$ are flexible enough to adapt to various uncertain statistical situation, such as

A. *robust estimation*. More precisely, consider the situation where we are given observation

$$\omega = \overline{A}x + u + \xi$$

and aim to estimate $w = \overline{B}x$. In our setting, this situation is captured by the special case where $A = [\overline{A} : I]$, $B = [\overline{B}, 0]$. The component u of our signal is then considered as a deterministic but unknown nuisance. Taking $\delta = \|u\|_2$, with δ also unknown, our total signal $z = [x; u]$ lies in $\mathcal{X} \times \mathcal{B}_2(\delta)$, with $\mathcal{B}_2(\delta)$ being the ball of radius δ of the Euclidean norm.

B. *signal set uncertainty*. We consider the situation where we are given a data-set made of N signals $(z_i)_{i=1}^N \in \mathbf{R}^n$ assumed to belong to an unknown subset $\mathcal{X}$. We can build, using K semidefinite matrices $(R_k)_{k=1}^K$ an *approximate* set $\mathcal{X}_N$ from the z_i's as follows:

$$\mathcal{X}_N = \bigcap_{k\leq K} \left\{z \in \mathbf{R}^n, z^T R_k z \leq \varrho_k\right\}, \quad \varrho_k := \max_{i\leq N} z_i^T R_k z_i.$$

When given observation (2.1) associated with unknown signal z, we can use the decomposition

$$z = \underbrace{z_{\mathcal{X}}}_{x(z)} + \underbrace{z - z_{\mathcal{X}}}_{u(z)}, \quad z_{\mathcal{X}} := \operatorname*{argmin}_{v\in\mathcal{X}} \|v - z\|_2^2.$$

Thus, in our formulation, the matrices of interest are A, B and $z \in \mathcal{X} + \mathcal{B}_2(\delta)$, with $\delta = \|u(z)\|_2$ unknown.

[1] Here and in what follows we adopt "MATLAB notation" for "horizontal" concatenation $[a, b]$ and "vertical" concatenation $[c; d]$ of matrices of appropriate sizes.

Given $\epsilon \in (0,1)$ and a norm $\|\cdot\|$ on $\mathbf{R}^\nu$, we quantify the quality of recovery of an estimation routine $\widehat{w}(\cdot)$ by its ϵ-risk[2]

$$\mathrm{Risk}_\epsilon[\widehat{w}|\mathfrak{m}(\delta)] := \sup_{z\in\mathcal{Z}(\delta),\ \mathcal{Z}(\delta)\in\mathfrak{m}(\delta)} \inf\left\{\rho : \mathrm{Prob}_\xi\{\|Bz - \widehat{w}(\omega)\| > \rho\} \le \epsilon\right\}. \tag{2.4}$$

We consider the problem described above for $\mathfrak{M}(\Delta) = \{\mathfrak{m}(\delta),\ \delta\in\Delta\}$, with Δ an interval, and assume that when given observation (2.1), signal $z \in \mathfrak{m}(\delta)$ with δ unknown to us. We are thus interested in estimates that are *adaptive* to δ, i.e., that do not require knowledge of its value to be computed, and are nearly optimal w.r.t. minimax risk, defined as

$$\mathrm{RiskOpt}_\epsilon[\mathfrak{m}(\delta)] := \inf_{\widehat{w}} \mathrm{Risk}_\epsilon[\widehat{w}|\mathfrak{m}(\delta)]. \tag{2.5}$$

Adaptive estimates have been widely studied in the literature, for instance in [5, 7, 10, 11, 65–68]. In our case, we consider the following approach. We approximate $\mathfrak{M}(\Delta)$ by

$$\mathfrak{M}(I) = \{\mathfrak{m}(\delta_i),\ i\in\{0,1\cdots,I\}\},$$

with $\delta_i \in \Delta$ for all $i \in [0, I]$. Given the collection of estimates

$$\mathfrak{W}(I) = \left\{\widehat{w}^{(i)},\ i\in\{0,1\cdots,I\}\right\},$$

we propose data-driven procedures that selects an estimate in $\mathfrak{W}(I)$ with ϵ-risk close to the smallest possible one in that collection. We then proceed to describe computational procedures to design $\widehat{x}^{(i)}$ such that $\mathrm{Risk}_\epsilon[\widehat{x}^{(i)}|\mathfrak{m}(\delta_i)]$ is close to the lowest possible ϵ-risk for the considered type of estimate.

More precisely, we will consider two type of estimates, the

- *linear estimates*, that we define as

$$\widehat{w}^H_{\mathrm{lin}}(\omega) := H^T\omega,\ H\in\mathbf{R}^{m\times\nu}. \tag{2.6}$$

  In part, due to their simplicity, they constitute a popular tool for solving estimation problems as described above. They have received much attention in the statistical literature (cf. [4, 23, 24, 28, 69–71] among many others). When designing a linear estimate, the emphasis is on how to specify the matrix H in order to obtain the lowest possible maximal over $\mathfrak{m}$ estimation risk, which is then compared to the minimax risk. "Near optimality" results for the case of indirect observations (where A and B are arbitrary) are the subject of recent papers [30, 31], where it was shown that in the *ellitopic case*, where the (known) set of signals $\mathcal{Z}$ and the unit ball $\mathcal{B}_*$ of the norm conjugate to $\|\cdot\|$ are *ellitopes*[3], a properly designed, via solving an explicit convex optimization problem, linear estimate is nearly optimal.

- *Polyhedral estimates.* The idea of a *polyhedral estimate* goes back to [72] where it was shown (see also [36, Chapter 2]) that such estimate is near-optimal when recovering smooth multivariate regression function known to belong to a given Sobolev ball from noisy observations taken along a regular grid. It has been recently reintroduced in [73] and [74] and extended to the setting to follow in [33]. In this setting, a polyhedral estimate $\omega\mapsto\widehat{w}^H_{\mathrm{poly}}(\omega)$ is specified by a *contrast matrix* $H\in\mathbf{R}^{m\times M}$ according to

$$\omega\mapsto\widehat{x}^H(\omega)\in\mathop{\mathrm{Argmin}}_{x\in\mathcal{X}}\|H^T(\omega-Ax)\|_\infty \mapsto \widehat{w}^H_{\mathrm{poly}}(\omega) := B\widehat{x}(\omega).$$

  In the ellitopic case, authors show how to compute efficiently a contrast such that the resulting estimate is nearly-minimax w.r.t. the ϵ-risk.

[2] Because the norm $\|\cdot\|$ is usually clear from the context, we omit the corresponding index of the risk.

[3] See [30] and [32, Section 4.2.1] below; as of now, an instructive example of ellitope is an intersection of finite family of ellipsoids/elliptic cylinders with common center.

2.2 Estimate aggregation routines

Throughout this section, $\Delta = [0, \overline{\delta}]$, with $\overline{\delta}$ known to us, and $\varepsilon \in (0,1)$. Given a parameter δ, model $\mathfrak{m}(\delta) \in \mathfrak{M}(\Delta)$ is of the form

$$\mathfrak{m}(\delta) = [A, B, \mathcal{Z}(\delta), (P_z, z \in \mathcal{Z}(\delta))],$$

with $\mathcal{Z}(\delta)$ computationally tractable[4], and such that the family $(\mathcal{Z}(\delta))_{0\le\delta}$ is monotone, i.e., for $\delta' \le \delta''$, one has $\mathcal{Z}(\delta') \subset \mathcal{Z}(\delta'')$. We are also given a *grid*, i.e. a collection of reals $(\delta_i)_{i=0}^I$, such that

$$0 = \delta_0 < \delta_1 < ... < \delta_I = \overline{\delta} \tag{2.7}$$

and we assume that given observation ω as in (2.1) stemming from the model $\mathfrak{m}(\delta_i)$, we can build a "presumably good" estimate of $w = Bz$. Our first objective is to describe two generic data-driven adaptive estimation procedures which allow to "aggregate" such estimates when the "true" value of $\delta \in \Delta$ is unknown, resulting in an estimate suited for $\mathfrak{m}(\delta)$.

2.2.1 Aggregating linear estimates

We first propose an adaptive routine for aggregating linear estimates $\widehat{w}_{\rm lin}^H$ of the form

$$\widehat{w}_{\rm lin}^H(\omega) = H^T\omega,$$

where $H \in \mathbf{R}^{m\times\nu}$ is the *contrast matrix*. We are given collection of contrast matrices

$$\mathfrak{H}_I = \{H_i,\ i = \{0, 1, \cdots, I\}\}$$

and associated collection of estimates

$$\mathfrak{W}_I = \left\{\widehat{w}_i = \widehat{w}_{\rm lin}^{H_i}(\omega),\ i = \{0, 1, ..., I\}\right\},$$

the latter being "tuned" to recover $w = Bz$ when $\mathfrak{m} = \mathfrak{m}(\delta_i)$. Our objective, under the assumption that $\mathfrak{m} \in \mathfrak{M}(\Delta)$, is to use observation ω to select the best estimate — the one with the smallest maximal over $\mathfrak{m}$ risk — among all $\widehat{w}_i$'s . To this end, we assume that given $\epsilon \in (0,1)$, $i \in \{0, ..., I\}$, and $j \in \{i+1, ..., I\}$, we can compute the quantiles upper bound $d_{ij}^{(i)}(\epsilon)$ such that if $z \in \mathcal{Z}(\delta_i)$

$$\mathrm{Prob}_\xi\{\|(H_i - H_j)^T(Az + \xi)\| \ge d_{ij}^{(i)}(\epsilon)\} \le \epsilon.$$

Let us consider the following construction.

1. For $0 \le i < j \le I$ and $\epsilon = \varepsilon/(I+1)$ compute the $d_{ij}^{(i)}(\epsilon)$'s.
2. We say that $i \in \{0, ..., I-1\}$ is admissible if

$$\forall j \in \{i+1, ..., I\},\ \ \|\widehat{w}_i - \widehat{w}_j\| \le d_{ij}^{(i)}(\epsilon).$$

   We define the aggregated estimate $\widehat{w}_{\rm lin}^{(a)}(\omega) = \widehat{w}_{\widehat{i}}(\omega)$ where $\widehat{i}$ is the smallest admissible i's, or we put $\widehat{w}_{\rm lin}^{(a)}(\omega) = \widehat{w}_I(\omega)$ when the set of such i is empty.

[4] By computationally tractable, we mean that a constraint of the type $x \in \mathcal{Z}(\delta)$ follows the criteria of Disciplined Convex Programming [16].

Let us denote $\rho_i(\epsilon)$ bounds for the ϵ-risk of estimates $\widehat{w}_i$ over $\mathfrak{m}(\delta_i)$, i.e.,

$$\mathrm{Risk}_\epsilon[\widehat{w}_i|\mathfrak{m}(\delta_i)] \le \rho_i(\epsilon), \quad i = 0, ..., I. \tag{2.8}$$

Proposition 2.2.1 *Suppose that "true model" $\mathfrak{m} \in \mathfrak{M}(\Delta)$ and let $\delta \in \Delta$ such that $\mathcal{Z} = \mathcal{Z}(\delta)$, and $\bar{i}$ be the smallest $i \in \{0, ..., I\}$ such that $\delta \le \delta_{\bar{i}}$. Estimate $\widehat{w}^{(a)}_{\mathrm{lin}}$ satisfies*

$$\mathrm{Risk}_\varepsilon[\widehat{w}^{(a)}_{\mathrm{lin}}|\mathfrak{M}(\delta)] \le \rho_{\bar{i}}(\epsilon) + \max_{i<\bar{i}} d^{(i)}_{i,\bar{i}}(\epsilon).$$

For proof, see 2.5.1.

Remark. One easily recognizes that estimate $\widehat{w}^{(a)}_{\mathrm{lin}}$ is nothing but the celebrated Lepski's adaptive estimate [5] adjusted for the situation. In the "classical" Lepski adaptation procedure we assume that the risk bounds $\rho_i(\epsilon)$, $i = 0, ..., I$, are available and the difference of estimates $w_i - w_j$ is compared to the sum $\rho_i(\epsilon) + \rho_j(\epsilon)$. One easily verifies that the quantity $\|(H_i - H_j)^T\omega\|$ may be bounded with $\rho_i(\epsilon) + \rho_j(\epsilon)$ for $j > i$. On the other hand, in some situations (and this is the case in the problem setting we discuss in Section 2.3.2), better bounds for the difference of estimates are readily available.

2.2.2 Aggregating polyhedral estimates

Polyhedral estimate, preliminaries Given an observation model $\mathfrak{m} = [A, B, \mathcal{Z}, (P_z, z \in \mathcal{Z})]$, $\epsilon \in (0, 1)$, and observation $\omega = Az + \xi$, a *polyhedral estimate* $\widehat{w}^H_{\mathrm{poly}}(\omega)$ [32, 33] of $w = Bz$ is

$$\widehat{w}^H_{\mathrm{poly}}(\omega) = B\widehat{z}(\omega), \quad \widehat{z}(\omega) \in \operatorname*{Argmin}_{z\in\mathcal{Z}} \|H^T[\omega - Az]\|_\infty. \tag{2.9}$$

We restrict its study to ϵ-admissible *contrast matrix* $H = [h_1, ..., h_m] \in \mathbf{R}^{m\times m}$ that is, matrices such that for all $z \in \mathcal{Z}$,

$$\mathrm{Prob}_\xi\{\|H^T\xi\|_\infty \ge 1\} \le \epsilon. \tag{2.10}$$

For such matrices, the polyhedral estimate's error allows for a straightforward upper bound.

Proposition 2.2.2 [33, Proposition 5.1] *Given $\epsilon > 0$ and a ϵ-admissible contrast matrix $H \in \mathbf{R}^{m\times m}$, let*

$$\mathfrak{p}_\epsilon[H] = \max_z \left\{\|Bz\| : z \in \mathcal{Z},\ \|H^TAz\|_\infty \le 1\right\}. \tag{2.11}$$

Then for every $z \in \mathcal{Z}$ the P_z-probability of the event

$$\|\widehat{w}(Az + \xi) - Bz\| > 2\mathfrak{p}_\epsilon[H]$$

does not exceed ϵ. Equivalently, the ϵ-risk of the polyhedral estimate $\widehat{w}^H_{\mathrm{poly}}(\cdot)$ does not exceed $2\mathfrak{p}_\epsilon[H]$.

We now return to the setting of the aggregation problem. The structure of the polyhedral estimate allows for a particularly simple implementation of the aggregated estimate in this case. Given $\varepsilon \in (0, 1)$, let H_i, $i = 0, ..., I$ be ϵ-admissible contrast matrices for $\epsilon = \varepsilon/(I + 1)$, and let

$$\overline{H} = [H_0, ..., H_I]; \tag{2.12}$$

note that $\overline{H}$ is ε-admissible. Consider the estimate $\widehat{w}^{(a)}_{\rm poly}$ as follows:

$$\widehat{w}^{(a)}_{\rm poly}(\omega) = B\widehat{z}, \quad [\widehat{\delta}, \widehat{z}] \in \mathop{\rm Argmin}_{\varkappa, v} \left\{ \varkappa : \|\overline{H}^T(\omega - Av)\|_\infty \leq 1, \ v \in \mathcal{Z}(\varkappa) \right\}. \tag{2.13}$$

We have the following straightforward result on its risk.

Proposition 2.2.3 *Let z be the "true"' signal vector, $z \in Z(\delta)$. Let also*

$$\mathfrak{p}_\varkappa[H] = \max_z \left\{ \|Bz\| : \ z \in \mathcal{Z}(\varkappa), \ \|H^T Az\|_\infty \leq 1 \right\}. \tag{2.14}$$

Then the ε-risk of the estimate $\widehat{w}^{(a)}_{\rm poly}$ satisfies

$$\mathrm{Risk}_\varepsilon[\widehat{w}^{(a)}_{\rm poly} | \mathfrak{M}(\delta)] \leq 2 \min_i \mathfrak{p}_\delta(H_i) \leq 2\mathfrak{p}_{\delta_{\bar{i}}}(H_{\bar{i}}) \tag{2.15}$$

where $\bar{i}$ is the smallest $i \in \{0, ..., I\}$ such that $\delta \leq \delta_i$.

For proof, see 2.5.1.

Remark 2.2.1 *Note that we do not aggregate estimates but contrasts here. The minimization procedure select the minimal set constraint $z \in \mathcal{Z}(\varkappa)$. In that case, we are sure that their is at least one feasible $z \in \mathcal{Z}(\overline{\delta})$ such that*

$$\forall i, \quad \left\|H_i^T Az\right\|_\infty \leq 1.$$

Provided that among the H_i's, one is tuned for the constraint $z \in \mathcal{Z}(\delta_{\widehat{i}})$, with $\widehat{i}$ the smallest i such that $\widehat{\delta} \leq \delta_i$, the resulting estimate enjoys the essentially the same risk upper bound as the Polyhedral Estimate associated with $H_{\widehat{i}}$. Finally, observe that in (2.15), in contrast with (2.8), we can replace $\mathfrak{p}_{\delta_{\bar{i}}}(H_{\bar{i}})$ by

$$\min_i \mathfrak{p}_\delta(H_i).$$

This hints at potentially better performances of the adaptive polyhedral estimate in practice compared to the polyhedral one that knows δ. In section 3.5 of chapter 3, we present numerical experiments demonstrating that this seems to be the case when signal sets $\mathcal{X}$ and $\mathcal{U}(\delta)$ are ellitopes.

2.3 Adaptive estimation on ellitopes

2.3.1 Robust estimation: signal set uncertainty

Before implementing in the present setting constructions of robust estimates $\widehat{w}^{(a)}$ described in Section 2.2, we provide the followingh refresher on Ellitopes.

Preliminaries: ellitopes

We provide here a refresher on a specific class of computationally tractable convex sets, called *ellitopes*, around wich computations of this section revolve.
Recall that, by definition [30, 32], a *basic ellitope* in $\mathbf{R}^n$ is a set of the form

$$\mathcal{X} = \{x \in \mathbf{R}^n : \ \exists t \in \mathcal{T} : \ z^T T_k z \leq t_k, \ k \leq K\}, \tag{2.16}$$

where $T_k \in \mathbf{S}^n_+$, $T_k \succeq 0$, $\sum_k T_k \succ 0$, and $\mathcal{T} \subset \mathbf{R}^K_+$ is a convex compact set with a nonempty interior which is monotone: whenever $0 \le t' \le t \in \mathcal{T}$ one has $t' \in \mathcal{T}$. We refer to K as *ellitopic dimension* of $\mathcal{X}$.

Clearly, every basic ellitope is a convex compact set with nonempty interior which is symmetric w.r.t. the origin. For instance,

A. Bounded intersection $\mathcal{X}$ of K centered at the origin ellipsoids/elliptic cylinders $\{x \in \mathbf{R}^n : x^T T_k x \le 1\}$ $[T_k \succeq 0]$ is a basic ellitope:

$$\mathcal{X} = \{x \in \mathbf{R}^n : \exists t \in \mathcal{T} := [0,1]^K : x^T T_k x \le t_k,\ k \le K\}$$

In particular, the unit box $\{x \in \mathbf{R}^n : \|x\|_\infty \le 1\}$ is a basic ellitope.

B. A $\|\cdot\|_p$-ball in $\mathbf{R}^n$ with $p \in [2, \infty]$ is a basic ellitope:

$$\{x \in \mathbf{R}^n : \|x\|_p \le 1\} = \big\{x : \exists t \in \mathcal{T} = \{t \in \mathbf{R}^n_+, \|t\|_{p/2} \le 1\} : \underbrace{x_k^2}_{x^T T_k x} \le t_k,\ k \le K\big\}.$$

In the present context, our interest for ellitopes is motivated by their special relationship with the optimization problem

$$\mathrm{Opt}_*(C) = \max_{x \in \mathcal{X}} x^T C x,\ C \in \mathbf{S}^n \tag{2.17}$$

of maximizing a homogeneous quadratic form over $\mathcal{X}$. As it is shown in [32], when $\mathcal{X}$ is an ellitope, (2.17) admits "reasonably tight" efficiently computable upper bound. Specifically,

Theorem 2.3.1 [32, Proposition 4.6] *Given ellitope (2.16) and matrix C, consider the quadratic maximization problem (2.17) along with its relaxation*[5]

$$\mathrm{Opt}(C) = \min_\lambda \Big\{\phi_{\mathcal{T}}(\lambda) : \lambda \ge 0, \sum\nolimits_k \lambda_k T_k - C \succeq 0\Big\} \tag{2.18}$$

The problem is computationally tractable and solvable, and $\mathrm{Opt}(C)$ is an efficiently computable upper bound on $\mathrm{Opt}_*(C)$. This upper bound is tight:

$$\mathrm{Opt}_*(C) \le \mathrm{Opt}(C) \le 3\ln(\sqrt{3}K)\mathrm{Opt}_*(C).$$

To the best of our knowledge, the first result of this type was established in [75] for $\mathcal{X}$ being an intersection of concentric elliptic cylinders/ellipsoids; in this case, (2.17) becomes a special case of quadratically constrained quadratic optimization problem, and (2.18) is the standard Shor's semidefinite relaxation (see, e.g., [13, Section 4.3]) of this problem. In [75] it is shown that the ratio $\mathrm{Opt}(C)/\mathrm{Opt}_*(C)$ indeed can be as large as $O(\ln(K))$, even when all $T_k = a_k a_k^T$ are of rank 1 and $\mathcal{X}$ is the polytope $\{x : |a_k^T x| \le 1, k \le \mathfrak{n}_x\}$.

Problem setting

Recall that we consider the problem of recovery of signal $z = [x; u]$ from the noisy observation (cf. (2.1))

$$\omega = A\underbrace{[x; u]}_{=z} + \xi \in \mathbf{R}^m, \quad A = [A_x, A_u] \in \mathbf{R}^{m \times n}.$$

[5] Here and below, we use notation $\phi_{\mathcal{S}}(\cdot)$ for the support function of a convex set $\mathcal{S} \subset \mathbf{R}^n$: for $y \in \mathbf{R}^n$,

$$\phi_{\mathcal{S}}(y) = \sup_{u \in \mathcal{S}} y^T s.$$

From now on we assume that the noise ξ is sub-Gaussian, $\xi \sim \mathcal{SG}(0, \sigma^2 I)$, i.e., for all $t \in \mathbf{R}^m$,

$$\mathbf{E}\left\{\exp\left(t^T\xi\right)\right\} \le \exp\left(\tfrac{\sigma^2}{2}\|t\|_2^2\right).$$

Given $\bar{\delta} > 0$ and unknown parameter $\delta \in \Delta = [0, \bar{\delta}]$, our aim is to estimate the linear image $w = Bz$, $B = [B_x, B_u] \in \mathbf{R}^{\nu\times n}$, of unknown signal $z = [x; u] \in \mathcal{Z}(\delta) = \mathcal{X} \times \mathcal{U}(\delta)$ where $\mathcal{X} \subset \mathbf{R}^{n_x}$ is a known ellitope

$$X = \left\{x \in \mathbf{R}^{n_x} : \ t \in \mathcal{T} : \ x^T T_k x \le t_k, k \le \mathfrak{n}_x\right\}, \tag{2.19}$$

while $\mathcal{U}(\delta) \subset \mathbf{R}^{n_u}$ belongs to a known family $\mathcal{U}(\cdot)$ of ellitopes

$$\mathcal{U}(\delta) = \left\{u \in \mathbf{R}^{n_u} : \ t \in \mathcal{V}(\delta) : \ u^T V_k u \le t_k, k \le \mathfrak{n}_u\right\}. \tag{2.20}$$

Here, $(\mathcal{V}(\delta))$, $\delta \in [0, \bar{\delta}]$, is a family of convex compact sets satisfying the standard requirements in the ellitope's definition and such that $\mathcal{V}(\delta) \subset \mathcal{V}(\delta')$ for $0 \le \delta < \delta' \le \bar{\delta}$.

We assume that the norm $\|\cdot\|$ is *co-ellitopic*; the latter means that the unit ball $\mathcal{B}$ of $\|\cdot\|$ is the polar of the given ellitope $\mathcal{B}_*$ (the unit ball of the norm $\|\cdot\|_*$ conjugate to $\|\cdot\|$),

$$\mathcal{B}_* = P\mathcal{Y}, \quad \mathcal{Y} = \left\{y \in \mathbf{R}^N, t \in \mathcal{M} : \ y^T M_k y \le s_k, k \le \mathfrak{n}_b\right\} \tag{2.21}$$

where P is surjective (i.e., is of rank ν).
Additionally, we will use the following notations associated with ellitopic descriptions (2.19),(2.20) and (2.21):

$$\forall \lambda \in \mathbf{R}^{\mathfrak{n}_x}_+, \ T(\lambda) := \sum_{k=1}^{\mathfrak{n}_x} \lambda_k T_k, \tag{2.22}$$

$$\forall \mu \in \mathbf{R}^{\mathfrak{n}_u}_+, \ V(\mu) := \sum_{k=1}^{\mathfrak{n}_u} \mu_k V_k, \tag{2.23}$$

$$\forall \kappa \in \mathbf{R}^{\mathfrak{n}_b}_+, \ M(\kappa) := \sum_{k=1}^{\mathfrak{n}_b} \kappa_k M_k. \tag{2.24}$$

For the present setting, we implement constructions of robust estimates $\widehat{w}^{(a)}$ described in Section 2.2.

2.3.2 Adaptive linear estimation

Linear estimates tuned for the ϵ-risk for $\mathfrak{m}(\delta)$

Let $H \in \mathbf{R}^{m\times\nu}$ and $\delta \in \Delta$ be fixed. For all $z \in \mathcal{Z}(\delta)$ the error of a linear estimate $\widehat{w}^H(\omega) = H^T\omega$ satisfies the equality

$$\|\widehat{w}^H(\omega) - w\| = \|H^T(Az + \xi) - Bz\| = \|[H^T A - B]z + H^T\xi\|.$$

Recall that $\mathcal{Z}(\delta) = \mathcal{X} \times \mathcal{U}(\delta)$ is a basic ellitope (as a direct product of basic ellitopes), its ellitopic description is given by

$$\mathcal{Z}(\delta) = \left\{ \begin{array}{ll} z = [x; u] \in \mathbf{R}^n : \exists r = [t; v] \in \mathcal{T} \times \mathcal{V}(\delta) : & x^T T_j x \le t_k, \ 1 \le k \le \mathfrak{n}_x, \\ & u^T V_k u \le v_l, \ 1 \le l \le \mathfrak{n}_u \end{array} \right\} \tag{2.25}$$

with

$$\phi_{\mathcal{T}\times\mathcal{V}(\delta)}([\lambda;\mu]) = \phi_{\mathcal{T}}(\lambda) + \phi_{\mathcal{V}(\delta)}(\mu), \quad \lambda \geq 0,\ \mu \geq 0.$$

Applying the result of Proposition 2.6.1 of Appendix 2.6 to matrices $V = H^TA - B$ and $W = H$, we obtain an upper bound for the ϵ-risk of the linear estimate $\widehat{w}^H_{\rm lin}$. Let

$$\mathfrak{r}_{\delta,\epsilon}(H) := \min_{\lambda,\mu,\chi,\kappa,\alpha,\Theta} \left\{ \phi_{\mathcal{T}}(\lambda) + \phi_{\mathcal{V}(\delta)}(\mu) + \phi_{\mathcal{M}}(\chi) + \phi_{\mathcal{M}}(\kappa) + \psi_\epsilon(\alpha,\Theta) :\ \lambda \geq 0, \chi \geq 0, \kappa \geq 0, \right. \tag{2.26}$$

$$\left. \left[\begin{array}{c|c|c} T(\lambda) & & \\ \hline & V(\mu) & \frac{1}{2}[H^TA - B]^TP \\ \hline \frac{1}{2}P^T[H^TA - B] & & M(\chi) \end{array}\right] \succeq 0,\ \left[\begin{array}{c|c} \Theta & \frac{1}{2}HP \\ \hline \frac{1}{2}P^TH^T & M(\kappa) \end{array}\right] \succeq 0,\ \alpha \geq 2\lambda_{\max}(\Theta) \right\}$$

where

$$\psi_\epsilon(\alpha,\Theta) := -\tfrac{\alpha}{2}\ln\mathrm{Det}(I - 2\alpha^{-1}\Theta) + \alpha\ln[\epsilon^{-1}]. \tag{2.27}$$

Proposition 2.3.1 *Assuming that $\delta > 0$ is known, let $\widehat{w}^H_{\rm lin}(\omega) = H^T\omega$ with some $H \in \mathbf{R}^{m\times\nu}$. Then*

$$\mathrm{Risk}_\epsilon[\widehat{w}^H_{\rm lin}|Z(\delta)] \leq \mathfrak{r}_{\delta,\epsilon}[H]. \tag{2.28}$$

Furthermore, the upper bound $\mathfrak{r}_{\delta,\epsilon}[H]$ is a convex, continuous and coercive function of the contrast matrix, and can be efficiently minimized w.r.t. H. Let now $H_*[= H_*(\delta,\epsilon)]$ be the corresponding minimizer. Then, the near-optimal linear estimate $\widehat{w}^{H_*}_{\rm lin}$ is such that

$$\mathrm{Risk}_\epsilon[\widehat{w}^{H_*}_{\rm lin}|Z(\delta)] \leq \mathfrak{r}_{\delta,\epsilon}\big[H_*(\delta,\epsilon)\big] =: \varrho_*(\delta,\epsilon).$$

Moreover, this estimate is nearly-minimax in the sense that

$$\varrho_*(\delta,\epsilon) \leq 36\sqrt{2\ln(32(\mathfrak{n}_x+\mathfrak{n}_u))\ln(32\mathfrak{n}_b)\ln(1/\epsilon)}\ \ \mathrm{RiskOpt}_\epsilon[Z(\delta)]. \tag{2.29}$$

For proof, see 2.5.2.

Quantile upper bound on the difference of two linear estimates

Next, let $\widehat{w}_1 = H_1^T\omega$ and $\widehat{w}_2 = H_2^T\omega$ where H_1 and H_2 are $m\times\nu$-matrices. Because

$$\|\widehat{w}_1 - \widehat{w}_2\| = \|H_1^T(Az+\xi) - H_2^T(Az+\xi)\| = \|(H_1-H_2)^TAz + (H_1-H_2)^T\xi\|,$$

when applying Proposition 2.6.1 to $V = (H_1-H_2)^TA$ and $W = H_1 - H_2$ we get

$$\mathrm{Prob}_\xi\left\{\|\widehat{w}_1 - \widehat{w}_2\| \geq \partial_{\delta,\epsilon}(H_1-H_2)\right\} \leq \epsilon$$

where

$$\partial_{\delta,\epsilon}(G) := \min_{\lambda,\mu,\chi,\kappa,\alpha,\Theta} \left\{ \phi_{\mathcal{T}}(\lambda) + \phi_{\mathcal{V}(\delta)}(\mu) + \phi_{\mathcal{M}}(\chi) + \phi_{\mathcal{M}}(\kappa) + \psi_\epsilon(\alpha,\Theta)) : \lambda \geq 0, \chi \geq 0, \kappa \geq 0, \right. \tag{2.30}$$

$$\left. \left[\begin{array}{c|c|c} T(\lambda) & & \frac{1}{2}GP \\ \hline & V(\mu) & \\ \hline \frac{1}{2}P^TG^T & & M(\chi) \end{array}\right] \succeq 0, \ \left[\begin{array}{c|c} \Theta & \frac{1}{2}GP \\ \hline \frac{1}{2}P^TG^T & M(\kappa) \end{array}\right] \succeq 0, \ \alpha \geq 2\lambda_{\max}(\Theta) \right\}$$

We are now equipped to implement the aggregated linear estimate following the program described in Section 2.2.1. Let us consider the following construction.

1. *[Initialization]* Given $\bar{\delta} > 0$, $\varepsilon \in (0,1)$, and $\vartheta > 0$, we compute upper risk bounds $\varrho_*(0,\epsilon) =: \underline{\rho}$ and $\varrho_*(\bar{\delta},\epsilon) =: \overline{\rho}$, I, and ϵ such that

$$\epsilon = \frac{\varepsilon}{I+1}, \quad I = \left\lceil \frac{\ln\left\lceil \overline{\rho}/\underline{\rho} \right\rceil}{\ln[1+\vartheta]} \right\rceil$$

   (here $\lceil a \rceil$ stands for the smallest integer $\geq a$); we compute values $\varrho_*(\delta_i,\epsilon) =: \rho_i$, $i = 1,...,I-1$ such that $\rho_0 = \underline{\rho}, ..., \rho_I = \overline{\rho}$ and

$$\rho_i/\rho_{i-1} \leq 1+\vartheta, \ \ i = 1,...,I. \tag{2.31}$$

   We also compute corresponding contrast matrices $H_i = H_*(\delta_i,\epsilon)$, estimates $w_i = H_i^T\omega$ of $w = Bz$ and values

$$d_{ij}^{(i)}(\epsilon) = \partial_{\delta_i,\epsilon}(H_i - H_j), \quad 0 \leq i < j \leq I.$$

   By the above, all corresponding quantities can be efficiently computed.

2. We run the aggregation routine from Section 2.2.1 and select the estimate $\widehat{w}_{\rm lin}^{(a)}$.

Proposition 2.3.2 *Suppose that $z \in \mathcal{Z}(\delta)$, and let $\bar{i}$ be the smallest i such that $\delta \leq \delta_i$. The estimate $\widehat{w}_{\rm lin}^{(a)}$ satisfies*

$$\mathrm{Risk}_\varepsilon[\widehat{w}_{\rm lin}^{(a)} | \mathcal{Z}(\delta)] \leq (3+2\vartheta)\, \varrho_*(\delta,\epsilon) \tag{2.32}$$

With a grid of size $I+1$ such as previously constructed, the following holds

$$\mathrm{Risk}_\varepsilon[\widehat{w}_{\rm lin}^{(a)} | \mathcal{Z}(\delta)] \leq (3+2\vartheta)\sqrt{\ln(e(1+I))}\varrho_*(\delta,\varepsilon). \tag{2.33}$$

For proof, see 2.5.2.

2.3.3 Design of the adaptive polyhedral estimate

Let $H \in \mathbf{R}^{m\times m}$, $H = [h_1,...,h_m]$; our current objective is to construct an efficiently computable bound for the quantity $2\mathfrak{p}_\epsilon[H]$ as defined in (2.11) which bounds the risk of the polyhedral estimate $\widehat{w}_{\rm poly}^H$ by Proposition 2.2.2. Note that

$$\|Bz\| = \max_{v\in\mathcal{B}_*} v^TBz = \max_{y\in\mathcal{Y}} y^TP^TBz,$$

and to bound the maximal risk of $\widehat{w}^H_{\rm poly}$ it suffices to approximate the maximum of the quadratic form

$$f([z;y]) = [z;y]^T \left[\begin{array}{c|c} & \frac{1}{2}B^TP \\ \hline \frac{1}{2}P^TB & \end{array}\right] [z;y]$$

over the set $\mathcal{F}(\delta) \times \mathcal{Y}$ where $\mathcal{F}(\delta) = \{z \in \mathcal{Z}(\delta) : \|H^TAz\|_\infty \le 1\}$. The crucial observation here is that $\mathcal{F}(\delta)$ is a basic ellitope; to see this, it suffices to note that it is an intersection of the ellitope $Z(\delta)$ and the elliptic cylinder

$$\{z : \exists p \in \mathcal{P} : (h_k^TAz)^2 \le p_k,\ k = 1, ..., m\}, \qquad \mathcal{P} := [0,1]^m.$$

When taking into account the ellitopic description (2.25) of $\mathcal{Z}(\delta)$, $\mathcal{F}(\delta)$ becomes

$$\begin{aligned}\mathcal{F}(\delta) = \Big\{z = [x;u] \in \mathbf{R}^n : \exists r = [t,v,p] \in \mathcal{T} \times \mathcal{V}(\delta) \times \mathcal{P} : x^TT_jx \le t_j,\ j = 1, ..., \mathfrak{n}_x,\\ u^TV_ku \le v_k,\ k = 1, ..., \mathfrak{n}_u,\ (h_i^TAz)^2 \le p_i,\ i = 1, ..., m\Big\}.\end{aligned}$$

When applying the result of Proposition 2.3.1 to the bounding of $f([z;y])$ on the basic ellitope $\mathcal{F}(\delta) \times \mathcal{Y}$ we arrive at

$$\begin{aligned}\mathfrak{p}_\epsilon[H] \le \min_{\lambda,\mu,\chi,\kappa} \Bigg\{ &\phi_{\mathcal{T}}(\lambda) + \phi_{\mathcal{V}(\delta)}(\mu) + \sum_i \chi_i + \phi_{\mathcal{M}}(\kappa) :\ \lambda \ge 0, \mu \ge 0, \chi \ge 0, \kappa \ge 0,\\ &\left[\begin{array}{c|c} \left[\begin{array}{c|c} T(\lambda) & \\ \hline & V(\mu) \end{array}\right] + \sum_i \chi_i A^Th_jh_j^TA & \frac{1}{2}B^TM \\ \hline \frac{1}{2}P^TB & M(\kappa) \end{array}\right] \succeq 0 \Bigg\}.\end{aligned}$$

Now, to build the "presumably good" contrast matrix H we need to minimize the just built upper bound over ϵ-admissible H. In the case of sub-Gaussian $\xi \sim \mathcal{SG}(0, I)$, the admissibility relationship (2.10) clearly holds for $H = [h_1, ..., h_m]$ satisfying $\|h_j\|_2 \le \varkappa^{-1}$, $j = 1, ..., m$, where $\varkappa = \sqrt{2\ln[2m/\epsilon]}$. On the other hand, let $\Theta \in \mathbf{S}^m_+$ and let $\Theta = \Gamma{\rm Diag}(\chi_j)\Gamma^T$ be the eigenvalue decomposition of Θ. When setting $H = \Gamma/\varkappa$ we ensure the ϵ-admissibility of H, with

$$A^T\Theta A = A^TH{\rm Diag}(v_j)H^TA = \sum_j \underbrace{\varkappa^2 v_j}_{\chi_j} Ah_jh_j^TA.$$

We conclude that if for $\Theta \succeq 0$,

$$\begin{aligned}\overline{\mathfrak{p}}_{\delta,\epsilon}[\Theta] := \min_{\lambda,\mu,\kappa} \Big\{ &\phi_{\mathcal{T}}(\lambda) + \phi_{\mathcal{V}(\delta)}(\mu) + \phi_{\mathcal{M}}(\kappa) + \varkappa^2{\rm Tr}(\Theta) :\ \lambda \ge 0, \mu \ge 0, \kappa \ge 0,\\ &\left[\begin{array}{c|c} \left[\begin{array}{c|c} T(\lambda) & \\ \hline & V(\mu) \end{array}\right] + A^T\Theta A & \frac{1}{2}B^TM \\ \hline \frac{1}{2}P^TB & K(\kappa) \end{array}\right] \succeq 0 \Big\}. \qquad (2.34)\end{aligned}$$

and $H = \varkappa^{-1}\Gamma$ where $\Theta = \Gamma{\rm Diag}(\chi_j)\Gamma^T$ is the eigenvalue decomposition of Θ, one has

$$\mathfrak{p}_\epsilon[H] \le \overline{\mathfrak{p}}_{\delta,\epsilon}[\Theta].$$

Note that as a result of partial minimization of a jointly convex problem (2.34) $\mathfrak{p}_{\delta,\epsilon}[\cdot]$ is a well structured convex function of $\Theta \succeq 0$ and can be efficiently minimized in Θ.

Proposition 2.3.3 *In the situation of this section, consider the optimization problem*

$$\overline{\mathfrak{p}}_*(\delta,\epsilon) = \min_{\Theta}\{\overline{\mathfrak{p}}_{\delta,\epsilon}[\Theta] : \Theta \succeq 0\}. \tag{2.35}$$

The problem is solvable; let Θ_* be an optimal solution, and let $H_*(\delta,\epsilon) = \varkappa\Gamma$ where $\Theta_* = \Gamma\mathrm{Diag}(\chi_j)\Gamma^T$ is the eigenvalue decomposition of Θ_*. Then

$$\mathfrak{p}_\epsilon[H_*(\delta,\epsilon)] \le \overline{\mathfrak{p}}_*(\delta,\epsilon).$$

Proof. Problem (2.35) is a convex optimization problem with positive homogeneous of degree 1 objective, and as such is solvable. The bound $\mathfrak{p}_\epsilon[H_*] \le \overline{\mathfrak{p}}_*(\delta,\epsilon)$ is a straightforward consequence of the derivation of this section. □

We can now implement the aggregated polyhedral estimate.

1. *[Initialization]* Given $\bar\delta > 0$, $\varepsilon \in (0,1)$, and $\vartheta > 0$, we compute upper risk bounds $\overline{\mathfrak{p}}_*(0,\epsilon) =: \underline{\rho}$ and $\overline{\mathfrak{p}}_*(\bar\delta,\epsilon) =: \overline{\rho}$, I, and ϵ such that

$$\epsilon = \frac{\varepsilon}{I+1}, \quad I = \left\lceil \frac{\ln\left[\overline{\rho}/\underline{\rho}\right]}{\ln[1+\vartheta]} \right\rceil;$$

we compute values $\mathfrak{p}_*(\delta_i,\epsilon) =: \rho_i$, $i = 1,...,I-1$ such that $\rho_0 = \underline{\rho}, ..., \rho_{I+1} = \overline{\rho}$ and

$$\rho_i/\rho_{i-1} \le 1+\vartheta, \quad i = 1,...,I. \tag{2.36}$$

We also compute corresponding contrast matrices $H_i = H_*(\delta_i,\epsilon)$ and assemble the matrix $\overline{H} = [H_0,...,H_I]$.

2. We compute the aggregated estimate $\widehat{w}^{(a)}_{\mathrm{poly}}(\omega)$ in (2.13).

Proposition 2.3.4 *Suppose that $z \in \mathcal{Z}(\delta)$, and let $\bar\imath$ be the smallest i such that $\delta \le \delta_{\bar\imath}$. The estimate $\widehat{w}^{(a)}_{\mathrm{poly}}(\omega)$ satisfies*

$$\mathrm{Risk}_\varepsilon[\widehat{w}^{(a)}_{\mathrm{poly}}|\mathcal{Z}(\delta)] \le 2\mathfrak{p}_{\delta,\varepsilon}[\overline{H}] \le 2\sqrt{1+\vartheta}\,\overline{\mathfrak{p}}_*(\delta\epsilon) \le 2\sqrt{(1+\vartheta)\ln(e(I+1))}\,\overline{\mathfrak{p}}_*(\delta,\varepsilon). \tag{2.37}$$

For $\varepsilon \le \frac{1}{8}$, this estimate is nearly-minimax, in the sense that

$$\frac{\mathrm{Risk}_\varepsilon[\widehat{w}^{(a)}_{\mathrm{poly}}|\mathcal{Z}(\delta)]}{\mathrm{RiskOpt}_\varepsilon(\mathcal{Z}(\delta))} \le 48\sqrt{2(1+\vartheta)\ln(e(I+1))\ln(2m/\varepsilon)\ln(32(n_x+n_u))\ln(32n_b)}. \tag{2.38}$$

For proof, see 2.5.2.

2.4 Adaptive estimate: case of Euclidean norm

Suppose that, similarly to what we have seen in Section 2.2, we are given $\bar\delta > 0$, $\varepsilon \in (0,1)$, a collection of either linear estimates $\mathfrak{W}_I = \{w_i,\ i = \{0,1,...,I\}\}$ or polyhedral matrix contrast $\mathfrak{H}_I = \{H_i,\ i = \{0,1,...,I\}\}$, tuned for subsets $(\mathcal{Z}(\delta_i))_{i=0}^I$, with

$$0 = \delta_0 < \delta_1 < ... < \delta_I = \bar\delta.$$

Our objective is to select in $\mathfrak{W}_I$ the estimate $\widehat{w}^{(a)}_{\mathrm{lin}}$ with the smallest *Euclidean risk*

$$\mathrm{Risk}_\epsilon^{\|\cdot\|_2}[\widehat{w}|\mathcal{Z}(\delta)] := \sup_{z\in\mathcal{Z}(\delta)} \inf\left\{\rho : \mathrm{Prob}_\xi\{\|Bz - \widehat{w}(\omega)\|_2 > \rho\} \le \epsilon\right\}. \tag{2.39}$$

2.4.1 Aggregating linear estimates

The proposed aggregation procedure relies upon the following simple observation. Let w_i, w_j be points in $\mathbf{R}^\nu$. Then, w_i is the closest to w among w_i and w_j, if and only if the projection of w on the line passing through w_i, w_j is closer to w_i than the middle point $\overline{w}_{ij}$ of the segment $[w_i, w_j]$. In other words, if and only if

$$r_{ij} = v_{ij}^T(w - \underbrace{\tfrac{1}{2}[w_i + w_j]}_{=:\overline{w}_{ij}}) \le 0, \quad v_{ij} = w_j - w_i.$$

In the present situation where w_i, w_j are linear estimates, $w_i = H_i^T\omega$ and $w_j = H_j^T\omega$, evaluating $\pi_{ij} = v_{ij}^T w$ amounts to estimating the quadratic (in variables z, ξ) functional

$$\pi_{ij} = [(H_j - H_i)^T\omega]^T Bz = z^T A^T(H_j - H_i)Bz + \xi^T(H_j - H_i)Bz. \tag{2.40}$$

This is why we assume for now that given $\epsilon \in (0,1)$, $i, j \in \{0, ..., I\}$, $j \ne i$, we can compute the quantities $\widehat{\pi}_{ij}^{(\ell)}$ and $\mathfrak{d}_{ij}^{(\ell)}(\epsilon)$ such that for ω stemming from $\mathfrak{m}(\delta_\ell)$,

$$\mathrm{Prob}_\xi\{|\widehat{\pi}_{ij}^{(\ell)} - \pi_{ij}| \ge \mathfrak{d}_{ij}^{(\ell)}(\epsilon)\} \le \epsilon. \tag{2.41}$$

Let us consider the following procedure.

1. For $0 \le i \ne j \le I$, $\ell \le I$, and $\epsilon = \frac{\varepsilon}{\overline{I}}$, $\overline{I} = 1 + \frac{1}{2}I^2(I-1)$, compute the quantities $\widehat{\pi}_{ij}^{(\ell)}$ and $\mathfrak{d}_{ij}^{(\ell)}(\epsilon)$.

2. For all pairs (i, ℓ), $0 \le i, \ell \le I$, and all $j \ne i$, we compute

$$\widehat{r}_{ij}^{(\ell)} = \widehat{\pi}_{ij}^{(\ell)} - v_{ij}^T\bar{w}_{ij} - \mathfrak{d}_{ij}^{(\ell)}(\epsilon),$$

   and we say that index i is ℓ-admissible if $\widehat{r}_{ij}^{(k)} \le 0$ for all $j \ne i$ and $k \ge \ell$. We define the score of i as the smallest ℓ such that i is ℓ-admissible.

3. We define the aggregated estimate $\widetilde{w}_{\mathrm{lin}}^{(a)}(\omega) = w_{\widetilde{i}}(\omega)$ where $\widetilde{i}$ has the lowest score, or we put, say, $\widehat{w}_{\mathrm{lin}}^{(a)}(\omega) = w_1(\omega)$ when the set of admissible i is empty.

Proposition 2.4.1 *Assuming that observation ω stems from z, let ℓ_* be the smallest $\ell \in \{1, ..., I\}$ such that $\mathcal{Z}(\delta) \in \mathcal{Z}(\delta_\ell)$. Denote $\mathfrak{d}_{\ell_*} = \max_{\substack{0 \le i,j \le I,\\ i \ne j}} \mathfrak{d}_{ij}^{(\ell_*)}$. One has*

$$\|w_{\widetilde{i}}(\omega) - w\|_2^2 \le \min_i \|w_i - w\|_2^2 + 4\mathfrak{d}_{\ell_*}.$$

with probability at least $1 - \varepsilon$. In particular, one has

$$\mathrm{Risk}_\varepsilon^{\|\cdot\|_2}[\widetilde{w}_{\mathrm{lin}}^{(a)}|\mathcal{Z}(\delta)] \le \mathrm{Risk}_\epsilon^{\|\cdot\|_2}[w_{\ell_*}|\mathcal{Z}(\delta_{\ell_*})] + 4\mathfrak{d}_{\ell_*}. \tag{2.42}$$

For proof, see 2.5.3.

Remark. The construction of this section may be modified for the case where the family of models $\mathfrak{M}(\delta)$ is not monotone in δ (cf., e.g., [76] and [77, Section 3.3]). On the other hand, in general, we cannot guarantee that the additive term $\mathfrak{d}_{\ell_*}$ in the risk bound of Proposition 2.4.1 is small if compared to the maximal estimation risk and bound (2.42) improves over the corresponding bound of Proposition 2.2.1 for the aggregation procedure of Section 2.2.1. The good news is that in the situation where the family of models is monotone and we can apply both approaches, we can have "the best of both worlds" by choosing the aggregated estimate from the intersection of sets of estimates which are admissible for both routines.

2.4.2 Construction of the Linear Estimate

In the setting of Section 2.3.2, construction of the robust estimate may be refined in the case of estimation error measured with Euclidean norm. Let $H \in \mathbf{R}^{\nu\times m}$, and let $\delta \in [0, \bar{\delta}]$ be fixed. We now apply the bound of Proposition 2.6.2 of the appendix to bound the squared error

$$\|\widehat{w}^H(\omega) - Bz\|_2^2 = \|[H^TA - B]z + H^T\xi\|_2^2$$

of the linear estimate $\widehat{w}^H_{\rm lin}(\omega) = H^T\omega$. When applied with $V = H^TA - B$ and $W = H$, with the basic ellitope $\mathcal{Z}$ replaced with $\mathcal{Z}(\delta)$ as in (2.25), Proposition 2.6.2 results in the following upper bound for the ϵ-risk of the linear estimate $\widehat{w}^H_{\rm lin}$.

$$\mathfrak{r}^{(2)}_{\delta,\epsilon}[H] = \min_{\lambda,\mu,\alpha,\Theta}\left\{\phi_{\mathcal{T}}(\lambda) + \phi_{\mathcal{V}(\delta)}(\mu) - \tfrac{\alpha}{2}\ln\mathrm{Det}(I - 2\alpha^{-1}\Theta) + \alpha\ln[\epsilon^{-1}] : \lambda \geq 0, \mu \geq 0, \right. \qquad (2.43)$$
$$\left. \alpha \geq 2\lambda_{\max}(\Theta), \left[\begin{array}{c|c} \Theta & H \\ \hline H^T & I \end{array}\right] \succeq 0, \left[\begin{array}{c|c} \begin{array}{c|c} T(\lambda) & \\ \hline & V(\mu) \end{array} & A^TH - B^T \\ \hline H^TA - B & I - 2\alpha^{-1}\Theta \end{array}\right] \succeq 0 \right\}.$$

and

$$\overline{\mathfrak{r}}^{(2)}_{\delta,\epsilon}[H] = \min_{\lambda,\mu,\alpha,\Theta,\Lambda,\Upsilon}\left\{\phi_{\mathcal{T}}(\lambda) + \phi_{\mathcal{V}(\delta)}(\mu) + \mathrm{Tr}[\Theta + \Lambda] + \alpha\ln[\epsilon^{-1}] : \lambda \geq 0, \mu \geq 0, \alpha \geq 2\lambda_{\max}(\Theta)\right.$$
$$\left. \left[\begin{array}{c|c} \Theta & H \\ \hline H^T & I \end{array}\right] \succeq 0, \left[\begin{array}{c|c} \Lambda & \Theta \\ \hline \Theta & \alpha I - 2\Theta \end{array}\right] \succeq 0, \left[\begin{array}{c|c} \begin{array}{c|c} T(\lambda) & \\ \hline & V(\mu) \end{array} & A^TH - B^T \\ \hline H^TA - B & I - 2\alpha^{-1}\Theta \end{array}\right] \succeq 0 \right\}.$$

Proposition 2.4.2 *Assuming that $\delta > 0$ is known, let $\widehat{w}^H_{\rm lin}(\omega) = H^T\omega$ with some $H \in \mathbf{R}^{m\times\nu}$. Then*

$$\mathrm{Risk}^{\|\cdot\|_2}_\epsilon[\widehat{w}^H_{\rm lin}|Z(\delta)] \leq \mathfrak{r}^{(2)}_{\delta,\epsilon}[H] \leq \overline{\mathfrak{r}}^{(2)}_{\delta,\epsilon}[H]. \qquad (2.44)$$

Furthermore, $\mathfrak{r}^{(2)}_{\delta,\epsilon}[H]$ and $\overline{\mathfrak{r}}^{(2)}_{\delta,\epsilon}[H]$ are convex, continuous and coercive functions of H, and can be efficiently minimized. Let now $H_*[= H_*(\delta,\epsilon)]$ and $\overline{H}_*[= \overline{H}_*(\delta,\epsilon)]$ be corresponding minimizers. Then

$$\mathrm{Risk}^{\|\cdot\|_2}_\epsilon\big[\widehat{w}^{H_*}_{\rm lin}|Z(\delta)\big] \leq \mathfrak{r}^{(2)}_{\delta,\epsilon}\big[H_*(\delta,\epsilon)\big][=: \varrho^{(2)}_*(\delta,\epsilon)].$$

and

$$\mathrm{Risk}^{\|\cdot\|_2}_\epsilon\big[\widehat{w}^{\overline{H}_*}_{\rm lin}|Z(\delta)\big] \leq \overline{\mathfrak{r}}^{(2)}_{\delta,\epsilon}\big[\overline{H}_*(\delta,\epsilon)\big][=: \bar{\varrho}^{(2)}_*(\delta,\epsilon)].$$

2.4.3 Estimating quadratic functionals over ellitopes

To streamline the presentation, we assume in this section that observation noise ξ is standard Gaussian, i.e., $\xi \sim \mathcal{N}(0, I)$.[6]

The construction in Section 2.4.1 allows to reduce building an aggregated estimate in the Euclidean case to estimating quadratic functionals (cf. (2.40))

$$\pi_{ij} = [(H_j - H_i)^T\omega]^T Bx = z^T A^T(H_j - H_i)Bz + \xi^T(H_j - H_i)Bz, \quad 0 \le i < j \le I,$$

uniformly over $z \in \mathcal{Z}(\delta)$. Let the pair i, j be fixed; for the estimate

$$\widehat{\pi}_{ij} = \omega^T F\omega - \mathrm{Tr}[F] + \varkappa, \quad F \in \mathbf{S}^m,\ \varkappa \in \mathbf{R},$$

we have

$$\begin{aligned}\widehat{\pi}_{ij} - \pi_{ij} = z^T \underbrace{\left[A^T FA - \tfrac{1}{2}(A^T(H_j - H_i)B + B^T(H_j - H_i)^T A)\right]}_{=:C(F)} z \\ + \xi^T \underbrace{[2FA - (H_j - H_i)B]}_{=:D(F)} z + \xi^T F\xi - \mathrm{Tr}[F] + \varkappa.\end{aligned}$$

Recall that $\mathcal{Z}(\delta) = \mathcal{X} \times \mathcal{U}(\delta)$ is a direct product of basic ellitopes and its ellitopic description is (cf. (2.25))

$$\mathcal{Z}(\delta) = \Big\{z = [x; u] \in \mathbf{R}^n : \exists r = [t, t'] \in \mathcal{T} \times \mathcal{V}(\delta) : x^T T_j x \le t_j,\ j \le \mathfrak{n}_x,\ u^T V_k u \le t'_k,\ k \le \mathfrak{n}_u\Big\}.$$

By Lemma 1 of the appendix,

$$\mathrm{Prob}\big\{|\widehat{\pi}_{ij} - \pi_{ij}| \ge \rho^\delta(F, \epsilon)\big\} \le \epsilon \quad \forall z \in \mathcal{Z}(\delta)$$

where

$$\rho^\delta(F, \epsilon) = \tfrac{1}{2}[\Phi^\delta_+(F; \epsilon) + \Phi^\delta_-(F; \epsilon)], \quad \varkappa = \tfrac{1}{2}[\Phi^\delta_-(F; \epsilon) - \Phi^\delta_+(F; \epsilon)]$$

and

$$\begin{aligned}\Phi^\delta_+(F; \epsilon) = \min_{\alpha,\lambda,\mu}\Big\{&-\tfrac{\alpha}{2}\ln\mathrm{Det}(I - 2F/\alpha) + \tfrac{1}{2}[\phi_{\mathcal{T}}(\lambda) + \phi_{\mathcal{V}(\delta)}(\mu)] + \alpha\ln(2/\epsilon) - \mathrm{Tr}[F], \\ &\alpha I \succeq 2F,\ \lambda \ge 0, \mu \ge 0,\ \left[\begin{array}{c|c|c} \left[\begin{array}{c|c} T(\lambda) & \\ \hline & V(\mu)\end{array}\right] - 2C(F) & D(F)^T \\ \hline D(F) & \alpha I - 2F\end{array}\right] \succeq 0\Big\}, \\ \Phi^\delta_-(F; \epsilon) \le \min_{\alpha,\lambda,\mu}\Big\{&-\tfrac{\alpha}{2}\ln\mathrm{Det}(I + 2F/\alpha) + \tfrac{1}{2}[\phi_{\mathcal{T}}(\lambda) + \phi_{\mathcal{V}(\delta)}(\mu)] + \alpha\ln(2/\epsilon) - \mathrm{Tr}[F], \\ &\alpha I \succeq -2F,\ \lambda \ge 0, \mu \ge 0,\ \left[\begin{array}{c|c|c} \left[\begin{array}{c|c} T(\lambda) & \\ \hline & V(\mu)\end{array}\right] + 2C(F) & D(F)^T \\ \hline D(F) & \alpha I + 2F\end{array}\right] \succeq 0\Big\}.\end{aligned}$$

[6] The argument to follow relies upon tight bounds on the concentration of indefinite quadratic forms of the observation noise ξ. In the sub-Gaussian setting, corresponding inequalities may be obtained if we assume that covariance matrix $\Sigma \succ 0$ of ξ is known; the required concentration bound may be obtained in this case using the decoupling argument (cf., e.g., [78] in the situation in which ξ_i's are independent, or [79] in the general case). Utilizing such inequalities results in the bounds for the accuracy of aggregation of the form which is similar to what is obtained below in the case of Gaussian ξ, albeit with significantly degraded constants.

Let us now consider the optimization problem

$$\rho_*^\delta(\epsilon) := \min_F \left[\rho^\delta(F,\epsilon) = \tfrac{1}{2}(\Phi_+^\delta(F;\epsilon) + \Phi_-^\delta(F;\epsilon))\right]. \tag{2.45}$$

Because $C(\cdot)$ and $D(\cdot)$ are affine, $\rho(F,\epsilon)$ is convex and continuous in F. Furthermore, observe that functions

$$\begin{aligned}\nu_+(F,\alpha;\epsilon) &= -\tfrac{\alpha}{2}\ln\mathrm{Det}(I-2F/\alpha) + \alpha\ln(2/\epsilon) - \mathrm{Tr}[F]\\ \nu_-(F,\alpha;\epsilon) &= -\tfrac{\alpha}{2}\ln\mathrm{Det}(I+2F/\alpha) + \alpha\ln(2/\epsilon) + \mathrm{Tr}[F]\end{aligned}$$

are positive homogeneous in F and α. We conclude that $\rho^\delta(F,\epsilon)$ is coercive in F, implying that (2.45) is solvable. We arrive at the following conclusion.

Proposition 2.4.3 *Assuming that observation ω stems from z, let ℓ_* be the smallest $\ell \in \{1,...,I\}$ such that $\mathfrak{m} \in \mathfrak{M}(\delta_\ell)$. Denote $\mathfrak{d}_{\ell_*} = \max_{\substack{0\le i,j\le I,\\ i\ne j}} \mathfrak{d}_{ij}^{(\ell_*)}$, with $\mathfrak{d}_{ij}^{(\ell_*)} := \min_F \left[\rho^{\delta_{\ell_*}}(H_i - H_j,\epsilon)\right]$. One has*

$$\|w_{\widetilde{i}}(\omega) - w\|_2^2 \le \min_i \|w_i - w\|_2^2 + 4\mathfrak{d}_{\ell_*}.$$

with probability at least $1-\varepsilon$. In particular, one has

$$\mathrm{Risk}_\varepsilon^2[\widetilde{w}_{\mathrm{lin}}^{(a)}|\mathfrak{M}(\delta)] \le \varrho_*^{(2)}(\delta_{\ell_*},\epsilon) + 4\mathfrak{d}_{\ell_*}. \tag{2.46}$$

Proof First part of the proof stems directly from construction of estimate $\widetilde{\omega}_{lin}^{(a)}$, while second part of it is completely analogous to proofs for corollaries 2.32 and 2.37. □

Remark. As constructed above, $\widehat{\pi}_{ij}$ is a specific estimate of essentially generic quadratic form $\pi_{ij} = z^TCz + z^TD\xi$ in the setting of this section, i.e., Gaussian observation (2.1), general matrices C and D, and $z \in \mathcal{Z}$ where $\mathcal{Z}$ ia a basic ellitope.[7] We do not know any "truly general" optimality result supporting upper accuracy bounds of Proposition 2.4.3 when estimating quadratic functionals from Gaussian observation (2.1). To the best of our knowledge, the most general (and close in spirit to our approach) are results of [28, 80] on estimating quadratic functions in white noise model in the setting where the matrix of the functional is diagonal and the set of signals is orthosymmetric and quadratically convex. In terms of assumptions of this section, those results correspond to the case of sensing matrix $A = I$ and the ellitope $\mathcal{Z}$ of the form

$$\mathcal{Z} = \{z \in \mathbf{R}^n : \exists q \in \mathcal{Q} : z_k^2 \le q_k,\ k \le n\}.$$

It can be shown that in this situation the risk bound yielded by Proposition 2.4.3 is within a logarithmic in $1/\epsilon$ and n factor (stemming from passing from expected risk to ϵ-risk) of the minimax risk (cf. Theorem 6 of [28]). On the other hand, it is well known that "quadratic estimates" like those built in this section are suboptimal in the case of general geometry of the signal set even when estimating signal energy (case of $C = I$ and $D = 0$) from direct observations (see, e.g., [81, 82]).

[7] The latter restriction is clearly not restrictive due to the total freedom of the choice of matrices C and D.

2.5 Proofs

2.5.1 Proofs of Section 2.2

Proof of 2.2.1. Let event Ξ be such that for all realizations $\xi \in \Xi$,

$$\|w_{\bar{i}}(\omega) - Bz\| \le \rho_{\bar{i}}(\epsilon) \ \text{ and } \ \|w_{\bar{i}}(\omega) - w_j(\omega)\| = \|(H_{\bar{i}} - H_j)^T\omega\| \le d_{\bar{i},j}(\epsilon),\ \forall j \in \{\bar{i}+1, ..., I\}.$$

By construction, we have $\mathrm{Prob}_\xi\{\Xi\} \ge 1 - (I+1)\epsilon = 1 - \varepsilon$. Let now $\xi \in \Xi$ be fixed. Because $z \in \mathcal{Z}(\delta) \subset \mathcal{Z}(\delta_{\bar{i}})$, $\bar{i}$ is admissible, implying that $\widehat{i} \le \bar{i}$. On the other hand,

$$\|w_{\widehat{i}} - Bz\| \le \|w_{\bar{i}} - Bz\| + \|w_{\widehat{i}} - w_{\bar{i}}\| \le \rho_{\bar{i}}(\epsilon) + d_{\widehat{i},\bar{i}}(\epsilon) \le \rho_{\bar{i}}(\epsilon) + \max_{i<\bar{i}} d_{i,\bar{i}}(\epsilon). \qquad \square$$

Proof of 2.2.3. Because $\overline{H}$ is ε-admissible, there is a set Ξ such that $\mathrm{Prob}_\xi\{\Xi\} \ge 1 - \varepsilon$ and $\|\overline{H}^T\xi\|_\infty \le 1$ for all $\xi \in \Xi$. Let now $\xi \in \Xi$, so that $\|\overline{H}^T(\omega - Az_*)\|_\infty = \|\overline{H}^T\xi\|_\infty \le 1$, implying that z is feasible for (2.13). We conclude that $\widehat{\delta} \le \delta$ and $\widehat{z} \in \mathcal{Z}(\widehat{\delta}) \subset \mathcal{Z}(\delta)$. Thus, $\widehat{z} - z \in 2\mathcal{Z}(\delta)$, and, by construction,

$$\|\overline{H}^T A(\widehat{z} - z)\|_\infty \le \|\overline{H}^T(A\widehat{z} - \omega)\|_\infty + \|\overline{H}\xi\|_\infty \le 2,$$

which in turn implies that risk of estimate $\widehat{w}^a_{\mathrm{poly}}$ is bounded by

$$\begin{aligned}
2\mathfrak{p}_\delta(\overline{H}) &= 2\max_z \left\{\|Bz\| :\ z \in \mathcal{Z}(\delta),\ \left\|\overline{H}^T Az\right\|_\infty \le 1\right\} \\
&= 2\max_z \left\{\|Bz\| :\ z \in \mathcal{Z}(\delta),\ \forall i \in \{0, 1, \cdots, I\},\ \left\|H_i^T Az\right\|_\infty \le 1\right\} \\
&\le 2\min_i \max_z \left\{\|Bz\| :\ z \in \mathcal{Z}(\delta),\ \left\|H_i^T Az\right\|_\infty \le 1\right\} \\
&= 2\mathfrak{p}_\delta(H_i) \le 2\mathfrak{p}_\delta(H_{\bar{i}}).
\end{aligned} \qquad \square$$

2.5.2 Proofs of section 2.3

Proof of 2.3.1. The proposed construction of the linear estimates directly implies (2.28). What is left to prove is near-optimality result (2.29). The latter is a direct consequence of [32, Proposition 4.5]. In their proof, authors show that for observations

$$\omega = Ax + \xi, \quad x \in \mathcal{X}, \quad \xi \sim \mathcal{N}(0, Q), \ \ Q \in \mathbf{S}^m_+,$$

and both $\mathcal{X}$ and $\mathcal{B}_{\|\cdot\|_*}$ being ellitopes of respective dimensions K and L,

$$\mathrm{RiskOpt}^{\|\cdot\|}_{\frac{1}{8}}[\mathcal{X}] \ge \frac{\mathrm{Opt}}{24\sqrt{\ln(32K)\ln(32L)}},$$

where

$$\mathrm{Opt} = \min_{H,\Theta,\kappa,\lambda,\chi} \left\{ \phi_{\mathcal{T}}(\lambda) + \phi_{\mathcal{M}}(\kappa) + \phi_{\mathcal{M}}(\chi) + \mathbf{Tr}(\Theta Q) : \begin{array}{l} \left[\begin{array}{c|c} R(\lambda) & \frac{1}{2}[B^T - A^T H]M \\ \hline \frac{1}{2}P^T[B - H^T A] & M(\kappa) \end{array}\right] \succeq 0, \\ \left[\begin{array}{c|c} \Theta & \frac{1}{2}HP \\ \hline \frac{1}{2}(HP)^T & M(\chi) \end{array}\right] \succeq 0, \ \ \lambda,\ \chi,\ \kappa \ge 0 \end{array} \right\}.$$

As the deviations upper bound $\psi_\varepsilon(\Theta)$, $\widetilde{\psi}_\varepsilon(\Theta)$, $\overline{\psi}_\varepsilon(\Theta)$ proposed in 2.6.1 are all smaller than $(1 + \sqrt{2\ln(1/\epsilon)})\mathbf{Tr}(\Theta)$, we have when considering $Q = \sigma^2 I$ and $\mathcal{X} = \mathcal{Z}(\delta)$ that

$$\mathrm{Opt} \ge \frac{\varrho_*(\delta, \epsilon)}{1 + \sqrt{2\ln(1/\epsilon)}} \ge \frac{2\varrho_*(\delta, \epsilon)}{3\sqrt{2\ln(1/\epsilon)}},$$

with last inequality stemming from $\sqrt{2\ln(1/\epsilon)} \geq 2$ when $\epsilon \leq \frac{1}{8}$. Since we assume $\epsilon \geq \frac{1}{8}$,

$$\mathrm{RiskOpt}_{\epsilon}^{\|\cdot\|}[\mathcal{Z}(\delta)] \geq \mathrm{RiskOpt}_{\frac{1}{8}}^{\|\cdot\|}[\mathcal{Z}(\delta)],$$

which in turn implies the desired result. □

Proof of 2.3.2. Recall that both terms $\overline{\alpha}(\cdot)$ and $\overline{\beta}(\cdot)$ of the bound (2.61) of Proposition 2.6.1 are subadditive; we first treat the case of $\delta_1 \leq \delta$, such that for notation of the proposition, for $i < \bar{i}$ (recall that $\mathcal{Z}(\delta_i) \subset \mathcal{Z}(\delta) \subset \mathcal{Z}(\delta_{\bar{i}})$ in this case)

$$\begin{aligned} d_{i\bar{i}}(\epsilon) &= \overline{\alpha}\big((H_i - H_{\bar{i}})^T A, \mathcal{Z}(\delta_i)\big) + \overline{\beta}\big(H_i - H_{\bar{i}}, \epsilon\big) \\ &\leq \overline{\alpha}\big((H_i^T A - B), \mathcal{Z}(\delta_i)\big) + \overline{\alpha}\big((H_{\bar{i}}^T A - B), \mathcal{Z}(\delta_i)\big) + \overline{\beta}\big(H_i, \epsilon\big) + \overline{\beta}\big(H_{\bar{i}}, \epsilon\big) \\ &= \rho_i + \rho_{\bar{i}} \leq (2+\vartheta)\varrho_*(\delta, \epsilon) \end{aligned}$$

due to $\rho_{\bar{i}} = \varrho_*(\delta_{\bar{i}}, \epsilon) \leq (1+\vartheta)\varrho_*(\delta, \epsilon)$ and $\rho_i \leq \varrho_*(\delta, \epsilon)$;this implies (2.32). When grid-size $I+1$ is as prescribed in 2.3.2, 2.70 implies that

$$\frac{\varrho_*(\delta, \epsilon)}{\varrho_*(\delta, \varepsilon)} \leq \sqrt{\ln(e\varepsilon/\epsilon)} = \sqrt{\ln(e(1+I))}.$$

□

Proof of 2.3.4. We first prove the last inequality in (2.37), the latter being a consequence of 2.70 for the function $\overline{\mathfrak{p}}$. Upper bound (2.38) is a direct consequence of 2.70. □

2.5.3 Proofs of section 2.4

Proof of 2.4.1. Let the set Ξ be such that for all $\xi \in \Xi$ one has

$$\big\{|\widehat{\pi}_{ij}^{(\ell_*)} - \pi_{ij}| \geq \mathfrak{d}_{ij}^{(\ell)}(\epsilon) \quad \forall i \neq j \leq I,\, \ell \geq \ell_*\big\}.$$

By construction, we have $\mathrm{Prob}_\xi\{\Xi\} \geq 1 - \frac{1}{2}I^2(I-1)\epsilon = 1 - \varepsilon$ (recall that one has $\widehat{\pi}_{ij}^{(\ell_*)} = -\widehat{\pi}_{ji}^{(\ell_*)}$, so that there are only $\frac{1}{2}I(I-1)$ distinct estimates in the above set for each $\ell \leq I$). Now, let $\xi \in \Xi$ be fixed, and let $i_* = i_*(\omega)$ be the index of one of the $\|\cdot\|_2$-closest to w points among $w_0, ..., w_I$. First, observe that i_* is ℓ_*-admissible. Indeed, due to $z \in \mathcal{Z}(\delta_{\ell_*}) \subset \mathcal{Z}(\delta_\ell)$ for $\ell \geq \ell_*$, one has

$$|\widehat{\pi}_{i_*j}^{(\ell)} - \pi_{i_*j}| \leq \mathfrak{d}_{i_*j}^{(\ell)} \quad \forall j \neq i_*,\, \ell \geq \ell_*,$$

and because $\pi_{i_*j} \leq v_{i_*j}^T \overline{w}_{i_*j}$, we conclude that

$$\widehat{r}_{i_*j}^{(\ell)} = [\widehat{\pi}_{i_*j}^{(\ell)} - \pi_{i_*j}] + [\pi_{i_*j} - v_{i_*j}^T \overline{w}_{i_*j}] - \mathfrak{d}_{i_*j}^{(\ell)} \leq 0 \quad j \neq i_*,\, \ell \geq \ell_*.$$

On the other hand, assuming that i' is ℓ_*-admissible, we have $\widehat{r}_{i'i_*}^{(\ell_*)} \leq 0$, so that

$$\pi_{i'i_*} \leq \widehat{\pi}_{i'i_*}^{(\ell_*)} + \mathfrak{d}_{i'i_*}^{(\ell_*)} \leq v_{i'i_*}^T \overline{w}_{i'i_*} + 2\mathfrak{d}_{i'i_*}^{(\ell_*)}.$$

Let now $\gamma_{i_*i'}$ be the orthogonal projection of w onto the line passing through $w_{i'}$ and w_{i_*}, and let $\tau_* = \tau(w_{i_*})$, $\tau_\gamma = \tau(\gamma_{i_*i'})$, and $\tau' = \tau(w_{i'})$ be coordinates of w_{i_*}, $\pi_{i_*i'}$, and $w_{i'}$ on this line, the

origin on the line being the midpoint $\overline{w}_{i_*i'}$ of the segment $[w_{i_*}, w_{i'}]$, its orientation given by $v_{i'i_*}$. One has

$$\tau_* = \tfrac{1}{2}\|w_{i'} - w_{i_*}\|_2, \quad \tau' = -\tfrac{1}{2}\|w_{i'} - w_{i_*}\|_2, \quad \tau_\gamma = \frac{\pi_{i'i_*} - v_{i'i_*}^T \overline{w}_{i'i_*}}{\|v_{i'i_*}\|_2} \leq \frac{2\mathfrak{d}_{i'i_*}^{(\ell_*)}}{\|w_{i_*} - w_{i'}\|_2},$$

and so

$$\begin{aligned}
&\|w_* - w_{i'}\|_2^2 - \|w_* - w_{i_*}\|_2^2 = \|\gamma_{i_*i'} - w_{i'}\|_2^2 - \|\gamma_{i_*i'} - w_{i_*}\|_2^2 \\
&= (\tau_\gamma - \tau')^2 - (\tau_\gamma - \tau_*)^2 = 2(\tau_* - \tau')\tau_\gamma \leq 2\|w_{i_*} - w_{i'}\|_2 \frac{2\mathfrak{d}_{i'i_*}^{(\ell_*)}}{\|w_{i_*} - w_{i'}\|_2} \leq 4\mathfrak{d}_{i'i_*}^{(\ell_*)}
\end{aligned}$$

what implies the first statement of the proposition. Let now $\overline{\Xi} = \Xi \cup \Xi_{\ell_*}$ where

$$\Xi_\ell = \{\|w_\ell(\omega) - Bz\| \leq \rho_\ell(\epsilon)\}.$$

One clearly has $\mathrm{Prob}_\xi\{\overline{\Xi}\} \geq 1 - \varepsilon$, and for all $\xi \in \overline{\Xi}$,

$$\|w_{i_*} - w\|_2 \leq \|w_{\ell_*} - w\|_2 \leq \rho_{\ell_*}(\epsilon),$$

what implies (2.42). □

2.6 Deviations of norms of sub-Gaussian random vectors

2.6.1 Preliminaries

For the reader's convenience, we discuss in this section some essentially known bounds for deviations of quadratic forms of Gaussian and sub-Gaussian random vectors.

1°. Let ξ be a d-dimensional normal vector, $\xi \sim \mathcal{N}(\mu, \Sigma)$. For all $h \in \mathbf{R}^d$ and $G \in \mathbf{S}^d$ such that $G \prec \Sigma^{-1}$ we have the well known relationship:

$$\begin{aligned}
\ln\left(\mathbf{E}_\xi\left\{e^{h^T\xi + \frac{1}{2}\xi^T G\xi}\right\}\right) &= -\tfrac{1}{2}\ln\mathrm{Det}(I - \Sigma^{1/2}G\Sigma^{1/2}) \\
&\quad + h^T\mu + \tfrac{1}{2}\mu^T G\mu + \tfrac{1}{2}[G\mu + h]^T\Sigma^{1/2}(I - \Sigma^{1/2}G\Sigma^{1/2})^{-1}\Sigma^{1/2}[G\mu - h].
\end{aligned} \tag{2.47}$$

Now, suppose that $\eta \sim \mathcal{SG}(0, \Sigma)$ where $\Sigma \in \mathbf{S}_+^d$, let also $g \in \mathbf{R}^d$ and $S \in \mathbf{R}^{d\times d}$ such that $S\Sigma S^T \prec I$. Then for $\xi \sim \mathcal{N}(g, I)$ one has

$$\mathbf{E}_\eta\left\{e^{[Sg]^T\eta + \frac{1}{2}\eta^T S^T S\eta}\right\} = \mathbf{E}_\eta\left\{\mathbf{E}_\xi\left\{e^{[S\eta]^T\xi}\right\}\right\} = \mathbf{E}_\xi\left\{\mathbf{E}_\eta\left\{e^{[S\eta]^T\xi}\right\}\right\} \leq \mathbf{E}_\xi\left\{e^{\frac{1}{2}\xi^T S^T\Sigma S\xi}\right\},$$

so that

$$\begin{aligned}
\ln\left(\mathbf{E}_\eta\left\{e^{[Sg]^T\eta + \frac{1}{2}\eta^T S^T S\eta}\right\}\right) &\leq \ln\left(\mathbf{E}_\xi\left\{e^{\frac{1}{2}\xi^T S^T\Sigma S\xi}\right\}\right) \\
&= -\tfrac{1}{2}\ln\mathrm{Det}(I - S\Sigma S^T) + \tfrac{1}{2}g^T S\Sigma^{1/2}(I - S\Sigma S^T)^{-1}\Sigma^{1/2}S^T g.
\end{aligned}$$

Next, let $\zeta \sim \mathcal{SG}(\mu, \Sigma)$, and let $h \in \mathbf{R}^d$, $G \in \mathbf{S}_+^d$, $G \prec \Sigma^{-1}$. When setting $\eta \sim \mathcal{SG}(0, \Sigma)$, $S^T S = G$ and $S^T g = h + G\mu$ we get

$$\begin{aligned}
&\ln\left(\mathbf{E}_\zeta\left\{e^{h^T\zeta + \frac{1}{2}\zeta^T G\zeta}\right\}\right) = h^T\mu + \tfrac{1}{2}\mu^T G\mu + \ln\left(\mathbf{E}_\eta\left\{e^{h^T\zeta + \frac{1}{2}\zeta^T G\zeta}\right\}\right) \\
&= h^T\mu + \tfrac{1}{2}\mu^T G\mu - \tfrac{1}{2}\ln\mathrm{Det}(I - S\Sigma S^T) + \tfrac{1}{2}g^T S\Sigma^{1/2}(I - S\Sigma S^T)^{-1}\Sigma^{1/2}S^T g \\
&= h^T\mu + \tfrac{1}{2}\mu^T G\mu - \tfrac{1}{2}\ln\mathrm{Det}(I - \Sigma^{1/2}G\Sigma^{1/2}) + \tfrac{1}{2}[G\mu + h]^T\Sigma^{1/2}(I - \Sigma^{1/2}G\Sigma^{1/2})^{-1}\Sigma^{1/2}[G\mu + h].
\end{aligned}$$

2^o. In particular, when $\zeta \sim \mathcal{SG}(0, I)$, one has

$$\ln\left(\mathbf{E}_\zeta\left\{e^{h^T\zeta+\frac{1}{2}\zeta^T G\zeta}\right\}\right) \le -\tfrac{1}{2}\ln\mathrm{Det}(I-G) + \tfrac{1}{2}h^T(I-G)^{-1}h =: \Phi(h,G).$$

Observe that $\Phi(h,G)$ is convex and continuous in $h \in \mathbf{R}^d$ and $0 \preceq G \prec I$ on its domain. Using the inequality (cf. [65, Lemma 1])

$$\forall v \in [0,1[\quad -\ln(1-v) \le v + \frac{v^2}{2(1-v)}, \tag{2.48}$$

we arrive at

$$\Phi(h,G) \le \tfrac{1}{2}\mathrm{Tr}[G] + \tfrac{1}{4}\mathrm{Tr}[G(I-G)^{-1}G] + \tfrac{1}{2}h^T(I-G)^{-1}h =: \widetilde{\Phi}(h,G).$$

Finally, using

$$\mathrm{Tr}[G(I-G)^{-1}G] \le (1-\lambda_{\max}(G))^{-1}\mathrm{Tr}[G^2], \quad h^T(I-G)^{-1}h \le (1-\lambda_{\max}(G))^{-1}h^Th,$$

we arrive at

$$\widetilde{\Phi}(h,G) \le \tfrac{1}{2}\mathrm{Tr}[G] + \tfrac{1}{4}(1-\lambda_{\max}(G))^{-1}(\mathrm{Tr}[G^2] + 2\|h\|_2^2) =: \overline{\Phi}(h,G).$$

3^o. In the above setting, let $Q \in \mathbf{S}^d_+$, $\alpha > 2\lambda_{\max}(Q)$, $G = 2Q/\alpha$, and let $h = 0$. By the Cramer argument we conclude that

$$\mathrm{Prob}\left\{\zeta^TQ\zeta \ge \alpha[\Phi(2Q/\alpha) + \ln\epsilon^{-1}]\right\} \le \epsilon \tag{2.49}$$

where $\Phi(\cdot) = \Phi(0,\cdot)$. In particular,

$$\mathrm{Prob}\left\{\zeta^TQ\zeta \ge \alpha[\overline{\Phi}(2Q/\alpha) + \ln\epsilon^{-1}]\right\} \le \epsilon,$$

so, when choosing $\alpha = 2\lambda_{\max}(Q) + \sqrt{\frac{\mathrm{Tr}(Q^2)}{\ln\epsilon^{-1}}}$ we arrive at the "standard bound"

$$\mathrm{Prob}\left\{\zeta^TQ\zeta \ge \mathrm{Tr}(Q) + 2\|Q\|_{\mathrm{Fro}}\sqrt{\ln\epsilon^{-1}} + 2\lambda_{\max}(Q)\ln\epsilon^{-1}\right\} \le \epsilon.$$

4^o. Let now $H \in \mathbf{R}^{m\times n}$ and $z \in \mathbf{R}^n$ with $h = 2Hz$ and $G = 2HH^T$. Then for $\zeta \sim \mathcal{SG}(0,I)$ and $\alpha > 2\lambda_{\max}(H^TH)$ one has

$$\begin{aligned}\alpha\ln\left(\mathbf{E}_\zeta\left\{e^{\frac{1}{\alpha}\|z+H^T\zeta\|_2^2}\right\}\right) &= z^Tz + \alpha\ln\left(\mathbf{E}_\zeta\left\{e^{\alpha^{-1}[h^T\zeta+\frac{1}{2}\zeta^TG\zeta]}\right\}\right) \le z^Tz + \alpha\Phi\left(\frac{h}{\alpha},\frac{G}{\alpha}\right)\\ &= z^T(I-2\alpha^{-1}HH^T)^{-1}z - \frac{\alpha}{2}\ln\mathrm{Det}(I-2\alpha^{-1}HH^T) =: \Psi(\alpha;z,H).\end{aligned} \tag{2.50}$$

Being the perspective transformation of Φ, Ψ, by construction, is convex and continuous in α, z and $G = HH^T$ on its domain. As a consequence, for all $\epsilon \in (0,1)$ and all $\alpha > 2\lambda_{\max}(HH^T)$,

$$\mathrm{Prob}\left\{\|z+H^T\eta\|_2^2 \ge \Psi(\alpha;z,H) + \alpha\ln[\epsilon^{-1}]\right\} \le \epsilon.$$

When utilizing (2.48) we arrive at

$$\begin{array}{rcl}\Psi(\alpha;z,H) & \leq & \overline{\Psi}(\alpha;z,H)\\ & := & z^T(I-2\alpha^{-1}HH^T)^{-1}z+\mathrm{Tr}[HH^T]+\alpha^{-1}\mathrm{Tr}[HH^T(I-2\alpha^{-1}HH^T)^{-1}HH^T]\end{array} \tag{2.51}$$

Similarly to $\Psi(\cdot)$, $\overline{\Psi}$ is convex, continuous and positive homogeneous in $G=HH^T$, z and α on its domain. Same as above, we may bound $\overline{\Psi}(\alpha;z,H)$ with

$$\widetilde{\Psi}(\alpha;z,H)=\mathrm{Tr}(HH^T)+\frac{\alpha\|z\|_2^2+\|HH^T\|_{\mathrm{Fro}}^2}{\alpha-2\lambda_{\max}(HH^T)},$$

so that

$$\mathrm{Prob}\left\{\|z+H^T\eta\|_2^2\geq\widetilde{\Psi}(\alpha;z,H)+\alpha\ln\epsilon^{-1}\right\}\leq\epsilon.$$

The choice of

$$\alpha=2\lambda_{\max}(HH^T)+\frac{\left(\|HH^T\|_{\mathrm{Fro}}^2+2\|z\|_2^2\lambda_{\max}(HH^T\right)^{1/2}}{\ln^{1/2}[\epsilon^{-1}]}$$

in the latter bound leads results in

$$\mathrm{Prob}\left\{\|z+H^T\eta\|_2^2\geq\widetilde{\Psi}_\epsilon(z,H)\right\}\leq\epsilon.$$

where

$$\begin{array}{rl}\widetilde{\Psi}_\epsilon(z,H):=&z^Tz+\mathrm{Tr}(HH^T)\\ &+2\Big([\|HH^T\|_{\mathrm{Fro}}^2+2\|z\|_2^2\lambda_{\max}(HH^T)]\ln[\epsilon^{-1}]\Big)^{1/2}+2\lambda_{\max}(HH^T)\ln[\epsilon^{-1}].\end{array} \tag{2.52}$$

2.6.2 Bounding the maximal probability of norm deviations

In this section, we assume to be given matrices $V\in\mathbf{R}^{\nu\times n}$, $W\in\mathbf{R}^{m\times\nu}$, a basic ellitope $\mathcal{Z}$, [8]

$$\mathcal{Z}=\{z\in\mathbf{R}^n:\exists q\in\mathcal{Q}:\,z^TQ_kz\leq q_k,\,k\leq\mathfrak{n}_z\}, \tag{2.53}$$

and a co-ellitopic norm $\|\cdot\|$ with the unit ball $\mathcal{B}_*$ (2.21) of the conjugate norm $\|\cdot\|_*$. Our objective is to derive for all $z\in\mathcal{Z}$ a "reasonably tight" high-probability upper bound for the quantity

$$\|Vz+W^T\xi\|,\quad\xi\sim\mathcal{SG}(0,I).$$

1^o. Our first observation is that

$$\|Vz+W^T\xi\|\leq\|Vz\|+\|W^T\xi\|=\underbrace{\max_{v\in\mathcal{B}_*}v^TVz}_{\alpha(V,z)}+\underbrace{\max_{w\in\mathcal{B}_*}w^TW^T\xi}_{\beta(W,\xi)}. \tag{2.54}$$

One has

$$\forall z\in\mathcal{Z}\quad\alpha(V,z)\leq\max_{y\in\mathcal{Y}}\sup_{z\in\mathcal{Z}}y^TP^TVz\leq\tfrac{1}{2}\max_{[z;y]\in\mathcal{Z}\times\mathcal{Y}}[z;y]^T\left[\begin{array}{c|c}&V^TM\\\hline P^TV&\end{array}\right][z;y] \tag{2.55}$$

[8] We assume that the data Q_k and $\mathcal{Q}$ of $\mathcal{Z}$ satisfy the standard requirements of the definitions in Section 2.3.1.

Observe that the set $\mathcal{G} := Z \times \mathcal{Y}$ is the direct product of basic ellitopes and is itself a basic ellitope,

$$\mathcal{G} = \left\{[z; y] \in \mathbf{R}^n \times \mathbf{R}^N : \exists r = [q, s] \in \mathcal{Q} \times \mathcal{M} : z^T Q_j z \le q_j,\ j \le \mathfrak{n}_z,\ y^T M_\ell y \le s_\ell,\ \ell \le \mathfrak{n}_y\right\}$$

In other words, the r.-h.s. of (2.55) is the maximum of a quadratic form over the basic ellitope $\mathcal{G}$; when applying Theorem 2.3.1 we arrive at the following bound for $\alpha(V, z)$:

$$\forall z \in \mathcal{Z} \quad \alpha(V, z) \le \overline{\alpha}(V, \mathcal{Z}) := \min_{\lambda, \chi} \left\{ \phi_{\mathcal{Q}}(\lambda) + \phi_{\mathcal{M}}(\chi) :\ \lambda \ge 0, \chi \ge 0, \left[\begin{array}{c|c} \sum_j \lambda_j Q_j & \frac{1}{2} V^T M \\ \hline \frac{1}{2} P^T V & M(\chi) \end{array}\right] \succeq 0 \right\}. \tag{2.56}$$

Observe first that problem (2.56) is jointly convex in λ, χ and V, and its objective is continuous and positive homogeneous of degree 1.[9] We conclude that (2.56) is solvable and its objective is convex in V.

2^o. To bound $\beta(W, \xi)$ we act as follows (cf. [30, Section 3.2]). Let $\Theta \in \mathbf{S}^\nu_+$ satisfy the constraint

$$\left[\begin{array}{c|c} \Theta & \frac{1}{2} W M \\ \hline \frac{1}{2} P^T W^T & M(\kappa) \end{array}\right] \succeq 0. \tag{2.57}$$

By (2.57),

$$\beta(W, \xi) \le \max_{y \in \mathcal{Y}} \left\{ \xi^T \Theta \xi + y^T M(\kappa) y \right\} \le \xi^T \Theta \xi + \phi_{\mathcal{M}}(\kappa).$$

When using the classical concentration inequality (2.49) for the homogeneous quadratic bound form $\xi^T \Theta \xi$ of the sub-Gaussian vector ξ we obtain, for any $\epsilon \in (0, 1)$,

$$\mathrm{Prob}_\xi \left\{ \xi^T \Theta \xi \le \psi_\epsilon(\Theta) \right\} \ge 1 - \epsilon \tag{2.58}$$

where

$$\psi_\epsilon(\Theta) = \min_\alpha \left\{ \psi_\epsilon(\alpha, \Theta) := -\frac{\alpha}{2} \log \mathrm{Det}(I - 2\alpha^{-1}\Theta) + \alpha \ln[\epsilon^{-1}] : 2\lambda_{\max}(\Theta) \le \alpha \right\}. \tag{2.59}$$

Thus,

$$\beta(W, \xi) \le \overline{\beta}(W, \epsilon) := \min_{\Theta, \kappa, \alpha} \left\{ \psi_\epsilon(\alpha, \Theta) + \phi_{\mathcal{M}}(\kappa) :\ \kappa \ge 0,\ 2\lambda_{\max}(\Theta) \le \alpha, \left[\begin{array}{c|c} \Theta & \frac{1}{2} W M \\ \hline \frac{1}{2} P^T W^T & M(\kappa) \end{array}\right] \succeq 0 \right\} \tag{2.60}$$

with probability $1 - \epsilon$. The optimization problem in the r.-h.s. of (2.60) is jointly convex in Θ, κ, α and W, its objective is continuous and positive homogeneous of degree 1, so (2.60) is solvable.

[9] One can easily see that when multiplying V, λ and χ by $\gamma > 0$ the objective of (2.59) is also multiplied by γ due to the positive homogeneity of $\phi_{\mathcal{Q}}$.

Remark. The term $\psi_\epsilon(\Theta)$ in the bound (2.58) and $\psi_\epsilon(\alpha,\Theta))$ in (2.60) may be replaced by their standard upper approximations; for instance, one may use

$$\begin{aligned}\psi_\epsilon(\Theta) &\le \widetilde{\psi}_\epsilon(\Theta) := \min_{\alpha}\Big\{\underbrace{\mathrm{Tr}(\Theta)+\mathrm{Tr}\left(\Theta(\alpha I-2\Theta)^{-1}\Theta\right)+\alpha\ln(\epsilon^{-1}):\ 2\lambda_{\max}(\Theta)\le\alpha}_{=:\widetilde{\psi}_\epsilon(\alpha,\Theta))}\Big\}\\ &\le \overline{\psi}_\epsilon(\Theta) := \mathrm{Tr}(\Theta)+2\|\Theta\|_{\mathrm{Fro}}\sqrt{\ln\epsilon^{-1}}+2\lambda_{\max}(\Theta)\ln\epsilon^{-1}.\end{aligned}$$

Although less precise, such approximations lead to simpler optimization problems to be solved to compute the resulting bound.

When substituting the bounds for $\alpha(V,z)$ and $\beta(W,\xi)$ into (2.54) we arrive at the following

Proposition 2.6.1 *In the setting of this section, let $\overline{\alpha}(V,\mathcal{Z})$ and $\overline{\beta}(W,\epsilon)$ be as in (2.56) and (2.60). Then*

$$\|Vz+W^T\xi\|\le\overline{\alpha}(V,\mathcal{Z})+\overline{\beta}(W,\epsilon) \tag{2.61}$$

with probability at least $1-\epsilon$. The r.-h.s. of (2.61) is a convex function of variables V and W.

Furthermore, the terms of the bound (2.61) satisfy the triangular inequality: for all $V_1,V_2\in\mathbf{R}^{\nu\times n}$ and $W_1,W_2\in\mathbf{R}^{\nu\times m}$

$$\begin{aligned}\overline{\alpha}(V_1+V_2,\mathcal{Z})&\le\overline{\alpha}(V_1,\mathcal{Z})+\overline{\alpha}(V_2,\mathcal{Z}),\\ \overline{\beta}(W_1+W_2,\epsilon)&\le\overline{\beta}(W_1,\epsilon)+\overline{\beta}(W_2,\epsilon).\end{aligned}$$

Indeed, by construction function $\overline{\alpha}(\cdot,\mathcal{Z})$ is positive homogeneous and convex in its argument. Thus,

$$\tfrac12\overline{\alpha}(V_1+V_2,\mathcal{Z})=\overline{\alpha}\big(\tfrac12[V_1+V_2],\mathcal{Z}\big)\le\tfrac12\overline{\alpha}(V_1,\mathcal{Z})+\tfrac12\overline{\alpha}(V_2,\mathcal{Z}).$$

Similar derivation implies that $\frac12\overline{\beta}(W_1+W_2,\epsilon)\le\frac12\overline{\beta}(W_1,\epsilon)+\overline{\beta}(W_2,\epsilon)$. □

We also have the following variant of Proposition 2.6.1 in the case of Euclidean norm $\|\cdot\|=\|\cdot\|_2$. Denote

$$\begin{aligned}\overline{\alpha}_2(V,W;\mathcal{Z},\epsilon)=\min_{\lambda,\alpha,\Theta}\Big\{&\phi_{\mathcal{Q}}(\lambda)-\tfrac{\alpha}{2}\ln\mathrm{Det}(I-2\alpha^{-1}\Theta)+\alpha\ln[\epsilon^{-1}]:\ \lambda\ge0,\mu\ge0,\alpha\ge2\lambda_{\max}(\Theta),\\ &\left[\begin{array}{c|c}\Theta&W\\\hline W^T&I\end{array}\right]\succeq0,\ \left[\begin{array}{c|c}\sum_k\lambda_kQ_k&V^T\\\hline V&I-2\alpha^{-1}\Theta\end{array}\right]\succeq0\Big\}.\end{aligned} \tag{2.62}$$

and

$$\begin{aligned}\overline{\beta}_2(V,W;\mathcal{Z},\epsilon)=\min_{\lambda,\alpha,\Theta,\Lambda,\Upsilon}\Big\{&\phi_{\mathcal{Q}}(\lambda)+\mathrm{Tr}[\Theta+\Lambda]+\alpha\ln[\epsilon^{-1}]:\ \lambda\ge0,\mu\ge0,\alpha\ge2\lambda_{\max}(\Theta),\\ &\left[\begin{array}{c|c}\Theta&W\\\hline W^T&I\end{array}\right]\succeq0,\ \left[\begin{array}{c|c}\Lambda&\Theta\\\hline\Theta&\alpha I-2\Theta\end{array}\right]\succeq0,\ \left[\begin{array}{c|c}\sum_k\lambda_kQ_k&V^T\\\hline V&I-2\alpha^{-1}\Theta\end{array}\right]\succeq0\Big\}.\end{aligned} \tag{2.63}$$

Proposition 2.6.2 *Given $\epsilon\in(0,1)$, let $V\in\mathbf{R}^{\nu\times N}$, $W\in m\times\nu$, let $\mathcal{Z}$ be defined in (2.53), and let $\overline{\alpha}_2(V,W;Z,\epsilon)$ and $\overline{\beta}_2(V,W;\mathcal{Z},\epsilon)$ be as in (2.62) and (2.63). Then for all $z\in\mathcal{Z}$*

$$\|Vz+W^T\xi\|_2^2\le\overline{\alpha}_2(V,W;\mathcal{Z},\epsilon)\le\overline{\beta}_2(V,W;\mathcal{Z},\epsilon) \tag{2.64}$$

with probability at least $1-\epsilon$.

Proof. Let

$$\psi_{2,\epsilon}(V,W,z) := \min_{\alpha>2\lambda_{\max}(WW^T)} z^TV^T(I-2\alpha^{-1}WW^T)^{-1}Vz - \frac{\alpha}{2}\ln\mathrm{Det}(I-2\alpha^{-1}WW^T) + \alpha\ln[\epsilon^{-1}]$$
$$= \min_{\alpha,\Theta,\Upsilon}\Big\{z^T\Upsilon z - \frac{\alpha}{2}\ln\mathrm{Det}(I-2\alpha^{-1}\Theta) + \alpha\ln[\epsilon^{-1}] : \alpha > 2\lambda_{\max}(WW^T)$$
$$\Theta \succeq WW^T,\ \Upsilon \succeq V^T(I-2\alpha^{-1}\Theta)^{-1}V\Big\}; \qquad (2.65)$$

by (2.50) one has

$$\mathrm{Prob}_\xi\{\|Vz+W^T\xi\|_2^2 \le \psi_{2,\epsilon}(V,W,z)\} \ge 1-\epsilon.$$

Using the result of Theorem 2.3.1 we may bound the first term in the r.-h.s. of (2.65) uniformly over $z \in \mathcal{Z}$:

$$\max_{z\in\mathcal{Z}} z^T\Upsilon z \le \phi_{\mathcal{Q}}(\lambda)$$

where $\lambda \ge 0$ is such that $\sum_k \lambda_k Q_k \succeq \Upsilon$. When summing up, we arrive at the bound

$$\forall z \in \mathcal{Z}) \quad \mathrm{Prob}_\xi\{\|V^Tz+W^T\xi\|_2^2 \le \overline{\alpha}_2(V,W;\mathcal{Z},\epsilon)\} \ge 1-\epsilon \qquad (2.66)$$

where $\overline{\alpha}_2(V,W;\mathcal{Z},\epsilon)$ is defined in (2.43).

Furthermore, we may upper bound the quantity $\overline{\alpha}_2(V,W;\mathcal{Z},\epsilon)$ using the inequality (2.51). Indeed, one has

$$\psi_{2,\epsilon}(V,W,z) \le \min_\alpha\Big\{z^TV^T(I-2\alpha^{-1}WW^T)^{-1}Vz + \mathrm{Tr}[WW^T]$$
$$+\alpha^{-1}\mathrm{Tr}[WW^T(I-2\alpha^{-1}WW^T)^{-1}WW^T] + \alpha\ln[\epsilon^{-1}] : \alpha \ge 2\lambda_{\max}(\Theta)\Big\}$$
$$= \min_{\alpha,\Theta,\Lambda,\Upsilon}\Big\{z^T\Upsilon z + \mathrm{Tr}[\Theta+\Lambda] + \alpha\ln[\epsilon^{-1}] :\ \alpha \ge 2\lambda_{\max}(\Theta)$$
$$\Theta \succeq WW^T,\ \Upsilon \succeq V^T(I-2\alpha^{-1}\Theta)^{-1}V,\ \Lambda \succeq \Theta(\alpha I-2\Theta)^{-1}\Theta\Big\},$$

and we conclude that $\overline{\alpha}_2(V,W;\mathcal{Z},\epsilon) \le \overline{\beta}_2(V,W;\mathcal{Z},\epsilon)$ with $\overline{\beta}_2(V,W;\mathcal{Z},\epsilon)$ as in (2.63). □

2.6.3 Bounding maximal over ellitope deviation of a quadratic form of Gaussian vector

Lemma 1 *Let $D \in \mathbf{R}^{d\times k}$, $C \in \mathbf{S}^k$, $F \in \mathbf{S}^d$, $\zeta \in \mathbf{R}^d$ be standard Gaussian, i.e., $\zeta \sim \mathcal{N}(0,I)$, and let $\mathcal{X} \subset \mathbf{R}^k$ be a basic ellitope,*

$$\mathcal{X} = \{x \in \mathbf{R}^k : t \in \mathcal{T} :\ z^TT_\ell z \le t_\ell,\ \ell \le \mathfrak{n}_x\}$$

with $T_k, k = 1,...,\mathfrak{n}_x$ and $\mathcal{T}$ satisfying the standard requirements. Let also

$$\Psi_+(C,D,F;\epsilon) = \min_{\alpha,\lambda}\Big\{-\tfrac{\alpha}{2}\ln\mathrm{Det}(I-2F/\alpha) + \tfrac{1}{2}\phi_{\mathcal{T}}(\lambda) + \alpha\ln(2/\epsilon) - \mathrm{Tr}[F],$$
$$\alpha I \succeq 2F,\ \lambda \ge 0,\ \left[\begin{array}{c|c}\sum_k\lambda_kT_k-2C & D^T\\ \hline D & \alpha I-2F\end{array}\right] \succeq 0\Big\},$$
$$\Psi_-(C,D,F;\epsilon) \le \min_{\alpha,\lambda}\Big\{-\tfrac{\alpha}{2}\ln\mathrm{Det}(I+2F/\alpha) + \tfrac{1}{2}\phi_{\mathcal{T}}(\lambda) + \alpha\ln(2/\epsilon) + \mathrm{Tr}[F],$$
$$\alpha I \succeq -2F,\ \lambda \ge 0,\ \left[\begin{array}{c|c}\sum_k\lambda_kT_k+2C & D^T\\ \hline D & \alpha I+2F\end{array}\right] \succeq 0\Big\}.$$

Then for all $x \in \mathcal{X}$,

$$\mathrm{Prob}_\zeta\left\{\left|x^TCx + \zeta^TDx + \zeta^TF\zeta - \mathrm{Tr}(F) + \varkappa\right| \le \rho\right\} \le \epsilon$$

where $[C, D, F, \rho, \varkappa]$ is a feasible solution to the system of convex constraints

$$\Psi_+(C, D, F; \epsilon) \le \rho - \varkappa, \quad \Psi_-(C, D, F; \epsilon) \le \rho + \varkappa. \tag{2.67}$$

Furthermore, one has

$$\Psi_+(C, D, F; \epsilon) \le \overline{\Psi}_+(C, D, F; \epsilon), \quad \Psi_-(C, D, F; \epsilon) \le \overline{\Psi}_-(C, D, F; \epsilon)$$

where

$$\begin{aligned}
\overline{\Psi}_+(C, D, F; \epsilon) := \min_{\alpha,\lambda}\Big\{&\alpha^{-1}\mathrm{Tr}(F^2) + \tfrac{1}{2}\phi_{\mathcal{T}}(\lambda) + \alpha\ln(2/\epsilon),\\
&\alpha I \succeq 2.930F,\ \lambda \ge 0,\ \left[\begin{array}{c|c}\sum_k \lambda_k T_k - 2C & D^T\\ \hline D & \alpha I - 2F\end{array}\right] \succeq 0\Big\},\\
\overline{\Psi}_-(C, D, F; \epsilon) := \min_{\alpha,\lambda}\Big\{&\alpha^{-1}\mathrm{Tr}(F^2) + \tfrac{1}{2}\phi_{\mathcal{T}}(\lambda) + \alpha\ln(2/\epsilon),\\
&\alpha I \succeq -2.930F,\ \lambda \ge 0,\ \left[\begin{array}{c|c}\sum_k \lambda_k T_k + 2C & D^T\\ \hline D & \alpha I + 2F\end{array}\right] \succeq 0\Big\}.
\end{aligned}$$

Consequently, when $\overline{\rho}$ and $\overline{\varkappa}$ are such that

$$\overline{\Psi}_+(C, D, F; \epsilon) \le \overline{\rho} - \overline{\varkappa}, \quad \overline{\Psi}_-(C, D, F; \epsilon) \le \overline{\rho} + \overline{\varkappa},$$

one has

$$\mathrm{Prob}_\zeta\left\{\left|x^TCx + \zeta^TDx + \zeta^TF\zeta - \mathrm{Tr}(F) + \overline{\varkappa}\right| \le \overline{\rho}\right\} \le \epsilon \quad \forall x \in \mathcal{X}.$$

Proof. Using (2.47) one has for $h \in \mathbf{R}^d$, $F \in \mathbf{S}^d$, and α such that $2F \prec \alpha I$,

$$\alpha\ln\mathbf{E}_\zeta\left\{e^{\alpha^{-1}\left(h^T\zeta + \zeta^TF\zeta\right)}\right\} \le -\tfrac{\alpha}{2}\ln\mathrm{Det}(I - 2F/\alpha) + \tfrac{1}{2}h^T(\alpha I - 2F)^{-1}h,$$

so that $\mathrm{Prob}_\zeta\left\{h^T\zeta + \zeta^TF\zeta - \mathrm{Tr}[F] \ge \psi_+(h, F; \epsilon)\right\} \le \epsilon/2$ where

$$\begin{aligned}
\psi_+(h, F; \epsilon) &:= \inf_{\alpha:\,\alpha I \succ F} \alpha\ln\mathbf{E}_\zeta\left\{e^{\alpha^{-1}\left(h^T\zeta + \zeta^TF\zeta - \mathrm{Tr}[F]\right)}\right\} + \alpha\ln(2/\epsilon) \qquad (2.68)\\
&= \inf_\alpha\left\{-\tfrac{\alpha}{2}\ln\mathrm{Det}(I - 2F/\alpha) + \tfrac{1}{2}h^T(\alpha I - 2F)^{-1}h + \alpha\ln(2/\epsilon) - \mathrm{Tr}[F],\ \alpha I \succ 2F\right\}.
\end{aligned}$$

A completely analogous derivation results in $\mathrm{Prob}_\zeta\left\{h^T\zeta + \zeta^TF\zeta - \mathrm{Tr}[F] \ge -\psi_-(h, F; \epsilon)\right\} \le \epsilon/2$ where

$$\psi_-(h, F; \epsilon) = \inf_\alpha\left\{-\tfrac{\alpha}{2}\ln\mathrm{Det}(I + 2F/\alpha) + \tfrac{1}{2}h^T(\alpha I + 2F)^{-1}h + \alpha\ln(2/\epsilon) + \mathrm{Tr}[F],\ -\alpha I \prec 2F\right\} \tag{2.69}$$

When setting $h = Dx$ we arrive at

$$\begin{aligned}
&\sup_{x\in\mathcal{X}}[x^TCx + \psi_+(Dx, F; \epsilon)]\\
&\le \inf_\alpha \sup_{z\in\mathcal{Z}}\left\{-\tfrac{\alpha}{2}\ln\mathrm{Det}(I - 2F/\alpha) + \tfrac{1}{2}z^T\left[2C + D^T(\alpha I - 2F)^{-1}D\right]z + \alpha\ln(2/\epsilon) - \mathrm{Tr}[F],\ \alpha I \succ 2F\right\}\\
&\le \min_{\alpha,\lambda}\Big\{-\tfrac{\alpha}{2}\ln\mathrm{Det}(I - 2F/\alpha) + \tfrac{1}{2}\phi_{\mathcal{T}}(\lambda) + \alpha\ln(2/\epsilon) - \mathrm{Tr}[F],\\
&\qquad \alpha I \succ 2F,\ \lambda \ge 0,\ \sum_k \lambda_k T_k \succeq 2C + D^T(\alpha I - 2F)^{-1}D \succeq 0\Big\}\\
&= \Psi_+(C, D, F; \epsilon).
\end{aligned}$$

Similarly,

$$\begin{aligned}&\sup_{z\in\mathcal{Z}}[-z^TCz+\psi_-(Dz,F;\epsilon)]\\&\leq\inf_{\alpha}\sup_{z\in\mathcal{Z}}\left\{-\tfrac{\alpha}{2}\ln\mathrm{Det}(I+2F/\alpha)+\tfrac{1}{2}z^T\left[-2C+D^T(\alpha I+2F)^{-1}D\right]z+\alpha\ln(2/\epsilon)-\mathrm{Tr}[F],\ \alpha I\succ 2F\right\}\\&\leq\Psi_-(C,D,F;\epsilon)..\end{aligned}$$

We conclude that if ρ and $\varkappa$ satisfy the system (2.67) then

$$\mathrm{Prob}_\zeta\left\{\left|x^TCx+\zeta^TDx+\zeta^TF\zeta-\mathrm{Tr}(F)+\varkappa\right|\leq\rho\right\}\leq\epsilon\quad\forall x\in\mathcal{X}.$$

Furthermore, using the inequality $|-\ln(1-v)-v|\leq v^2$ for $v\leq 0.683$ we may upper bound the quantities ψ_+ and ψ_- in (2.68) and (2.69) as follows:

$$\begin{aligned}\psi_+(h,F;\epsilon)\leq\overline{\psi}_+(h,F;\epsilon):=\inf_\alpha\left\{\alpha^{-1}\mathrm{Tr}(F^2)+\tfrac{1}{2}h^T(\alpha I-2F)^{-1}h+\alpha\ln[2/\epsilon],\ \alpha I\succeq 2.930F\right\},\\\psi_-(h,F;\epsilon)\leq\overline{\psi}_-(h,F;\epsilon):=\inf_\alpha\left\{\alpha^{-1}\mathrm{Tr}(F^2)+\tfrac{1}{2}h^T(\alpha I+2F)^{-1}h+\alpha\ln[2/\epsilon],\ -\alpha I\preceq 2.930F\right\}.\end{aligned}$$

As a result,

$$\Psi_+(C,D,F;\epsilon)\leq\overline{\Psi}_+(C,D,F;\epsilon)\ \text{ and }\ \Psi_-(C,D,F;\epsilon)\leq\overline{\Psi}_-(C,D,F;\epsilon)$$

what implies the last statement of the lemma. □

2.6.4 Variations of risk upper bounds in δ and ε

Proposition 2.6.3 *Let $\delta>0$, $\vartheta>0$, $0<\epsilon<\varepsilon<1$, and $\mathfrak{r}_{\delta,\varepsilon}(\cdot),\mathfrak{r}^{(2)}_{\delta,\varepsilon}(\cdot),\overline{\mathfrak{p}}_{\delta,\varepsilon}(\cdot)$ defined respectively as in 2.26, 2.43 and 2.34.*
For any $H\in\mathbf{R}^{m\times\nu}$ and any $\Theta\in\mathbf{S}_m$,

$$\frac{\mathfrak{r}_{\delta,\epsilon}(H)}{\mathfrak{r}_{\delta,\varepsilon}(H)}\leq\sqrt{\ln(e\varepsilon/\epsilon)},\quad\frac{\mathfrak{r}^{(2)}_{\delta,\epsilon}(H)}{\mathfrak{r}^{(2)}_{\delta,\varepsilon}(H)}\leq\ln(e\varepsilon/\epsilon),\quad\frac{\overline{\mathfrak{p}}_{\delta,\epsilon}(\Theta)}{\overline{\mathfrak{p}}_{\delta,\varepsilon}(\Theta)}\leq\sqrt{\ln(e\varepsilon/\epsilon)}.\tag{2.70}$$

Additionnaly, when $\mathcal{U}(\delta):=\mathfrak{f}(\delta)\mathcal{U}$, with

$$f:\mathbf{R}_+\to\mathbf{R}_+$$

increasing in δ and $\mathcal{U}$ an ellitope, we define $\mathfrak{b}(\vartheta):=\max_{\delta>0}\left\{\frac{\mathfrak{f}((1+\vartheta)\delta)}{\mathfrak{f}(\delta)}\right\}$ and

$$\mathcal{C}_2(\vartheta,\varepsilon/\epsilon):=\max\left\{\mathfrak{b}(\vartheta);\sqrt{\ln(e\varepsilon/\epsilon)}\right\}.$$

One has

$$\frac{\mathfrak{r}_{(1+\vartheta)\delta,\epsilon}(H)}{\mathfrak{r}_{\delta,\varepsilon}(H)}\leq\mathcal{C}_2(\vartheta,\varepsilon/\epsilon),\quad\frac{\mathfrak{r}^{(2)}_{(1+\vartheta)\delta,\epsilon}(H)}{\mathfrak{r}^{(2)}_{\delta,\varepsilon}(H)}\leq\mathcal{C}_2(\vartheta,\varepsilon/\epsilon)^2,\quad\frac{\overline{\mathfrak{p}}_{(1+\vartheta)\delta,\epsilon}(\Theta)}{\overline{\mathfrak{p}}_{\delta,\varepsilon}(\Theta)}\leq\mathcal{C}_2(\vartheta,\varepsilon/\epsilon).\tag{2.71}$$

Proof. We first give a detailed proof of results 2.70 and 2.71 with respect to function $\mathfrak{r}_{\delta,\varepsilon}(\cdot)$.

1^o. Given a matrix $H \in \mathbf{R}^{m\times\nu}$, define variables $[\lambda_*, \mu_*, \chi_*, \kappa_*, \alpha_*, \Theta_*]$ that are optimal for problem 2.26 when computing $\mathfrak{r}_{\delta,\varepsilon}(H)$. We first proove that the latter is of the form

$$\mathfrak{r}_{\delta,\varepsilon}(H) = 2\left(\sqrt{\phi_{\mathcal{M}}(\chi_*)(\phi_{\mathcal{T}}(\lambda_*) + \phi_{\mathcal{V}(\delta)}(\mu_*))} + \sqrt{\phi_{\mathcal{M}}(\kappa_*)\psi_\varepsilon(\alpha_*, \Theta_*)}\right). \tag{2.72}$$

Indeed, for variables $[\lambda, \mu, \chi, \kappa, \alpha, \Theta]$ feasible for computation of $\mathfrak{r}_{\delta,\epsilon}(H)$ (note that for any other δ', ϵ', these are also feasible for computation of $\mathfrak{r}_{\delta',\epsilon'}(H)$) and positive γ, γ', variables $[\gamma\lambda, \gamma\mu, \chi/\gamma, \kappa/\gamma', \gamma'\alpha, \gamma'\Theta]$ are also feasible. 1-homogeneity of functions $\phi_{\mathcal{T}}, \phi_{\mathcal{V}(\delta)}, \phi_{\mathcal{M}}$ then leads to

$$\begin{array}{rll}\min_{\gamma,\gamma'} & \left\{r_{\gamma,\gamma'}(\lambda,\mu,\chi,\kappa,\alpha,\Theta) := & \gamma(\phi_{\mathcal{T}}(\lambda) + \phi_{\mathcal{V}(\delta)}(\mu)) + \phi_{\mathcal{M}}(\chi)/\gamma + \phi_{\mathcal{M}}(\kappa)/\gamma' + \gamma'\psi_\epsilon(\alpha,\Theta)\right\} \\ & = & 2\sqrt{\phi_{\mathcal{M}}(\chi)(\phi_{\mathcal{T}}(\lambda) + \phi_{\mathcal{V}}(\mu)} + 2\sqrt{\phi_{\mathcal{M}}(\kappa)\psi_\epsilon(\alpha,\Theta)}.\end{array}$$

We conclude using the fact that $\mathfrak{r}_{\delta,\varepsilon}(H) = \min_{\gamma,\gamma'} \quad r_{\gamma,\gamma'}(\lambda_*, \mu_*, \chi_*, \kappa_*, \alpha_*, \Theta_*)$.

2^o. Consider variables $[\lambda_*, \mu_*, \chi_*, \kappa_*, \alpha_*, \Theta_*]$ for computation of $\mathfrak{r}_{\delta,\epsilon}(H)$. Their sub-optimality for this problem entails

$$\frac{\mathfrak{r}_{\delta,\epsilon}(H)}{\mathfrak{r}_{\delta,\varepsilon}(H)} \le \frac{\left(\phi_{\mathcal{M}}(\chi_*)(\phi_{\mathcal{T}}(\lambda_*)+\phi_{\mathcal{V}(\delta)}(\mu_*))\right)^{1/2}+(\phi_{\mathcal{M}}(\kappa_*)\psi_\epsilon(\alpha_*,\Theta_*))^{1/2}}{\left(\phi_{\mathcal{M}}(\chi_*)(\phi_{\mathcal{T}}(\lambda_*)+\phi_{\mathcal{V}(\delta)}(\mu_*))\right)^{1/2}+(\phi_{\mathcal{M}}(\kappa_*)\psi_\varepsilon(\alpha_*,\Theta_*))^{1/2}} \le \sqrt{\frac{\psi_\epsilon(\alpha_*,\Theta_*)}{\psi_\varepsilon(\alpha_*,\Theta_*)}}.$$

Observing that $\psi_\epsilon(\alpha_*, \Theta_*) = \psi_\varepsilon(\alpha_*, \Theta_*) + \alpha_* \log(\varepsilon/\epsilon) \le \log(e\varepsilon/\epsilon)\psi_\varepsilon(\alpha_*, \Theta_*)$ yields 2.70.

3^o. In situations where $\mathcal{U}(\delta) = \mathfrak{f}(\delta)\mathcal{U}$, one has that $\phi_{\mathcal{V}(\delta)}(\cdot) = \phi_{\mathfrak{f}(\delta)^2\mathcal{V}}(\cdot) = \mathfrak{f}(\delta)^2\phi_{\mathcal{V}}(\cdot)$, with $\mathcal{V}$ defined in the ellitopic description of $\mathcal{U} = \left\{u \in \mathbf{R}^{n_u} : t \in \mathcal{V} : u^T V_k u \le t_k, k \le \mathfrak{n}_u\right\}$. Thus, using the same arguments as for proof of 2.70, we arrive at

$$\frac{\mathfrak{r}_{(1+\vartheta)\delta,\epsilon}(H)}{\mathfrak{r}_{\delta,\varepsilon}(H)} \le \frac{\left(\phi_{\mathcal{M}}(\chi_*)(\phi_{\mathcal{T}}(\lambda_*)+\mathfrak{f}((1+\vartheta)\delta)^2\phi_{\mathcal{V}}(\mu_*))\right)^{1/2}+(\phi_{\mathcal{M}}(\kappa_*)\psi_\epsilon(\alpha_*,\Theta_*))^{1/2}}{(\phi_{\mathcal{M}}(\chi_*)(\phi_{\mathcal{T}}(\lambda_*)+\mathfrak{f}(\delta)^2\phi_{\mathcal{V}}(\mu_*)))^{1/2}+(\phi_{\mathcal{M}}(\kappa_*)\psi_\varepsilon(\alpha_*,\Theta_*))^{1/2}} \le \max\left\{\mathfrak{b}(\vartheta); \sqrt{\log(e\varepsilon/\epsilon)}\right\}.$$

4^o. Recalling that

$$\mathfrak{r}^{(2)}_{\delta,\varepsilon}(H) = \phi_{\mathcal{T}}(\lambda_1) + \phi_{\mathcal{T}(\delta)}(\mu_1) + \psi_\varepsilon(\alpha_1, \Theta_1),$$

one can apply arguments from **2^o** and **3^o** to arrive at results 2.70 and 2.71. For function $\overline{\mathfrak{p}}$, recall that it is of the form

$$\overline{\mathfrak{p}}_{\delta,\varepsilon}(\Theta) = \phi_{\mathcal{M}}(\kappa_2) + \phi_{\mathcal{T}}(\lambda_2) + \phi_{\mathcal{T}(\delta)}(\mu_2) + 2\log(2m\varepsilon^{-1})\mathbf{Tr}(\Theta)$$

for some optimal variables $[\kappa_2, \lambda_2, \mu_2]$. Notice that with the same argument as in **1^o**, we arrive at the fact that optimal value of 2.34 is of the form

$$\overline{\mathfrak{p}}_{\delta,\varepsilon}(\Theta) = \sqrt{\phi_{\mathcal{M}}(\kappa_2)\left[\phi_{\mathcal{T}}(\lambda_2) + \phi_{\mathcal{T}(\delta)}(\mu_2) + 2\log(2m\varepsilon^{-1})\mathbf{Tr}(\Theta)\right]}.$$

One can then apply the reasonning of **2^o** and **3^o** to conclude. □

Chapter 3

Robust Counterparts of Linear Inverse Problems: Extensions

3.1 Simplified computations in the case of $\mathcal{U}(\delta) = \delta\mathcal{U}$

We treat in this subsection a special case where problem structure allows for a simple computation of the grid on parameter δ that ensures items (2.31) and (2.36) in the construction of adaptive linear and polyhedral estimates for $\vartheta = 1$. In this section, we set $\mathcal{U}(\delta) = \delta\mathcal{U}$, i.e. $\mathcal{U}(\delta)$ is the homothety of ellitope

$$\mathcal{U} = \left\{u \in \mathbf{R}^{n_u} : \ t \in \mathcal{V} : \ u^T V_k u \leq t_k, k \leq \mathfrak{n}_u\right\}. \tag{3.1}$$

For such ellitope, support function $\phi_{\mathcal{V}}(\delta)(\cdot) = \delta^2\phi_{\mathcal{V}}(\cdot)$. Indeed, for a given $u \in \mathcal{U}$ and $\mu \in \mathbf{R}^{\mathfrak{n}_u}_+$, their exist $t \in \mathcal{V}$ such that

$$(\delta u)^T V(\mu)(\delta u) = \delta^2 \sum_k^{\mathfrak{n}_u} \mu_k u^T V_k u \leq \delta^2 u^T t \leq \delta^2 \max_{v\in\mathcal{V}} v^T\mu = \delta^2\phi_{\mathcal{V}}(\mu).$$

Thanks to the latter, it is easy to see that if the $\delta_i/\delta_{i-1} \leq 2$ for $i > 1$, then (2.31) and (2.36) are also true for $\vartheta = 1$. For instance, recall that for linear estimates, our risk upper bound is actually of the form:

$$\varrho_*(\delta,\varepsilon) := \sqrt{\phi_{\mathcal{M}}(\chi_*)\left(\phi_{\mathcal{T}}(\lambda_*) + \delta^2\phi_{\mathcal{V}}(\mu_*)\right)} + \sqrt{\phi_{\mathcal{M}}(\kappa_*)\psi_\varepsilon(\alpha_*,\Theta_*)},$$

with $(\chi_*, \lambda_*, \mu_*, \Theta_*, \kappa_*, \alpha_*)$ optimal for computation of $\varrho_*(\delta,\varepsilon)$. For grids with first element strictly greater than 0, the work is done. If this is not the case, one has to carefully select δ_1 to ensure that for $0 \leq \delta < \delta_1$ and some constant C,

$$\varrho_*(\delta_1,\varepsilon) \leq C\varrho_*(\delta,\varepsilon)$$

holds when considering linear estimates. Note that since our upper bounds are increasing in δ, it is sufficient that the latter holds for 0 instead of δ. Thankfully, it is possible to compute quantities $r_0(\varepsilon)$ and $\mathfrak{l}_0(\varepsilon)$ such that for all $\delta \geq 0$, and $\varepsilon \in (0,1)$,

$$\varrho_*(\delta,\varepsilon) \leq \varrho_*(0,\varepsilon) + \delta^2 r_0(\varepsilon) \tag{3.2}$$

and

$$\mathfrak{p}_*(\delta,\varepsilon) \leq \mathfrak{p}_*(0,\varepsilon) + \delta^2\mathfrak{l}_0(\varepsilon), \tag{3.3}$$

and simply setting $\delta_1^2 = \varrho_*(0,\varepsilon)/r_0(\varepsilon)$ or $\mathfrak{p}_*(0,\varepsilon)/\mathfrak{l}_0(\varepsilon)$ yields the desired relationship. Indeed, for (3.2), consider variables $(H_0, \Theta_0, \alpha_0, \lambda_0, \chi_0)$ optimal for computation of $\varrho_*(0,\varepsilon)$ as proposed in (2.26). Then, for any $\mu \in \mathbf{R}^{\mathfrak{n}_u}_+$ such that

$$\left[\begin{array}{c|c|c} T(\lambda_0) & & \frac{1}{2}[H_0^TA - B]^TP \\ \hline & V(\mu) & \\ \hline \frac{1}{2}P^T[H_0^TA - B] & & M(\chi_0) \end{array}\right] \succeq 0, \quad \left[\begin{array}{c|c} \Theta_0 & \frac{1}{2}H_0P \\ \hline \frac{1}{2}P^TH_0^T & M(\kappa_0) \end{array}\right] \succeq 0, \tag{3.4}$$

variables $(\lambda_0, \mu, \Theta_0, \kappa_0)$ are admissible for the minimization program involved for the computation of $\varrho_*(\delta,\varepsilon)$. Using their sub-optimality, one directly obtains

$$\begin{array}{rl} \varrho_*(\delta,\varepsilon) & \leq \phi_{\mathcal{T}}(\lambda_0) + \phi_{\mathcal{M}}(\chi) + \phi_{\mathcal{M}}(\kappa_0) + \psi_\varepsilon(\alpha_0,\Theta_0) + \delta^2\phi_{\mathcal{V}}(\mu) \\ & = \varrho_*(0,\varepsilon) + \delta^2\phi_{\mathcal{V}}(\mu). \end{array}$$

Notice that first semidefinite constraint in (3.4) can be reformulated into

$$\left[\begin{array}{c|c} T(\lambda_0) & \\ \hline & V(\mu) \end{array}\right] \succeq \frac{1}{4}[H_0^T A - B]^T P M(\chi_0)^{-1} P^T [H_0^T A - B] =: V_{\rm lin}(H_0, \kappa_0).$$

Minimizing $\phi_{\mathcal{V}}(\mu)$ over μ satisfying this constraint allows to compute $r_0(\varepsilon)$ as

$$r_0(\varepsilon) := \min_{\mu \in \mathbf{R}_+^{n_u}} \left\{ \phi_{\mathcal{V}}(\mu) : \left[\begin{array}{c|c} T(\lambda_0) & \\ \hline & V(\mu) \end{array}\right] \succeq V_{\rm lin}(H_0, \chi_0) \right\}. \tag{3.5}$$

Observe that the "new" semidefinite constraint is just the Schurr complement of the first constraint in (3.4). As the variable $\widetilde{\mu}_0$ optimal for (3.5) satisfies the first semidefinite constraint in (3.4), (3.2) holds.

When considering the polyhedral estimate, the reasinoning presented above can also be applied. Let $(\Theta_0, \mu_0, \lambda_0, \kappa_0)$ be optimal variables in (2.35) for computation of $\mathfrak{p}_*(0, \varepsilon)$. Considering a vector μ that satisfies the semidefinite constraint,

$$\left[\begin{array}{c|c} \left[\begin{array}{c|c} T(\lambda_0) & \\ \hline & V(\mu) \end{array}\right] + A^T \Theta A & \frac{1}{2} B^T P \\ \hline \frac{1}{2} P^T B & M(\kappa_0) \end{array}\right] \succeq 0, \tag{3.6}$$

one has, using that $(\Theta_0, \lambda_0, \mu, \kappa_0)$ are feasible and sub-optimal for computation of $\mathfrak{p}_*(\delta, \varepsilon)$ yields

$$\begin{array}{rl} \mathfrak{p}_*(\delta, \varepsilon) & \leq \phi_{\mathcal{T}}(\lambda_0) + \phi_{\mathcal{M}}(\kappa_0) + \sigma^2 \ln(2m/\varepsilon) \mathrm{Tr}(\Theta) + \delta^2 \phi_{\mathcal{V}}(\mu) \\ & = \mathfrak{p}_*(0, \varepsilon) + \delta^2 \phi_{\mathcal{V}}(\mu). \end{array}$$

Thus, computing $\mathfrak{l}_0(\varepsilon)$ in (3.3) boils down to solving the minimization problem

$$\begin{array}{rl} \mathfrak{l}_0(\varepsilon) & := \min_{\mu \in \mathbf{R}_+^{n_u}} \left\{ \phi_{\mathcal{V}}(\mu) : \left[\begin{array}{c|c} T(\lambda_0) & \\ \hline & V(\mu) \end{array}\right] \succeq V_{\rm poly}(\Theta_0, \kappa_0) \right\}, \\ V_{\rm poly} & := \frac{1}{4} B^T P M(\kappa_0)^{-1} P^T B - A^T \Theta_0 A, \end{array} \tag{3.7}$$

where we have again used Schurr's lemma to reformulate constraint (3.6).

Remark 3.1.1 *In the special case where $\|\cdot\| = \|\cdot\|_2$, the reasonning and definitions of $r_0^{(2)}$ and $\mathfrak{l}_0^{(2)}$ are almost the same. Indeed, one only needs to replace $M(\chi_0)$ in(3.5) and $M(\kappa_0)$ in (3.7) by respectively $I - 2(\alpha_0)^{-1}\Theta_0$ and I, where α_0 is optimal in the minimization problem associated with the definition of $\varrho_*^{(2)}(0, \varepsilon)$.*

We can now consider the following grid construction.

1. For the adaptive linear estimate, set

$$I = \left\lceil \log_2 \left(\overline{\delta} \sqrt{\frac{r_0(\varepsilon)}{\varrho_*(0, \varepsilon)}} \right) \right\rceil, \tag{3.8}$$

and for the adaptive polyhedral estimate, set

$$I = \left\lceil \log_2 \left(\overline{\delta} \sqrt{\frac{\mathfrak{l}_0(\varepsilon)}{\mathfrak{p}_*(0, \varepsilon)}} \right) \right\rceil. \tag{3.9}$$

2. Set $\epsilon = \frac{\varepsilon}{I+1}$.

3. Set $\delta_1 = 2^{-I}\overline{\delta}$, and for all $i = 2, ..., I$, $\delta_i = 2\delta_{i-1}$.

Proposition 3.1.1 *With the following construction, one has*

$$\mathrm{Risk}_\varepsilon[\widehat{w}_{\mathrm{lin}}^{(a)}|\mathcal{Z}(\delta)] \leq 5\sqrt{\ln(e\log_2(2\overline{\delta}\sqrt{r_0(\varepsilon)/\varrho_*(0,\varepsilon)}))}\varrho_*(\delta,\varepsilon), \tag{3.10}$$

$$\mathrm{Risk}_\varepsilon[\widehat{w}_{\mathrm{lin}}^{(a)}|\mathcal{Z}(\delta)] \leq 4\sqrt{\ln(e\log_2(2\sqrt{\overline{\delta}\mathfrak{l}_0(\varepsilon)/\mathfrak{p}_*(0,\varepsilon)}))}\mathfrak{p}_*(\delta,\varepsilon). \tag{3.11}$$

In the case of the recovery norm being the Euclidean norm, replacing $r_0(\varepsilon)$ and $\varrho_*(0,\varepsilon)$ in (3.8) by $r_0^{(2)}(\varepsilon)$ and $\varrho_*^{(2)}(0,\varepsilon)$ leads to the following performances of the adaptive linear estimate presented in section 2.4:

$$\mathrm{Risk}_\varepsilon^{\|\cdot\|_2}[\widetilde{w}_{\mathrm{lin}}^{(a)}|\mathfrak{M}(\delta)]^2 \leq 2\ln\left(e\log_2\left(2\overline{\delta}r_0^{(2)}/\varrho_*^{(2)}(0,\varepsilon)\right)\right)\varrho_*^{(2)}(\delta,\varepsilon) + 4\mathfrak{d}_{l_*}, \tag{3.12}$$

where

$$r_0^{(2)} := \min_{\mu\geq 0}\left\{\phi_{\mathcal{V}}(\mu),\quad \left[\begin{array}{c|c} \left[\begin{array}{c|c} T(\lambda_0) & \\ \hline & V(\mu)\end{array}\right] & A^TH_0 - B^T \\ \hline (H_0)^TA - B & I - 2(\alpha_0)^{-1}\Theta_0\end{array}\right] \succeq 0\right\}. \tag{3.13}$$

and variables $[\lambda_0, H_0, \Theta_0, \alpha_0]$ are optimal for computation of $\varrho_*^{(2)}(0,\varepsilon)$.

For proof, see 3.4.1

3.2 Adaptive estimates robust to sparse perturbation

We assume in this section that observations are of the form

$$\omega = Ax + u + \xi \tag{3.14}$$

with $A \in \mathbf{R}^{m\times n}$, $x \in \mathcal{X}$ with $\mathcal{X}$ defined by (2.16), $\xi \sim \mathcal{N}(0, \sigma^2 I)$ and, for an unknown integer $s \geq 0$, deterministic but unknown s-sparse perturbation u, i.e., $\|u\|_0 = s$. As a consequence, our goal is to produce estimates of Bx that are adaptive to the sparsity level s, i.e. with the smallest possible ϵ-maximal risk over all x and u, measured in a co-ellitopic recovery norm $\|\cdot\|$:

$$\mathrm{Risk}_\epsilon[\widehat{w}|\mathfrak{m}(s)] := \sup_{x\in\mathcal{X},\ u:\ \|u\|_0=s} \inf\left\{\rho : \mathrm{Prob}_\xi\{\|Bz - \widehat{w}(Ax+u+\xi)\| > \rho\} \leq \epsilon\right\}. \tag{3.15}$$

Before building such estimates, we present the sparse recovery framework we will use to do so.

3.2.1 Verifyable design conditions, penalized and regular ℓ_1 recovery.

In this subsection, we recall two recovery routine of sparse vector u given observation

$$\omega = Au + v + \xi,$$

where $u \in \mathbf{R}^n$ is s-sparse, $v \in \mathbf{R}^m$ is a deterministic perturbation and ξ is a random vector. We also assume that $v \in \mathcal{V}$, where $\mathcal{V}$ is a closed convex and bounded subset of $\mathbf{R}^m$. For a given *contrast* matrix H, we define the following two ℓ_1 *recovery* routines jointly introduced in [83].

Definition 15 ([32], Chapter 1) *Given a contrast parameter $H \in \mathbf{R}^{m\times n}$ and $\mu > 0$, the ℓ_1-regular recovery estimate is defined as*

$$\widehat{x}^{H,\mu}_{reg}(\omega) \in \underset{z}{Argmin} \left\{ \|z\|_1 : \ \left\|H^T\left[\omega - Az\right]\right\|_\infty \leq \mu \right\}. \tag{3.16}$$

In the special cases of $\xi \sim \mathcal{N}(0, \sigma^2 \boldsymbol{I})$ or $\xi \sim \mathcal{SG}(0, \sigma^2 \boldsymbol{I})$, it is defined, for $\epsilon \in (0,1)$, by $\widehat{x}^{H}_{reg}(\omega) := \widehat{x}^{H,\Lambda(H)}_{reg}(\omega)$, where

$$\Lambda(H) := \max_i \left\{ \sigma\sqrt{2\ln(2n/\epsilon)}\, \|\mathrm{Col}_i(H)\|_2 + \max_{v\in\mathcal{V}} v^T \mathrm{Col}_i(H) \right\}. \tag{3.17}$$

Remark 3.2.1 *Note that $\Lambda(H)$ is simply an upper bound on the ϵ quantile of random quantity $\left\|H^T[v+\xi]\right\|_\infty$ (see [83]).*

Definition 16 ([32], Chapter 1) *Given a contrast parameter $H \in \mathbf{R}^{m\times m}$ and a sparsity parameter $s \geq 1$, the ℓ_1-penalized recovery is defined by*

$$\widehat{x}^{H,s}_{pen}(\omega) \in \underset{z}{Argmin} \left\{ \|z\|_1 + 2s \left\|H^T\left[\omega - Az\right]\right\|_\infty \right\}. \tag{3.18}$$

Results on the ϵ-risk over s-sparse vector of these two estimates are available in [83], provided that contrast matrix H satisfies the following condition, also introduced in [83].

Definition 17 ([32], Chapter 1) *Given a design matrix $A \in \mathbf{R}^{m\times n}$, a matrix $H \in \mathbf{R}^{m\times n}$ is said to satisfy the $\boldsymbol{Q}_\infty(s,\kappa)$ condition if their exist $\kappa \in (0, 1/2)$ such that the following holds:*

$$\forall z, \quad \|z\|_\infty \leq \left\|H^T Az\right\|_\infty + \frac{\kappa}{s}\|z\|_1. \tag{3.19}$$

Proposition 3.2.1 ([83], Proposition 2 and 4) *Let $H \in \mathbf{R}^{m\times n}$ satisfy the $\boldsymbol{Q}_\infty(s,\kappa)$ condition, noise ξ be either $\mathcal{N}(0,\sigma^2 I)$ or $\mathcal{SG}(0,\sigma^2 I)$, and $\epsilon \in (0,1)$. Then the following bound holds with probability greater than $1-\epsilon$, as soon as x is s-sparse,*

$$\begin{aligned} \left\|\widehat{x}^{H}_{reg} - x\right\|_1 &\leq \tfrac{4s}{1-2\kappa}\Lambda(H) \\ \left\|\widehat{x}^{H}_{reg} - x\right\|_\infty &\leq \tfrac{4}{1-2\kappa}\Lambda(H). \end{aligned} \tag{3.20}$$

Similarly, on an event of probability greater than $1-\epsilon$, the penalised ℓ_1-recovery enjoys the same rate under the $\boldsymbol{Q}_\infty(s,\kappa)$:

$$\begin{aligned} \left\|\widehat{x}^{H}_{pen} - x\right\|_1 &\leq \tfrac{4s}{1-2\kappa}\Lambda(H) \\ \left\|\widehat{x}^{H}_{pen} - x\right\|_\infty &\leq \tfrac{4}{1-2\kappa}\Lambda(H). \end{aligned} \tag{3.21}$$

Our interest in these estimates and the aforementionned condition resides in the fact that, in contrast with classical design conditions such as RIP or RE, condition $\mathbf{Q}_\infty(s,\kappa)$ is stated given design matrix A, and as the next proposition will show, one can efficiently compute a contrast H satisfying it if one exists.

Proposition 3.2.2 ([32], Proposition 1.10) *Consider the following LP programs*

$$\forall i \in [1:n], \quad \mathrm{Opt}_i := \min_{\gamma,h} \left\{\gamma : \ \left\|A^T h - e_i\right\|_\infty \leq \gamma\right\}. \tag{3.22}$$

Given $\kappa \in (0, \frac{1}{2})$, *if* $\gamma_* := \max_i \mathrm{Opt}_i < \kappa$, *let* $\overline{s} := \left\lfloor \frac{\kappa}{\gamma_*} \right\rfloor$. *Consider the following clearly solvable convex minimization program*

$$\mathrm{Opt} := \min_{H \in \mathbf{R}^{m\times n}} \left\{\Lambda(H) : \ \forall i \in [1;n], \ \left\|A^T \mathrm{Col}_i(H) - e_i\right\|_\infty \leq \gamma_*\right\}, \tag{3.23}$$

along with optimal variable H_*. *The latter satisfies* $\boldsymbol{Q}_\infty(\kappa, s)$ *for all* $s \leq s_*$.

Therefor, using this verifyable condition, it is possible, given a design matrix A, to compute the best parameter H leading to "minimal" bounds (3.21),(3.20), and the maximal sparsity level for which these hold.

3.2.2 A polyhedral estimate adaptive to sparsity

Assuming that u has at least one non-vanishing component, we propose the following strategy. First, define

$$\gamma_* := \max_i \mathrm{Opt}_i, \quad \forall i \in [1:m], \ \ \mathrm{Opt}_i = \min_{\gamma,h} \left\{\gamma : \ \|h - e_i\|_\infty \leq \gamma\right\}. \tag{3.24}$$

It is obvious that $\gamma_* = 0$, which implies using Proposition 3.2.2 that in this specific setting, the set of contrast satisfying $\mathbf{Q}_\infty(1/4, s)$ is nonempty for all $s \leq m$. As a consequence, we consider the following grid for the unknown sparsity level s:

$$\forall i \in [1:I], \ \ s_i = 2^{i-1}, \ \ s_I = m, \ \ I = \lfloor \log_2(m) \rfloor. \tag{3.25}$$

We then consider the following m minimization programs defined for all $i \in [1:m]$ by

$$\overline{\mathrm{Opt}}_i = \min_{g \in \mathbf{R}^m} \left\{\sigma\sqrt{2\log(4m/\varepsilon)}\|g\|_2 + \phi_{\mathcal{X}}\left(g^T A\right), \ \ \|g - e_i\|_\infty \leq (4m)^{-1}\right\},$$

and define contrast $\overline{G} = [g_1, ..., g_m] \in \mathbf{R}^{m\times m}$, where g_i is optimal for $\overline{\mathrm{Opt}}_i$. Observe that by definition, $\Lambda(\overline{G}) := \max_i \overline{\mathrm{Opt}}_i$ is minimal over all matrices satisfying the $\mathbf{Q}_\infty(1/4, m)$ condition. We are now ready to define an adaptive polyhedral estimate robust to sparse perturbation. For matrix contrast $(H_1, ..., H_I)$, let $\widehat{w}^{(a)}_{Poly} := B\widehat{x}$ where

$$(\widehat{x}, \widehat{u}) \in \mathrm{Argmin}_{z,v} \left\{ \|v\|_1 : \begin{array}{c} \forall i \in [1:I], \ \left\|H_i^T\left[\omega - Az - v\right]\right\|_\infty \leq 1, \\ \left\|\overline{G}^T\left[\omega - v\right]\right\|_\infty \leq \Lambda(\overline{G}), \quad w \in \mathcal{X} \end{array} \right\}. \tag{3.26}$$

The next proposition specifies a way to build the contrasts $H_1, ..., H_I$ and the risk upper bound associated with that construction.

Proposition 3.2.3 *Assume that $1 \leq s \leq m$ and let $\bar{i}$ be the minimal $i \in \{1, ..., I\}$ such that $s \leq s_i$. For this specific situation, we set $\varkappa(\varepsilon) = \sigma\sqrt{2\ln(4m/\varepsilon)}$, $\Lambda := \Lambda(\overline{G})$ and define*

$$\overline{\mathfrak{p}}_{s,\varepsilon}[\Theta] := \min_{\lambda,\mu,\overline{\mu},\kappa} \left\{ \begin{array}{l} \phi_{\mathcal{T}}(\lambda) + 64\Lambda^2(2s\overline{\mu} + \|\mu\|_1) + \phi_{\mathcal{M}}(\kappa) + \varkappa(\varepsilon)^2\mathrm{Tr}(\Theta) : \\ \\ \left[\begin{array}{c|c|c} \left[\begin{array}{c|c} T(\lambda) & \\ \hline & \mathrm{Diag}(\mu) + \overline{\mu} I \end{array}\right] + A^T\Theta A & \frac{1}{2}B^TP \\ \hline \frac{1}{2}P^TB & M(\kappa) \end{array}\right] \succeq 0, \\ \\ \lambda \geq 0, \mu \geq 0, \overline{\mu} \geq 0, \kappa \geq 0. \end{array} \right\}$$

and $\mathfrak{p}_*(s,\varepsilon) := \min\limits_{\Theta \succeq 0} \overline{\mathfrak{p}}_{s,\varepsilon}[\Theta]$.

Compute for all $i = \{1, ..., I\}$, $\Theta_i \in \mathop{Argmin}\limits_{\Theta \succeq 0} \overline{\mathfrak{p}}_{s_i,\varepsilon/(2I)}[\Theta]$ and $H_i = \varkappa^{-1}\Gamma_i$, where $\Theta_i = \Gamma_i\mathrm{Diag}(\chi)\Gamma_i^T$ is Θ_i's SVD. The outlined adaptive polyhedral estimate has a risk such that

$$\mathrm{Risk}_\varepsilon[\widehat{w}^{(a)}_{\mathrm{poly}}|\mathcal{Z}(s)] \leq 2\overline{\mathfrak{p}}_*(s_{\bar{i}}, \varepsilon(2I)^{-1}) \tag{3.27}$$

$$\leq 2\max\left\{\sqrt{2}; \sqrt{\ln\left(2e\log_2(2m)\right)}\right\}\overline{\mathfrak{p}}_*(s,\varepsilon). \tag{3.28}$$

For proof, see 3.4.2

3.2.3 Lepski's procedure applied to linear estimate and robust estimation to unknown sparse perturbations

We propose, under essentially the same setting, a linear estimate version of the previous developments. Given the same grid of levels of sparsity as in previous subsection, consider, for all $j = \{1, ..., m\}$ the following minimization programs defined by

$$\overline{\mathrm{Opt}}_j(s,\varepsilon) = \min_{g \in \mathbf{R}^m} \left\{\sigma\sqrt{2\log(2m/\varepsilon)}\|g\|_2 + \phi_{\mathcal{X}}\Big(g^TA\Big), \ \ \|g - e_j\|_\infty \leq (4s)^{-1}\right\},$$

and define contrast $\overline{G}^{(i)} = [g_1^{(i)}, ..., g_m^{(i)}] \in \mathbf{R}^{m\times m}$, with $g_j^{(i)}$ optimal for $\overline{\mathrm{Opt}}_j(s_i, \varepsilon/(2I))$. Also, with notations $\upsilon(\varepsilon) = \sigma(1 + \sqrt{2\ln(2/\varepsilon)})$ and $\Lambda(s,\varepsilon) := \max_j \ \overline{\mathrm{Opt}}_j(s,\varepsilon)$, we define for contrast matrix $H \in \mathbf{R}^{m\times\nu}$, function

$$\overline{\varrho}_{s,\varepsilon}(H) := \min_{\lambda,\mu,\overline{\mu},\kappa,\Theta} \left\{ \begin{array}{l} \phi_M(\chi) + \phi_{\mathcal{T}}(\lambda) + 64\Lambda(s,\varepsilon)^2(2s\overline{\mu} + \|\mu\|_1) + \phi_{\mathcal{M}}(\kappa) + \upsilon(\varepsilon)^2\mathrm{Tr}(\Theta) : \\ \\ \left[\begin{array}{c|c|c} \left[\begin{array}{c|c} T(\lambda) & \\ \hline & \mathrm{Diag}(\mu) + \overline{\mu} I \end{array}\right] & \frac{1}{2}[B - H^TA]^TP \\ \hline \frac{1}{2}P^T[B - H^TA] & M(\chi) \end{array}\right] \succeq 0, \\ \\ \left[\begin{array}{c|c} \Theta & HP/2 \\ \hline P^TH^T/2 & M(\kappa) \end{array}\right] \succeq 0, \\ \\ \lambda \geq 0, \mu \geq 0, \overline{\mu} \geq 0, \kappa \geq 0. \end{array} \right\} \tag{3.29}$$

along with its minimum over all matrices H, $\varrho_*(s,\varepsilon)$.
We propose the following adaptive procedure. Given observation ω, define for all $i=\{1,...,I\}$

- matrix contrasts $H_i \in \operatorname*{Argmin}_H \overline{\varrho}_{s_i,\varepsilon/(2I)}(H)$ along with associated $\varrho_*(s_i,\varepsilon/(2I)) = \overline{\varrho}_{s_i,\varepsilon/(2I)}(H_i)$.

- new observation $\omega_i = \omega - \widehat{u}_i$, where

$$\widehat{u}_i := \widehat{x}_{pen}^{G_i,s_i}(\omega) \tag{3.30}$$

- candidate estimate $w_i = H_i^T[\omega_i]$.

- admissible set

$$\mathcal{A} := \left\{ i \in [1:I],\ \forall j \in [i+1:I],\ \|w_i - w_j\| \le \overline{\varrho}_{s_i,\varepsilon/(2I)}(H_i) + \overline{\varrho}_{s_j,\varepsilon/(2I)}(H_j) \right\},$$

  and adaptive linear estimate $\widehat{w}_{\rm lin}^{(a)} := w_{\widehat{i}}$, with $\widehat{i} = \min\ \mathcal{A}$ is the latter is not empty, or w_I otherwise.

We have the following equivalent of Proposition 3.2.3.

Proposition 3.2.4 *Assume that $1 \le s \le \overline{s}$, and let $\bar{i}$ be the minimal $i \in \{1,...,I\}$ such that $s \le s_i$. The estimate $\widehat{w}_{lin}^{(a)}$ is such that*

$$\begin{aligned} \mathrm{Risk}_\varepsilon[\widehat{w}_{lin}^{(a)} | \mathcal{Z}(s)] &\le 2\varrho_*(s_{\bar{i}},\varepsilon/(2I)) + \varrho_*(s_{\bar{i}-1},\varepsilon/(2I)) &(3.31)\\ &\le (2\sqrt{2}+1)\sqrt{\ln\left(e\log_2(2m)\right)}\overline{\varrho}_*(s,\varepsilon). &(3.32)\end{aligned}$$

For proof, see 3.4.2

3.3 Comprehensive robust counterpart

In this section, we assume the same setting as in section 3.1. Given $\epsilon \in (0,1)$, let r be an upper bound of the risk of the near-optimal linear estimate in the case of $\delta = 0$, i.e., $r \ge \varrho_*(0,\epsilon)$ as defined in Proposition 2.3.1. Our objective is to design the contrast $H = H(r,\tau)$ satisfying two requirements:

(i) The ε-risk of recovery by estimate $\widehat{w}^H(\omega) = H^T\omega$ of the signal $w = Bz$ for $z \in \mathcal{Z}(0)$ does not exceed r;

(ii) Whenever $u \in \delta\mathcal{U}$ for some $\delta > 0$, the risk of the estimate $\widehat{w}^H$ does not exceed $r + \tau\delta$;

and we look for $H(r,\tau)$ such that τ is the smallest possible. Following [84] we refer to the corresponding setting as *comprehensive robust counterpart* of the original estimation problem, and we call τ *global sensitivity* of the estimate $\widehat{w}^H$.

3.3.1 Case of euclidean recovery norm

Let us consider first the case of Euclidean norm $\|\cdot\| = \|\cdot\|_2$. Given $r \geq \varrho_*^{(2)}(0,\epsilon)$ where $\varrho_*^{(2)}(0,\epsilon)$ is the bound of Proposition 2.4.2 for the squared Euclidean risk of nearly optimal linear estimate, denote

$$\tau_2(\epsilon) = \min_{\lambda,\mu,\alpha,\Theta,H} \left\{ \phi_{\mathcal{V}}(\mu) : \ \phi_{\mathcal{T}}(\lambda) + \psi_\epsilon(\alpha,\Theta) \leq r,\ \lambda \geq 0, \mu \geq 0, \right. \tag{3.33}$$
$$\left. \alpha \geq 2\lambda_{\max}(\Theta),\ \left[\begin{array}{c|c} \Theta & H \\ \hline H^T & I \end{array}\right] \succeq 0,\ \left[\begin{array}{c|c} T(\lambda) & \\ \hline & V(\mu) \\ \hline H^TA - B & \end{array} \begin{array}{c} \\ \hline \\ \hline \end{array} \begin{array}{c|} A^TH - B^T \\ \hline \\ \hline I - 2\alpha^{-1}\Theta \end{array}\right] \succeq 0 \right\}$$

where

$$\psi_\epsilon(\alpha,\Theta) := -\tfrac{\alpha}{2}\log\mathrm{Det}(I - 2\alpha^{-1}\Theta) + \alpha\ln[\epsilon^{-1}].$$

Let now $\delta \geq 0$, and let $(\lambda_*, \mu_*, \alpha_*, \Theta_*, H_*)$ be an optimal solution to (3.33). Because $(\lambda_*, \mu_*, \alpha_*, \Theta_*)$ also form a feasible solution to (2.43) with $H = H_*$, we conclude that the optimal value $\mathfrak{r}^{(2)}_{\delta,\epsilon}[H_*]$ of (2.43) satisfies

$$\mathfrak{r}^{(2)}_{\delta,\epsilon}[H_*] \leq \phi_{\mathcal{T}}(\lambda_*) + \delta^2\phi_{\mathcal{V}(\delta)}(\mu_*) - \tfrac{\alpha_*}{2}\ln\mathrm{Det}(I - 2\alpha_*^{-1}\Theta_*) + \alpha_*\ln[\epsilon^{-1}] \leq r + \delta^2\tau_2(\epsilon).$$

Applying the result of Proposition 2.4.2, we conclude that the risk of the estimate $\widehat{w}^{H_*}(\omega) = H_*^T\omega$ satisfies

$$\forall \delta \geq 0 \quad \mathrm{Risk}^2_\epsilon[\widehat{w}^{H_*}|Z(\delta)] \leq r + \delta^2\tau_2(\epsilon).$$

3.3.2 Case of general co-ellitopic recovery norm

Similar result in the case of the general co-ellitopic norm requires a bit more work—we need to replace the bound $\mathfrak{r}_{\delta,\epsilon}[H]$ for the risk of the linear estimate $\widehat{w}^H(\omega) = H^T\omega$, $H \in \mathbf{R}^{m\times\nu}$, we somewhat less precise bound which is more convenient for our purposes. Let us denote $A = [A_x, A_u]$ and $B = [B_x, B_u]$ with $A_x \in \mathbf{R}^{m\times n_x}$, $A_u \in \mathbf{R}^{m\times n_u}$, $B_x \in \mathbf{R}^{\nu\times n_x}$, $B_u \in \mathbf{R}^{\nu\times n_u}$. For all $z \in \mathcal{Z}(\delta)$, $\delta \geq 0$ the error of estimate $\widehat{w}^H(\omega)$ satisfies

$$\begin{aligned}\|H^T\omega - Bz\| &= \|[H^TA - B][x;u] + H^T\xi\| \leq \|[H^TA_x - B_x]x + [H^TA_u - B_u]u + H^T\xi\| \\ &\leq \|[H^TA_x - B_x]x\| + \|[H^TA_u - B_u]u\| + \|H^T\xi\|.\end{aligned}$$

To bound the terms in the r.-h.s. we act as follows. When denoting $V_x = [H^TA_x - B_x]$ (cf. (2.55)) we have

$$\forall x \in \mathcal{X} \quad a(V_x;x) := \|V_Xx\| \leq \max_{y\in\mathcal{Y}}\sup_{z\in\mathcal{Z}} y^TM^TV_xx \leq \tfrac{1}{2}\max_{[x;y]\in\mathcal{Z}\times\mathcal{Y}}[x;y]^T\left[\begin{array}{c|c} & [V_x^TM \\ \hline M^TV_x & \end{array}\right][x;y].$$

Because the set $\mathcal{G} = X \times \mathcal{Y}$ is a basic ellitope (as direct product of basic ellitopes), we may use Theorem 2.3.1 to bound the maximum of the quadratic form on $\mathcal{G}$, i.e.,

$$\forall x \in \mathcal{X} \quad a(V_x,x) \leq \overline{a}_x(V_x,\mathcal{X}) := \min_{\lambda,\chi}\left\{\phi_{\mathcal{T}}(\lambda) + \phi_{\mathcal{M}}(\chi) : \ \lambda \geq 0, \chi \geq 0, \right.$$
$$\left.\left[\begin{array}{c|c} T(\lambda) & \frac{1}{2}V_x^TM \\ \hline \frac{1}{2}M^TV_x & M(\chi) \end{array}\right] \succeq 0\right\}.$$

Similarly, we have for $V_u = [H^T A_u - B_u]$

$$\forall u \in \delta\mathcal{U} \quad a(V_u; u) = \|V_u u\| \le \overline{a}_u(V_x, \delta\mathcal{U}) = \min_{\lambda,\chi} \left\{ \delta^2 \phi_{\mathcal{V}}(\mu) + \phi_{\mathcal{M}}(\upsilon) : \ \lambda \ge 0, \upsilon \ge 0, \right.$$
$$\left. \left[\begin{array}{c|c} V(\lambda) & \frac{1}{2} V_u^T M \\ \hline \frac{1}{2} M^T V_u & M(\upsilon) \end{array} \right] \succeq 0 \right\}.$$

Together with the bound $\overline{\beta}(H, \epsilon)$ by Proposition 2.6.1 for the term $\|H^T \xi\|$, the above bounds result in

$$\mathrm{Risk}_\epsilon[\widehat{w}^H | Z(\delta)] \le \widetilde{\mathfrak{r}}_{\delta,\epsilon}[H]$$

where

$$\widetilde{\mathfrak{r}}_{\delta,\epsilon}[H] := \min_{\lambda,\mu,\chi,\upsilon\kappa,\alpha,\Theta} \left\{ \phi_{\mathcal{T}}(\lambda) + \phi_{\mathcal{M}}(\chi) + \delta^2 \phi_{\mathcal{V}}(\mu) + \phi_{\mathcal{M}}(\upsilon) + \phi_{\mathcal{M}}(\kappa) + \psi_\epsilon(\alpha, \Theta) : \right. \tag{3.34}$$
$$\lambda \ge 0, \chi \ge 0, \mu \ge 0, \upsilon \ge 0, \kappa \ge 0, \ \alpha \ge 2\lambda_{\max}(\Theta),$$
$$\left[\begin{array}{c|c} T(\lambda) & \frac{1}{2}[H^T A - B]^T M \\ \hline \frac{1}{2} M^T [H^T A - B] & M(\chi) \end{array} \right] \succeq 0,$$
$$\left. \left[\begin{array}{c|c} V(\mu) & \frac{1}{2}[H^T A - B]^T M \\ \hline \frac{1}{2} M^T [H^T A - B] & M(\upsilon) \end{array} \right] \succeq 0, \ \left[\begin{array}{c|c} \Theta & \frac{1}{2} HM \\ \hline \frac{1}{2} M^T H^T & M(\kappa) \end{array} \right] \succeq 0 \right\}.$$

Now we are almost done. Indeed, let

$$\tau(\epsilon) := \min_{\lambda,\mu,\chi,\kappa,\alpha,\Theta,H} \left\{ \tfrac{1}{2}(\phi_{\mathcal{V}}(\mu) + \phi_{\mathcal{M}}(\upsilon)) : \ \phi_{\mathcal{T}}(\lambda) + \phi_{\mathcal{M}}(\chi) + \phi_{\mathcal{M}}(\kappa) + \psi_\epsilon(\alpha, \Theta) \le r, \right. \tag{3.35}$$
$$\lambda \ge 0, \chi \ge 0, \mu \ge 0, \upsilon \ge 0, \kappa \ge 0, \ \alpha \ge 2\lambda_{\max}(\Theta),$$
$$\left[\begin{array}{c|c} T(\lambda) & \frac{1}{2}[H^T A - B]^T M \\ \hline \frac{1}{2} M^T [H^T A - B] & M(\chi) \end{array} \right] \succeq 0,$$
$$\left. \left[\begin{array}{c|c} V(\mu) & \frac{1}{2}[H^T A - B]^T M \\ \hline \frac{1}{2} M^T [H^T A - B] & M(\upsilon) \end{array} \right] \succeq 0, \ \left[\begin{array}{c|c} \Theta & \frac{1}{2} HM \\ \hline \frac{1}{2} M^T H^T & M(\kappa) \end{array} \right] \succeq 0 \right\}.$$

Similarly to (3.33), the problem in (3.35) is a well structured convex optimization problem which is solvable; let us denote $(\lambda_*, \mu_*, \chi_*, \upsilon_*, \kappa_*, \alpha_*, \Theta_*, H_*)$ its optimal solution. We note that for any $\delta \ge 0$, $(\lambda_*, \mu_*, \chi_*, \upsilon_*, \kappa_*, \alpha_*, \Theta_*)$ is also a feasible solution to (3.34) with $H = H_*$. Next, observe that if a pair (μ_*, υ_*) is feasible for (3.35) then so is $(\gamma\mu_*, \gamma^{-1}\upsilon_*)$ for any $\gamma > 0$, other components $(\lambda_*, \chi_*, \kappa_*, \alpha_*, \Theta_*)$ of the solution being unchanged. When minimizing w..r.t. γ, using homogeneity of $\phi_{\mathcal{V}}$ and $\phi_{\mathcal{M}}$, we obtain for $\bar{\gamma} = \delta^{-1}$:

$$\delta^2 \phi_{\mathcal{V}}(\bar{\gamma}\mu_*) + \phi_{\mathcal{M}}(\bar{\gamma}^{-1}\upsilon_*) = \delta^2 \bar{\gamma} \phi_{\mathcal{V}}(\mu_*) + \bar{\gamma}^{-1} \phi_{\mathcal{M}}(\upsilon_*) = \delta(\phi_{\mathcal{V}}(\mu_*) + \phi_{\mathcal{M}}(\upsilon_*)) = \delta\tau(\epsilon).$$

As a consequence,

$$\widetilde{\mathfrak{r}}_{\delta,\epsilon}[H_*] \le \phi_{\mathcal{T}}(\lambda_*) + \phi_{\mathcal{M}}(\chi_*) + \delta(\phi_{\mathcal{V}}(\mu_*) + \phi_{\mathcal{M}}(\upsilon_*)) + \phi_{\mathcal{M}}(\kappa_*) + \psi_\epsilon(\alpha_*, \Theta_*) \le r + \delta\tau(\epsilon),$$

and we conclude that the ϵ-risk of the estimate $\widehat{w}^{H_*}(\omega) = H_*^T \omega$ satisfies

$$\mathrm{Risk}_\epsilon[\widehat{w}^{H_*} | Z(\delta)] \le \widetilde{\mathfrak{r}}_{\delta,\epsilon}[H_*] \le r + \delta\tau(\epsilon).$$

3.3.3 Comprehensive polyhedral estimate in the Euclidean recovery norm case

Using a similar strategy, one can build a comprehensive polyhedral estimate in the case of Euclidian recovery norm. For $\delta \geq 0$, we define $\mathfrak{p}_*^{(2)}(\delta, \epsilon)$ as the minimum over semidefinite Θ of

$$\overline{\mathfrak{p}}_{\delta,\epsilon}^{(2)}(\Theta) = \min_{\lambda,\mu} \Big\{ \phi_{\mathcal{T}}(\lambda) + \delta^2 \phi_{\mathcal{V}}(\mu) + 2\sigma^2 \ln(2m/\epsilon)\mathrm{Tr}(\Theta),\ \lambda \geq 0, \mu \geq 0, \quad \left[\begin{array}{c|c} T(\lambda) & \\ \hline & V(\mu) \end{array}\right] \succeq B^T B - A^T \Theta A \Big\}. \tag{3.36}$$

Let r be an upper bound on $\mathfrak{p}_*^{(2)}(0, \epsilon)$, and consider

$$\mathfrak{q}_2(\epsilon) = \min_{\lambda,\mu,\Theta} \Big\{ \phi_{\mathcal{V}}(\mu) :\ \phi_{\mathcal{T}}(\lambda) + 2\sigma^2 \ln(2m/\epsilon)\mathrm{Tr}(\Theta) \leq r,\ \lambda \geq 0, \mu \geq 0, \quad \left[\begin{array}{c|c} T(\lambda) & \\ \hline & V(\mu) \end{array}\right] \succeq B^T B - A^T \Theta A \Big\}, \tag{3.37}$$

along with $[\lambda_*, \mu_*, \Theta_*]$ optimal for the latter. Since they are also feasible for computation of $\mathfrak{p}_*^{(2)}(0, \epsilon)$, one has that

$$\overline{\mathfrak{p}}_{\delta,\epsilon}^{(2)}(\Theta_*) \leq \phi_{\mathcal{T}}(\lambda_*) + 2\sigma^2 \ln(2m/\epsilon)\mathrm{Tr}(\Theta_*) + \delta^2 \phi_{\mathcal{V}}(\mu_*) \leq r + \mathfrak{q}_2(\epsilon).$$

3.4 Proofs

3.4.1 Proofs of section 3.1

Proof of 3.1.1: We can use the same strategy as for 2.3.2's proof. In cases where $\delta \geq \delta_1$, 2.71 (with $\mathfrak{f}(\delta) = \delta$ implying $\mathfrak{b}(\vartheta) = 1 + \vartheta$) yields the upper bound

$$\rho_{\bar{i}}/\varrho_*(\delta,\varepsilon) \leq \max\left\{2, \sqrt{\ln(e(I+1))}\right\} \leq 2\sqrt{\ln(e(I+1))}$$

and 2.70 gives

$$\rho_{\bar{i}-1}/\varrho_*(\delta,\varepsilon) \leq \sqrt{\ln(e(I+1))}.$$

When $\delta < \delta_1$, we first control the ratio $\varrho_*(\delta_1,\varepsilon)/\varrho_*(0,\varepsilon)$: Consider the sub-optimal variables $(H_0, \Theta_0, \alpha_0, \lambda_0, \chi_0)$ for computation of $\varrho_*(\delta_1,\varepsilon)$, along with μ feasible for 3.5. Clearly, one has

$$\varrho_*(\delta_1,\varepsilon) \leq \varrho_*(0,\varepsilon) + \delta_1\sqrt{\phi_{\mathcal{M}}(\chi_0)\phi_{\mathcal{V}}(\mu)}.$$

Minimization of the RHS for feasible μ and 2.70 yields

$$\rho_1/\varrho_*(\delta,\varepsilon) \leq \sqrt{\ln(e(I+1))}\left(1 + \frac{\delta_1 r_0}{\varrho_*(0,\varepsilon)}\right).$$

Finally, inequality $\mathrm{Risk}_\varepsilon[\widehat{w}^{(a)}_{\mathrm{lin}}|\mathcal{Z}(\delta)] \leq 2\rho_{\bar{i}} + \rho_{\bar{i}-1}$ allows to conclude that in all cases, one has

$$\mathrm{Risk}_\varepsilon[\widehat{w}^{(a)}_{\mathrm{lin}}|\mathcal{Z}(\delta)] \leq 5\sqrt{ln(e(I+1))}\varrho_*(\delta,\varepsilon).$$

When looking at the adaptive polyhedral estimate, using (2.71) for $\overline{\mathfrak{p}}$ naturally yields, for any semidefinite Θ, that

$$\overline{\mathfrak{p}}_{\delta_{\bar{i}},\epsilon}(\Theta) \leq \max\left\{\sqrt{\ln(e(I+1))}; 2\right\}\overline{\mathfrak{p}}_{\delta,\varepsilon}(\Theta),$$

implying

$$\overline{\mathfrak{p}}_*(\delta_{\bar{i}},\epsilon) \leq \max\left\{\sqrt{\ln(e(I+1))}; 2\right\}\overline{\mathfrak{p}}_*(\delta,\varepsilon).$$

For the case of $\delta < \delta_1$, proof is completely analogous to the developments above when replacing r_0 by $\mathfrak{l}_0$. This entails the upper bound

$$\begin{array}{rcl}
\overline{\mathfrak{p}}_*(\delta_{\bar{i}},\epsilon)/\overline{\mathfrak{p}}_*(\delta,\varepsilon) \leq & \max\left\{\sqrt{\ln(e(I+1))}; 2; \sqrt{\ln(e(I+1))}(1 + \frac{\delta_1\mathfrak{l}_0}{\overline{\mathfrak{p}}_*(0,\varepsilon)})\right\} \\
\leq & \sqrt{\ln(e(I+1))}\left(1 + \max\left\{1; 2^{-I}\frac{\bar{\delta}\mathfrak{l}_0}{\overline{\mathfrak{p}}_*(0,\varepsilon)}\right\}\right) \\
= & 2\sqrt{\ln(e(I+1))} \\
\leq & 2\sqrt{\ln\left(e\log_2\left(2\bar{\delta}\sqrt{\mathfrak{l}_0(\varepsilon)/\mathfrak{p}_*(0,\varepsilon)}\right)\right)}
\end{array}$$

□

3.4.2 Proofs of section 3.2

Proof of 3.2.3:

Consider the event $\mathcal{E} := \mathcal{E}_a \cap \mathcal{E}_b$, where

$$\mathcal{E}_a = \left\{\xi : \quad \|\overline{G}^T[Ax + \xi]\|_\infty \leq \Lambda(\overline{G})\right\}$$

and

$$\mathcal{E}_b = \left\{\xi : \quad \forall i \in [I], \quad \|H_i^T \xi\|_\infty \leq 1\right\}.$$

By construction, probability of $\mathcal{E}$ is greater than $1-\varepsilon$. Indeed, one has that

$$\text{Prob}\left[\mathcal{E}_a\right] \geq 1 - \varepsilon/2$$

and

$$\text{Prob}\left[\mathcal{E}_b\right] \geq 1 - \sum_{i=1}^{I} \text{Prob}\left[\left\|H_i^T \xi\right\|_\infty\right] \geq 1 - \sum_{i=1}^{I} \frac{\varepsilon}{2I} = 1 - \varepsilon/2.$$

On event $\mathcal{E}$, notice that (x, u) is admissible for problem 3.26. Since u is $s-$sparse, we can use 3.20 and the fact that $\overline{G}$ satisfies the $\mathbf{Q}_\infty(1/4, m)$ condition to state that

$$\|\hat{u} - u\|_1 \leq 16 s \Lambda(\overline{G}) \quad \text{and } \|\hat{u} - u\|_\infty \leq 8\Lambda(\overline{G}),$$

which in turn implies that

$$\|\hat{u} - u\|_2 \leq 8\sqrt{2s}\Lambda(\overline{G}).$$

From this, on event $\mathcal{E}$, $\hat{u} - u$ belongs to simple ellitope $8\Lambda(\overline{G})\mathcal{U}(s)$

$$\mathcal{U}(s) := \left\{z \in \mathbf{R}^m : \ \exists (t, \bar{t}) \in ([0:1]^m \times [0:2s]), \ \ \|z\|_2^2 \leq \bar{t} \text{ and } \forall i \in [1:m], \ \ z_i^2 \leq t_i\right\}.$$

Observe that if $s \leq s_i$, one also has $\hat{u} - u \in 2\Lambda_i \mathcal{U}(s_i)$. As a consequence, the convex upper bound

$$(\hat{u} - u)^T (\text{Diag}(\mu) + \overline{\mu} I)(\hat{u} - u) \leq 64\Lambda^2(\|\mu\|_1 + 2 s_i \overline{\mu})$$

holds for $\mu \in \mathbf{R}_+^m$ and non-negative $\overline{\mu}$, and the contrast constructions of this proposition along with the risk upper bound (3.27) are direct byproducts of subsection 2.3.3. The rest of the proof follows proof of 3.11, where one uses in that specific case the fact that $s \equiv \delta^2$, and that for all $i \in [1 : I-1]$, $s_i \leq 2 s_{i+1}$.

Proof of 3.2.4:

Consider the event $\mathcal{E} := \mathcal{E}_a \cap \mathcal{E}_b$, where

$$\mathcal{E}_a = \left\{\xi : \quad \forall i \in [1:I], \quad \|\overline{G}_i^T [Ax + \xi]\|_\infty \leq \Lambda_i\right\}$$

and

$$\mathcal{E}_b = \left\{\xi : \quad \forall i \in [I], \quad \|H_i^T \xi\|_\infty \leq 1\right\},$$

where $\Lambda_i := \Lambda(s_i, \varepsilon/(2I))$. By construction, probability of $\mathcal{E}$ is greater than $1-\varepsilon$. Indeed, one has that

$$\text{Prob}\left[\mathcal{E}_a\right] \geq 1 - \sum_{i=1}^{I} \text{Prob}\left[\left\|\overline{G}_i^T [Ax + \xi] \geq \Lambda_i\right\|_\infty\right] \geq 1 - \sum_{i=1}^{I} \frac{\varepsilon}{2I} = 1 - \varepsilon/2,$$

and

$$\text{Prob}\left[\mathcal{E}_b\right] \geq 1 - \sum_{i=1}^{I} \text{Prob}\left[\left\|H_i^T \xi\right\|_\infty\right] \geq 1 - \sum_{i=1}^{I} \frac{\varepsilon}{2I} = 1 - \varepsilon/2.$$

On event $\mathcal{E}$, notice that if u is s_i-sparse, the fact that $\overline{G}_i$ satisfies the $\mathbf{Q}_\infty(1/4, s_i)$ condition allows to state that

$$\|\hat{u}_i - u\|_1 \leq 16 s_i \Lambda_i \quad \text{and } \|\hat{u}_i - u\|_\infty \leq 8\Lambda_i,$$

which in turn implies that

$$\|\hat{u} - u\|_2 \leq 8\sqrt{2s_i}\Lambda_i.$$

From this, on event $\mathcal{E}$, if u is s_i-sparse, $\hat{u}_i - u$ belongs to simple ellitope $8\Lambda_i\mathcal{U}(s_i)$

$$\mathcal{U}(s) := \left\{z \in \mathbf{R}^m : \ \exists(t, \bar{t}) \in ([0:1]^m \times [0:2s]), \ \ \|z\|_2^2 \leq \bar{t} \text{ and } \forall i \in [1:m], \ \ z_i^2 \leq t_i\right\}.$$

As a consequence, the convex upper bound

$$(\hat{u} - u)^T(\mathrm{Diag}(\mu) + \overline{\mu}I)(\hat{u} - u) \leq 64\Lambda_i^2(\|\mu\|_1 + 2s_i\overline{\mu})$$

holds for $\mu \in \mathbf{R}^m_+$ and non-negative $\overline{\mu}$, and the contrast constructions of this proposition along with the risk upper bound (3.31) are direct byproducts of subsection 2.3.2. The rest of the proof follows proof of 3.10, where one uses in that specific case the fact that $s \equiv \delta^2$, and that for all $i \in [1 : I-1]$, $s_i \leq 2s_{i+1}$. □

3.5 Numerical Experiments

In this subsection, we provide numerical experiments illustrating results of Proposition 3.1.1, under the *uncertainty observation model* described as follows.

3.5.1 Uncertainty observation model and data points generation

For $i \in [0:6]$, we consider the grid

$$i \in [0:6], \quad \delta_0 = 0, \quad \delta_i = 2^{i-1}.$$

As far as uncertainty observation models are concerned, we work with the models

$$\mathfrak{m}(\delta_i) = [A, B, \mathcal{X} + \delta_i \mathcal{U}, \mathcal{N}(0, \sigma^2 I)],$$

where $\mathcal{X}$ is the intersection of 3 ellipsoids (a simple ellitope)

$$\mathcal{X} = \left\{ z \in \mathbf{R}^{50}, \quad \exists t \in [0:1]^3, \quad z^T R_i z \leq t_i \right\}.$$

and $\mathcal{U}$ is the following simple ellitope

$$\mathcal{U} = \left\{ z \in \mathbf{R}^{50}, \quad \exists t \in \mathbf{B}_2^6, \quad z^T V_j z \leq t_j \right\}, \quad \mathbf{B}_2^6 = \left\{ l \in \mathbf{R}^6; \|t\|_2 \leq 1 \right\}.$$

Matrices $A \in \mathbf{R}^{40\times 50}$ and $B \in \mathbf{R}^{35\times 50}$ are generated by drawing random matrices $\widetilde{A}, \widetilde{B}$ with coefficients following a $\mathcal{N}(0,1)$, computing the SVDs $\widetilde{A} = U_a D_a V_a^T$ and $\widetilde{B} = U_b D_b V_b^T$ and setting:

$$\begin{array}{llll} A = U_a Diag(\lambda(A)) V_a^T, & \lambda_k(A) = & 0.1 + (1-0.1) * (1 - \big(k/39\big)), & k \in [0:39] \\ B = U_b Diag(\lambda(B)) V_b^T, & \lambda_k(B) = & 0.1 + (1-0.1) * (1 - \big(k/35\big)), & k \in [0:34]. \end{array}$$

The same strategy is applied for generating matrices involved in ellitopic descriptions of $\mathcal{X}$ and $\mathcal{U}$. We draw a random square matrix S_i with $\mathcal{N}(0,1)$ coefficients, compute the SVD $S_i S_i^T = \Gamma_i D_i \Gamma_i^T$, and set

$$R_i = \Gamma_i Diag(\lambda^{(i)}) \Gamma_i^T, \quad \lambda_k^{(i)} = 0.1 + (1-0.1) * (1 - \big(k/49\big)^{1/i}), \ k \in [0:49].$$

When concerned with model $\mathfrak{m}(\delta_i)$, we generate signal $x + \delta_i u$ along with observation

$$\omega = A(x + \delta_i u) + \sigma\xi \in \mathbf{R}^{40} \tag{3.38}$$

in the following way. We draw $\widetilde{x} \sim \mathcal{N}(0, \mathbf{I}_{50})$, $\widetilde{u} \sim \mathcal{N}(0, \mathbf{I}_{50})$ and set

$$x = \widetilde{x} \left(\max_{i \leq 3} \widetilde{x}^T R_i \widetilde{x} \right)^{-\frac{1}{2}}, \quad u = \widetilde{u} \left(\sum_{j \leq 6} (\widetilde{u}^T V_j \widetilde{u})^2 \right)^{-\frac{1}{2}}.$$

Noise ξ is drawn from a $\mathcal{N}(0, I_{40})$. Finally, performance of estimates is calculated through quantity $\|B(x+u) - \widehat{w}^{Ada}(\omega)\|$, where $\|\cdot\|$ can either be $\|\cdot\|_2$ or $\|\cdot\|_1$, and $\widehat{w}^{Ada}(\omega)$ is the considered adaptive estimate.

3.5.2 plan of experiment

We repeat the proposed experiments for values of σ being $1, 0.1, 0.01$ and for the recovery norm $\|\cdot\|_1$ and $\|\cdot\|_2$.
For all $i \in [0;6]$, we compute

- the best contrast for linear and polyhedral estimates w.r.t. 0.01 and 0.01/7-maximal risk over $\mathfrak{m}(\delta_i)$. We respectively denote them H_i, G_i, $\overline{H}_i$ and $\overline{G}_i$.
- their associated upper bounds $\varrho_*(\delta_i, 0.01)$,$\mathfrak{p}_*(\delta_i, 0.01)$, $\varrho_*(\delta_i, 0.01/7)$ and $\mathfrak{p}_*(\delta_i, 0.01/7)$.

We then generate, for all $i \in [0;6]$ and all $k \in [1;100]$, observations $\omega_k^{(i)}$ of the form (3.38) under model $\mathfrak{m}(\delta_i)$, and compute

- the adaptive linear and polyhedral estimates $\widehat{w}_{Lin}^{Ada}(\omega_k^{(i)})$ and $\widehat{w}_{Poly}^{Ada}(\omega_k^{(i)})$ using contrasts $(\overline{H}_i)_{i=0}^I$ and $\overline{G} = [H_0; ...; H_I]$.
- estimates $\widehat{w}_{Lin}^{H_i}(\omega_k^{(i)})$, $\widehat{w}_{Poly}^{G_i}(\omega_k^{(i)})$.
- quantities $\|\widehat{w}_{Lin}^{Ada}(\omega_k^{(i)}) - w_k^{(i)}\|$, $\|\widehat{w}_{Poly}^{Ada}(\omega_k^{(i)}) - w_k^{(i)}\|$, $\|\widehat{w}_{Lin}^{H_i}(\omega_k^{(i)}) - w_k^{(i)}\|$ and $\|\widehat{w}_{Poly}^{G_i}(\omega_k^{(i)}) - w_k^{(i)}\|$, where $w_k^{(i)} = B(x_k^{(i)} + \delta_i u_k^{(i)})$ is the targeted signal.

We summarize the results of these experiments in the plots of the subsections to follow. Each box-plot represents the distribution of error $\|B(x+u) - \widehat{\omega}^{Ada}\|$, in the situation where noise is of magnitude σ, and signals are from $\mathcal{X} + \delta_i \mathcal{U}$. The whiskers of each box-plots represents the 0.05 and 0.95 quantiles, while the red bars represent the theoretical upper bounds previously computed for the 0.99 quantile of each estimate. We also include plots for the distribution of the ratio

$$\frac{\|B(x+u) - \widehat{w}^{Ada}\|}{\|B(x+u) - \widehat{w}^{(i)}\|},$$

where $\widehat{w}^{(i)}$ is the linear or polyhedral estimate that knows δ_i's value.

3.5.3 $\|\cdot\|_1$-recovery error

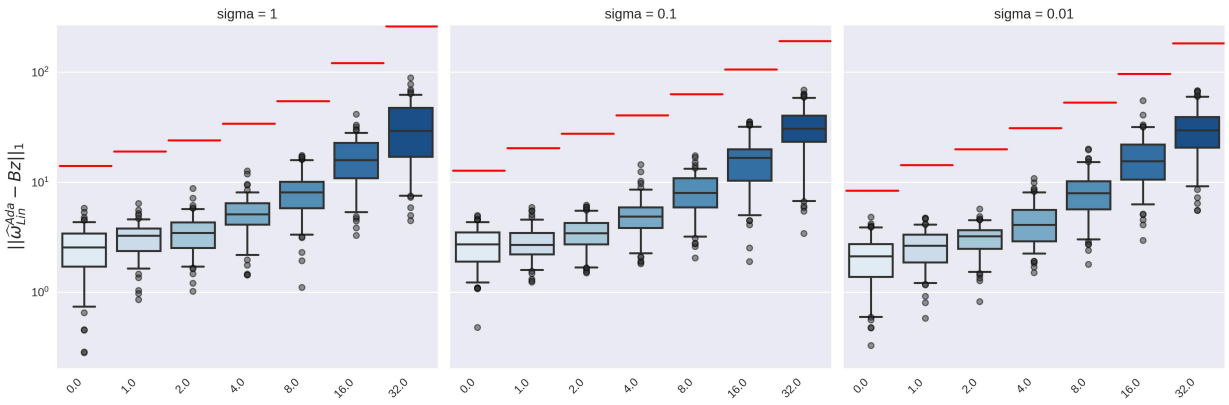


Figure 3.1: Evolution of the $\|\cdot\|_1$ error of the adaptive linear estimate as a function of parameter δ and noise σ

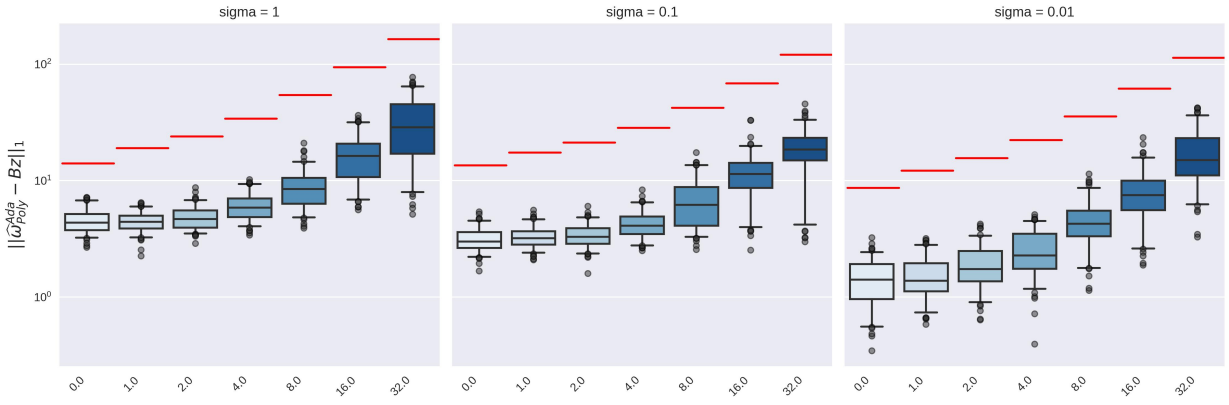


Figure 3.2: Evolution of the $\|\cdot\|_1$ error of the adaptive polyhedral estimate as a function of parameter δ and noise σ

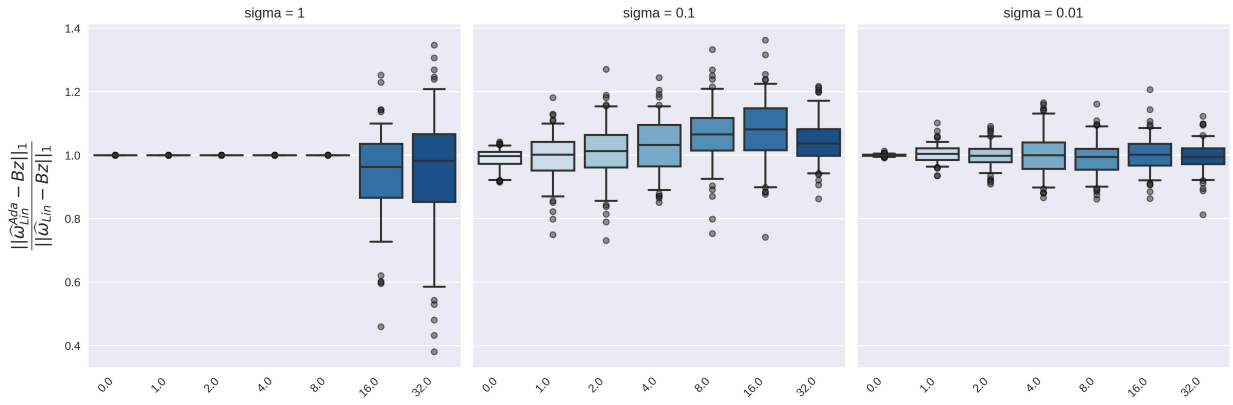


Figure 3.3: Evolution of $\|B(x+u) - \widehat{\omega}_{Lin}^{Ada}\|_1 / \|B(x+u) - \widehat{\omega}_{Lin}^{H_\delta}\|_1$ as a function of parameter δ and noise σ

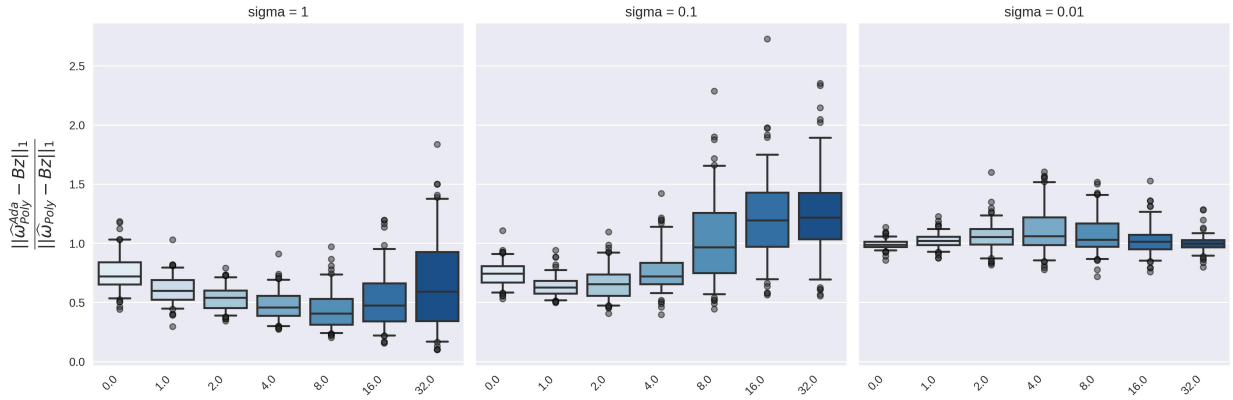


Figure 3.4: Evolution of $\|B(x+u) - \widehat{\omega}_{Poly}^{Ada}\|_1 / \|B(x+u) - \widehat{\omega}_{Poly}^{H_\delta}\|_1$ as a function of parameter δ and noise σ

3.5.4 $\|\cdot\|_2$-recovery error

When performance is measured using Euclidean norm, two selection routines are proposed to produce adaptive linear estimates. We thus compute all contrasts that are necessary in order to apply (2.4)'s method and Lepski's method.

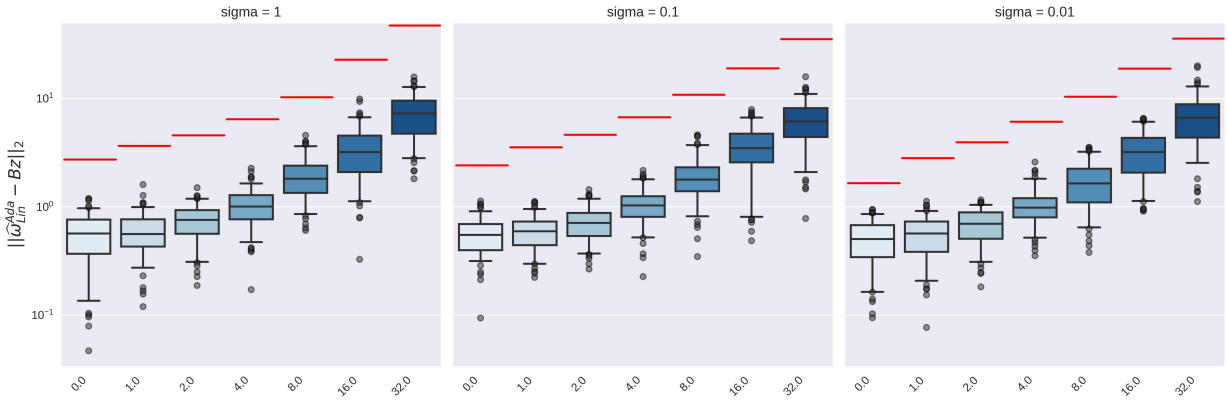


Figure 3.5: Evolution of the $\|\cdot\|_2$ error of the adaptive linear estimate as a function of parameter δ and noise σ

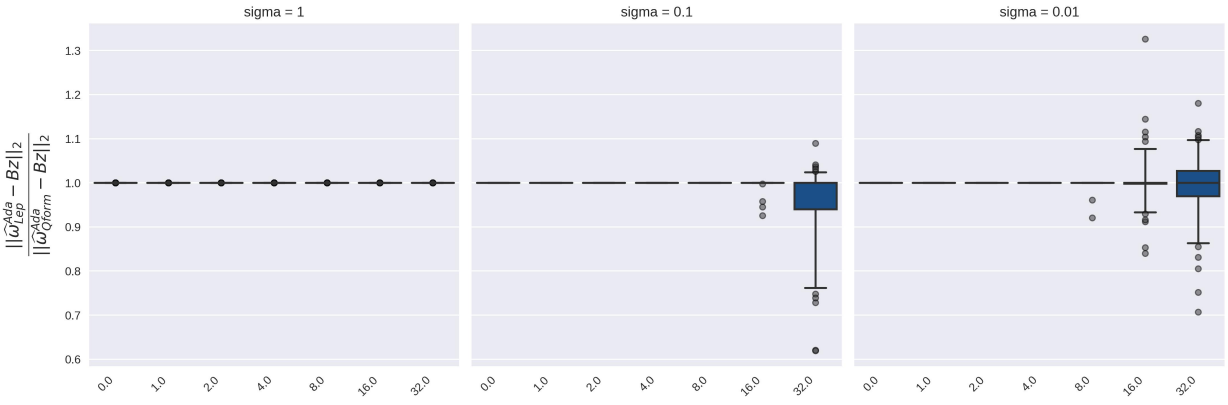


Figure 3.6: Evolution of $\|B(x+u) - \widehat{\omega}^{Ada}_{Lep}\|_2 / \|B(x+u) - \widehat{\omega}^{H_\delta}_{Qform}\|_2$ as a function of parameter δ and noise σ

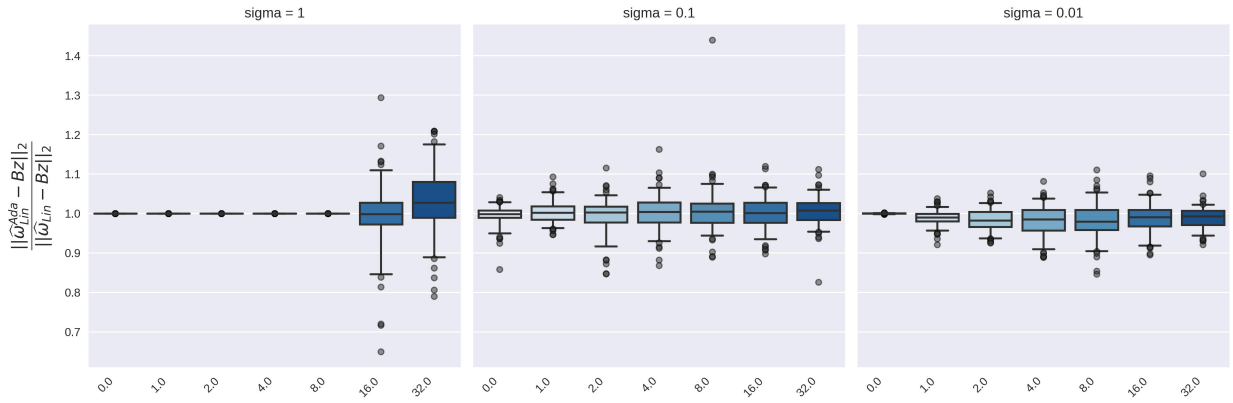


Figure 3.7: Evolution of $\|B(x+u) - \widehat{\omega}^{Ada}_{Lin}\|_2 / \|B(x+u) - \widehat{\omega}^{H_\delta}_{Lin}\|_2$ as a function of parameter δ and noise σ

The first take-away of 3.6 is that both estimates are roughtly equivalent in terms of their statistical performance. Indeed, for heavy noise, both routines select the same estimate. Where these two methods diverge is in the fact that one needs to compute $\mathcal{O}(I^3)$ upper bounds for quadratic form aggregation, whereas "only" $\mathcal{O}(I^2)$ are required for Lepski's procedure. Moreover, risk-upper bounds of contrasts tuned for the first are bigger than those for the latter, as they are respectively for $\frac{2\varepsilon}{I(I+1)}$-risk and $\frac{\varepsilon}{I+1}$-risk.

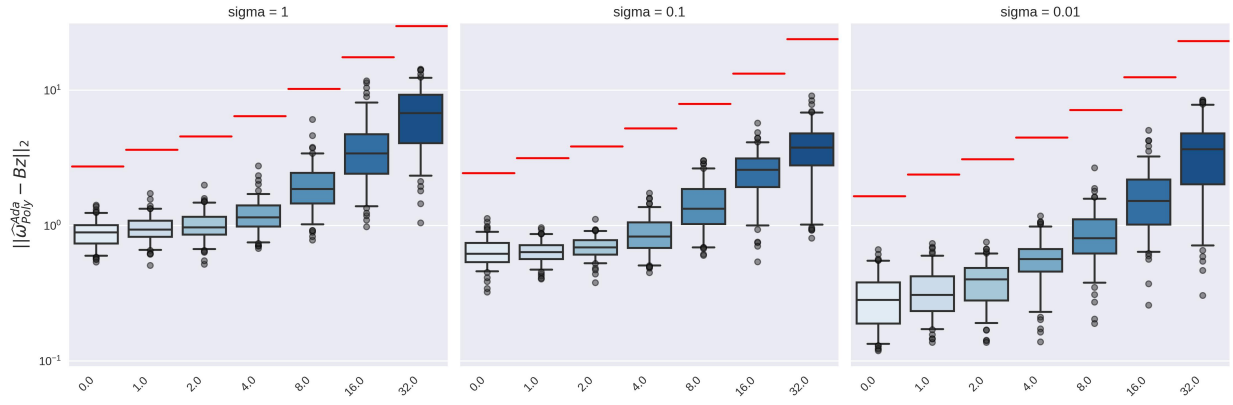


Figure 3.8: Evolution of the $\|\cdot\|_2$ error of the adaptive polyhedral estimate as a function of parameter δ and noise σ

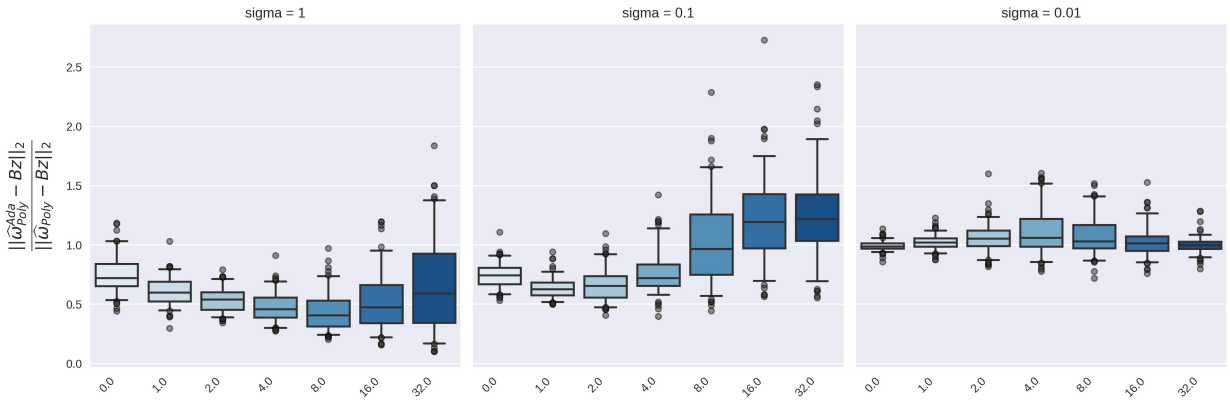


Figure 3.9: Evolution of $\|B(x+u) - \widehat{\omega}^{Ada}_{Poly}\|_2 / \|B(x+u) - \widehat{\omega}^{H_\delta}_{Poly}\|_2$ as a function of parameter δ and noise σ

Chapter 4

First order algorithms for computing linear and polyhedral estimates

Abstract

It was recently shown [30, 33] that "properly built" linear and polyhedral estimates nearly attain minimax accuracy bounds in the problem of recovery of unknown signal from noisy observations of linear images of the signal when the signal set is an *ellitope*. However, design of nearly optimal estimates relies upon solving semidefinite optimization problems with matrix variables, what puts the synthesis of such estimates beyond the rich of the standard Interior Point algorithms of semidefinite optimization even for moderate size recovery problems. Our goal is to develop First Order Optimization algorithms for the computationally efficient design of linear and polyhedral estimates. In this paper we (a) explain how to eliminate matrix variables, thus reducing dramatically the design dimension when passing from Interior Point to First Order optimization algorithms and (b) develop and analyse a dedicated algorithm of the latter type—Composite Truncated Level method.

4.1 Introduction

In this paper we discuss numerical algorithms for construction of "presumably good" estimates in linear inverse problems. Specifically, consider the estimation problem as follows. We are given an observation $\omega \in \mathbf{R}^m$,

$$\omega = Ax + \xi \tag{4.1}$$

where $\xi \in \mathbf{R}^m$ is zero mean random noise, $A \in \mathbf{R}^{m\times n}$ is observation matrix, and x is unknown signal known to belong to a given convex set $\mathcal{X} \subset \mathbf{R}^n$. Our objective is to recover the linear image $w = Bx$, $B \in \mathbf{R}^{\nu\times n}$, of x.

Our focus is on *linear* and *polyhedral estimates* for solving the problem in question.
When applied to the estimation problem above, *linear estimate* $\widehat{w}^H_{\rm lin}(\omega)$ of w is of the form $\widehat{w}^H_{\rm lin}(\omega) = H^T\omega$ where *contrast matrix* $H \in \mathbf{R}^{m\times\nu}$ is the estimate's parameter. A polyhedral estimate $\widehat{w}^H_{\rm poly}(\omega)$ is specified by a *contrast matrix* $H \in \mathbf{R}^{m\times M}$ according to

$$\omega \mapsto \widehat{x}^H(\omega) \in \operatorname*{Argmin}_{x\in\mathcal{X}} \left\{\|H^T(\omega - Ax)\|_\infty\right\}, \quad \widehat{w}^H_{\rm poly}(\omega) := B\widehat{x}^H(\omega).$$

Our interest in these two types of estimates stems from the fact that, as it was shown in [30, 32, 33], in the Gaussian case ($\xi \sim \mathcal{N}(0, \sigma^2 I_m)$), linear and polyhedral estimates with properly designed efficiently computable contrast matrices are near-minimax optimal in terms of their risks over a rather general class of loss functions and signal sets which we call *ellitopes*. [1] In this paper, our goal is to investigate numerical algorithms for design of near-optimal linear and polyhedral estimates. Specifically, we aim at developing numerical routines for efficient computation of contrast matrices H, the principal parameters of the estimates of both types.
As it was shown in [30, 32, 33], given the problem data—matrices A, B, the signal ellitope $\mathcal{X}$, and the *co-ellitopic norm* $\|\cdot\|$ in which estimation error is measured, computing the contrast matrices of linear and polyhedral estimates amounts to solving a well-structured convex optimization problem with linear objective and linear matrix inequality constraints. State-of-the-art optimization software, e.g., CVX [16] which relies upon Interior Point Semidefinite Programming (SDP) algorithms may be used to compute high-accuracy solutions to these problems. However, the structure of the

[1] Exact definitions of these sets are reproduced in the main body of the paper. For the time being, it suffices to point out an instructive example: a bounded intersections of finitely many sets of the form $\{x : \|Px\|_p \le 1\}$, $p \ge 2$, is an ellitope.

optimization problems in question (presence of "dense" matrix arguments) results in prohibitively long processing times by IPM algorithms already for rather moderate problem dimensions (signal dimension n in the range of few dozens). In this paper we discuss an alternative approach to solving the problem of designing linear and polyhedral estimates which rely upon first order algorithm, namely, the Composite Truncated Level method (CTL) of the bundle family. In particular, we show how matrix arguments can be eliminated from the contrast optimization problem and how the problem can be cast in the form amenable for first order algorithms.

The paper is organized as follows. In Section 4.2 we present the precise setting of the estimation problem and define optimization problems underlying the contrast design for linear and polyhedral estimates. To set up the first order optimization algorithm, we demonstrate how the problem of contrast computation for linear estimate can be reduced to that for the polyhedral one, and then explain in Section 4.3 how the latter problem can be rewritten in the form not involving matrix arguments and convenient for solving using first order algorithms. A small simulation study presented in Section 4.3.2 illustrates numerical performance of the proposed algorithms. We present the details of the Composite Truncated Level algorithm in Section 4.4 of the appendix. Proofs of technical statements are put to Section 4.5.

4.2 Linear and polyhedral estimates

4.2.1 Estimation problem

Consider the problem of recovering linear image $w = Bx$ of unknown signal $x \in \mathbf{R}^n$ from noisy observation

$$\omega = Ax + \sigma\xi \tag{4.2}$$

where $B \in \mathbf{R}^{\nu\times n}$ and $A \in \mathbf{R}^{m\times n}$ are given matrices, $\sigma > 0$ is known, and x is known to belong to a given signal set $\mathcal{X}$. Throughout the paper ξ is $(0, I_m)$-sub-Gaussian, denoted $\xi \sim \mathcal{SG}(0, I_m)$, i.e., for all $t \in \mathbf{R}^m$,

$$\mathbf{E}\left\{e^{t^T\xi}\right\} \le \exp\left(\tfrac{1}{2}\|t\|_2^2\right). \tag{4.3}$$

Given a norm $\|\cdot\|$ on $\mathbf{R}^\nu$ and reliability tolerance $\epsilon \in (0,1)$, we quantify the performance of a candidate estimate $\widehat{w}(\cdot)$ by its ϵ-*risk*

$$\mathrm{Risk}_\epsilon[\widehat{w}|\mathcal{X}] \quad = \sup_{x\in\mathcal{X}} \inf_\rho \left\{\rho : \mathrm{Prob}_\xi\left\{\|\widehat{w}(Ax+\xi) - Bx\| > \rho\right\} \le \epsilon\right\}. \tag{4.4}$$

We assume from now on that the signal set $\mathcal{X}$ and the *polar* $\mathcal{B}_*$ of the unit ball of $\|\cdot\|$ are basic ellitopes (see, e.g., [30] and [32, Section 4.2]). Specifically, we set

$$\begin{array}{rcl} \mathcal{X} & = & \{x \in \mathbf{R}^n : \exists t \in \mathcal{T} : x^T T_k x \le t_k, k \le K\}, \\ \mathcal{B}_* & = & \{y \in \mathbf{R}^\nu : \exists s \in \mathcal{S} : y^T S_\ell y \le s_\ell, \ell \le L\}. \end{array} \tag{4.5}$$

Here $T_k \succeq 0$ with $\sum_k T_k \succ 0$ (respectively, $S_\ell \succeq 0$ with $\sum_\ell S_\ell \succ 0$, and $\mathcal{T} \subset \mathbf{R}^K_+$ (respectively, $\mathcal{S} \subset \mathbf{R}^L_+$) is a convex compact set which is monotone (i.e., $0 \le t \le t' \in \mathcal{T} \Rightarrow t' \in \mathcal{T}$)[2] and possesses a nonempty interior. We refer to K (respectively, to L) as *ellitopic dimension* of $\mathcal{X}$ (respectively, of $\mathcal{B}_*$).

Every basic ellitope is a convex compact set with nonempty interior which is symmetric w.r.t. the origin. "Standard" examples of basic ellitopes are:

[2] Here and in the sequel, relationships $t \le s$ (or $t < s$) between $t, s \in \mathbf{R}^K$ are understood entrywise, i.e., as $t_i \le s_i$ (respectively, as $t_i < s_i$), $i = 1, ..., K$.

- A bounded intersection $\mathcal{X}$ of K centered at the origin ellipsoids/elliptic cylinders $\{x \in \mathbf{R}^n : x^T T_k x \le 1\}$ $[T_k \succeq 0]$:

$$\mathcal{X} = \{x \in \mathbf{R}^n : \exists t \in \mathcal{T} := [0,1]^K : x^T T_k x \le t_k,\ k \le K\}$$

  In particular, the unit box $\{x \in \mathbf{R}^n : \|x\|_\infty \le 1\}$ is a basic ellitope.

- A $\|\cdot\|_p$-ball in $\mathbf{R}^n$ with $p \in [2,\infty]$:

$$\{x \in \mathbf{R}^n : \|x\|_p \le 1\} = \big\{x : \exists t \in \mathcal{T} = \{t \in \mathbf{R}^n_+, \|t\|_{p/2} \le 1\} : \underbrace{x_k^2}_{x^T T_k x} \le t_k,\ k \le n\big\}.$$

4.2.2 The estimates

The interest in ellitopes in the present context is motivated by the results of [30, 32, 33] which state that, in the situation in question, "near-optimal"—with risks within logarithmic in K and L factors from the minimax risk—estimates can be found among *linear* and *polyhedral* ones.

Linear estimate

Linear estimate is specified by an $m \times \nu$ contrast matrix H according to

$$\widehat{w}_H(\omega) = H^T\omega.$$

Let

$$\mathfrak{r}_\varkappa[H] := \min_{\lambda,\mu,\Theta}\left\{\phi_{\mathcal{T}}(\mu) + \phi_{\mathcal{S}}(\lambda) + \sigma^2\varkappa^2\mathrm{Tr}(\Theta) :\ \lambda \ge 0, \mu \ge 0 \right. \tag{4.6}$$
$$\left.\left[\begin{array}{c|c|c} \sum_\ell \lambda_\ell S_\ell & \frac{1}{2}(B - H^T A) & \frac{1}{2}H^T \\ \hline \frac{1}{2}(B - H^T A)^T & \sum_k \mu_k T_k & \\ \hline \frac{1}{2}H & & \Theta \end{array}\right] \succeq 0\right\}$$

where for $\mathcal{G} \subset \mathbf{R}^p$

$$\phi_{\mathcal{G}}(z) = \sup_{g \in \mathcal{G}} z^T g :\ \mathbf{R}^p \to \mathbf{R} \cup \{+\infty\}$$

is the support function of $\mathcal{G}$.

Proposition 4.2.1 *(cf. [32, Proposition 4.14]) Let $\widehat{w}^H_{\rm lin}(\omega) = H^T\omega$ with some $H \in \mathbf{R}^{m\times\nu}$.*

(i) Then

$$\sup_{x \in \mathcal{X}} \mathbf{E}\left\{\|\widehat{w}(Ax+\xi)\widehat{w}^H_{\rm lin}(\omega) - Bx\|\right\} \le \mathfrak{r}_1[H]$$

(ii) Furthermore, let

$$\varkappa = 1 + \sqrt{2\ln[\epsilon^{-1}]}.$$

Then

$$\mathrm{Risk}_\epsilon[\widehat{w}^H_{\rm lin}|\mathcal{X}] \le \mathfrak{r}_\varkappa[H]. \tag{4.7}$$

Furthermore, function $\mathfrak{r}_\varkappa[H]$ is a convex, continuous and coercive function of the contrast matrix, and can be efficiently minimized w.r.t. H.

Remarks.

- As a consequence of the statement (i) of the proposition, the optimal value $\mathfrak{r}^*$ of the (clearly solvable) convex optimization problem

$$\mathfrak{r}^* := \min_H \mathfrak{r}_1[H] = \min_{H,\lambda,\mu,\Theta} \left\{ \phi_{\mathcal{T}}(\mu) + \phi_{\mathcal{S}}(\lambda) + \sigma^2 \mathrm{Tr}(\Theta) : \lambda \geq 0, \mu \geq 0 \right. \\ \left. \left[\begin{array}{c|c|c} \sum_\ell \lambda_\ell S_\ell & \frac{1}{2}(B - H^T A) & \frac{1}{2} H^T \\ \hline \frac{1}{2}(B - H^T A)^T & \sum_k \mu_k T_k & \\ \hline \frac{1}{2} H & & \Theta \end{array} \right] \succeq 0 \right\} \tag{4.8}$$

  is an upper bound on the expected risk

$$\mathrm{Risk}[\widehat{w}^*_{\mathrm{lin}} | \mathcal{X}] = \sup_{x \in \mathcal{X}} \mathbf{E} \left\{ \| \widehat{w}^*_{\mathrm{lin}}(Ax + \xi) - Bx \| \right\}$$

  of the estimate $\widehat{w}^*_{\mathrm{lin}}(\omega) = H_*^T \omega$ yielded by the H-component H_* of an optimal solution to the problem. Note that we do not need to assume that the noise ξ is sub-Gaussian for the above bound to hold: it would suffice to suppose that $\mathbf{E}\{\xi\xi^T\} \preceq \sigma^2 I$. Moreover (cf. [32, Proposition 4.16]), the value $\mathfrak{r}^*$ is within moderate factor of the minimax-optimal $\| \cdot \|$-risk

$$\mathrm{RiskOpt}[\mathcal{X}] = \inf_{\widehat{w}(\cdot)} \mathrm{Risk}[\widehat{w} | \mathcal{X}]$$

  (here inf is taken over all estimates, linear and non-linear alike):

$$\mathfrak{r}^* \leq O(1) \sqrt{\ln(K+1) \ln(L+1)} \mathrm{RiskOpt}[\mathcal{X}] \tag{4.9}$$

  (from now on, $O(1)$'s stands for an absolute constant).

- Observe that the ϵ-risk of the estimate $\widehat{w}^*_{\mathrm{lin}}$ may be evaluated using the statement (ii) of Proposition 4.2.1.

  Corollary 4.2.1 *Let $\epsilon \in (0,1)$. The ϵ-risk of the estimate $\widehat{w}^*_{\mathrm{lin}}(\omega)$ satisfies*

$$\mathrm{Risk}_\epsilon[\widehat{w}^*_{\mathrm{lin}} | \mathcal{X}] \leq \left(1 + \sqrt{2 \ln[\epsilon^{-1}]} \right) \mathfrak{r}^*.$$

- The bound $\overline{\psi}_\epsilon(\Theta) = \varkappa^2 \mathrm{Tr}(\Theta)$ for the $1 - \epsilon$ quantile of the quadratic form $\xi^T \Theta \xi$ in the expression (4.6) can be replaced by the following tighter but harder to process bounds (cf, e.g., [38, Proposition A.1])

$$\begin{array}{rl} & \psi_\epsilon(\Theta) := \min_\alpha \left\{ -\frac{\alpha}{2} \log \mathrm{Det}(I - 2\alpha^{-1}\Theta) + \alpha \ln[\epsilon^{-1}],\ \alpha \geq 2\lambda_{\max}(\Theta) \right\} \\ \leq & \psi'_\epsilon(\Theta) := \min_\alpha \left\{ \mathrm{Tr}(\Theta) + \mathrm{Tr}\left(\Theta(\alpha I - 2\Theta)^{-1} \Theta \right) + \alpha \ln(\epsilon^{-1}) : \alpha \geq 2\lambda_{\max}(\Theta) \right\} \\ \leq & \widetilde{\psi}_\epsilon(\Theta) := \mathrm{Tr}(\Theta) + 2\|\Theta\|_{\mathrm{Fro}} \sqrt{\ln[\epsilon^{-1}]} + 2\lambda_{\max}(\Theta) \ln[\epsilon^{-1}] \leq \overline{\psi}_\epsilon(\Theta). \end{array} \tag{4.10}$$

Polyhedral estimate

Polyhedral estimate is specified by $m \times m$ contrast matrix H satisfying

$$\sigma\chi\|\mathrm{Col}_j[H]\|_2 \le 1,\ 1 \le j \le m, \tag{4.11}$$

with

$$\chi = [\chi(\epsilon/m) =]\sqrt{2\ln[2\epsilon/m]};$$

in other words, H is normalized by the requirement that

$$\mathrm{Prob}_{\xi\sim\mathcal{SG}(0,I)}\{\sigma\|H^T\xi\|_\infty > 1\} \le \epsilon. \tag{4.12}$$

The associated with H polyhedral estimate $\widehat{w}^H_{\mathrm{poly}}(\omega)$ is

$$\widehat{w}^H_{\mathrm{poly}}(\omega) = B\overline{x}(\omega),\ \overline{x}(\omega) = \operatorname*{argmin}_x \left\{\|H^T(\omega - Ax)\|_\infty : x \in \mathcal{X}.\right\} \tag{4.13}$$

It is shown in [32, Setion 5.1.5] that the ϵ-risk of $\widehat{w}^H_{\mathrm{poly}}$ is upper-bounded by the quantity

$$\min_{\lambda,\mu,\upsilon}\left\{2\Big[\phi_{\mathcal{S}}(\lambda) + \phi_{\mathcal{T}}(\mu) + \sum_j \upsilon_j\Big] : \lambda \ge 0, \mu \ge 0, \upsilon \ge 0 \right. \tag{$P[H]$}$$
$$\left.\left[\begin{array}{c|c} \sum_\ell \lambda_\ell S_\ell & \frac{1}{2}B \\ \hline \frac{1}{2}B^T & A^TH\mathrm{Diag}\{\upsilon\}H^TA + \sum_k \mu_k T_k \end{array}\right] \succeq 0\right\}$$

A presumably good synthesis of the contrast H is yielded by feasible solutions to the convex optimization problem

$$\mathfrak{p}^*_\chi = \min_\Theta\left\{\mathfrak{p}_\chi[\Theta] := \min_{\lambda,\mu}\left\{2\big[\phi_{\mathcal{S}}(\lambda) + \phi_{\mathcal{T}}(\mu) + \sigma^2\chi^2(\epsilon/m)\mathrm{Tr}(\Theta)\big] : \lambda \ge 0,\ \mu \ge 0,\ \Theta \succeq 0\right.\right. \tag{4.14}$$
$$\left.\left.\left[\begin{array}{c|c} \sum_\ell \lambda_\ell S_\ell & \frac{1}{2}B \\ \hline \frac{1}{2}B^T & A^T\Theta A + \sum_k \mu_k T_k \end{array}\right] \succeq 0\right\}\right.$$

Given a feasible solution (λ, μ, Θ) to (4.14), we set $H = [\sigma\chi(\epsilon/m)]^{-1}U$ where $\Theta = U\mathrm{Diag}\{\nu\}U^T$ is the eigenvalue decomposition of Θ and $\upsilon = [\sigma\chi(\epsilon/m)]^2\nu$. Note that such H satisfies (4.11) and, moreover, $(\lambda, \mu, \upsilon, H)$ is a feasible solution to $(P[H])$ with values of respective objectives of both problems at these feasible solutions being equal to each other. As a result, the ϵ-risk of the polyhedral estimate $\widehat{w}_{\mathrm{poly}}$ stemming, in the just explained fashion, from an optimal solution to the (clearly solvable) problem (4.14) is upper-bounded by $\mathfrak{p}^*_\chi$. As shown in [32, Proposition 5.10], the resulting polyhedral estimate is nearly minimax-optimal:

$$\mathrm{Risk}_\epsilon[\widehat{w}_{\mathrm{poly}}|\mathcal{X}] \le \mathfrak{p}^*_\chi \le O(1)\sqrt{\ln(K+1)\ln(L+1)\ln(2m/\epsilon)}\,\mathrm{RiskOpt}_\epsilon[\mathcal{X}]$$

where $\mathrm{RiskOpt}_\epsilon[\mathcal{X}]$ is the minimax ϵ-risk.

4.2.3 From polyhedral to linear estimate and back

Observe that problems (4.8) and (4.14) responsible for the design of nearly minimax-optimal under the circumstances linear and polyhedral estimates are well structured convex problems. State-of-the-art Interior Point Semidefinite Programming (SDP) algorithms may be use to compute high-accuracy solutions to these problems in a wide range of geometries of $\mathcal{T}$ and $\mathcal{S}$. However, the

presence of matrix variables H and Θ results in large design dimensions of the SDP's to be solved and make prohibitively time consuming processing problem instances of sizes m, n in the range of hundreds. The first goal of this paper is to show that matrix variables may be eliminated from (4.8) and (4.14) allowing for processing by dedicated First Order algorithms, resulting in significant extension of the ranges of problem sizes amenable for numerical processing.

Our first observation is that problems (4.8) and (4.14) are "nearly reducible" to each other. Indeed, let $(\lambda, \mu, H, \Theta)$ be feasible to (4.8). We clearly have $\Theta \succeq 0$. Let $G = \left[\begin{array}{c|c|c} I & & \\ \hline & I & A^T \end{array}\right]$. By multiplying the semidefinite constraint of (4.8) by G on the left and G^T on the right we conclude that (λ, μ, Θ) is a feasible solution to (4.14) with the corresponding objective value

$$2[\phi_{\mathcal{S}}(\lambda) + \phi_{\mathcal{T}}(\mu) + \sigma^2\chi^2\mathrm{Tr}(\Theta)] = 2\mathfrak{r}_\chi[H].$$

The converse is also true.

Lemma 2 *Let (λ, μ, Θ) be a feasible solution to (4.14). Then it can be augmented to the feasible solution $(2\lambda, \mu, H, \Theta)$ of (4.8) with the corresponding objective value*

$$2\phi_{\mathcal{S}}(\lambda) + \phi_{\mathcal{T}}(\mu) + \sigma^2\varkappa^2\mathrm{Tr}(\Theta) \le \mathfrak{p}_\varkappa[\Theta].$$

4.3 Designing polyhedral estimates by a First Order method

According to the results from the previous section, when speaking about numerical design of linear and polyhedral estimates, we can focus solely on solving problem (4.14).[3] Next, projecting, if necessary, the observation ω onto the image space of A, we can assume w.l.o.g. that $m \le n$ and the image space of A is the entire $\mathbf{R}^m$. In fact, we make here a stronger assumption:[4]

AssO: Matrix $A \in \mathbf{R}^{n\times n}$ in (4.2) is nonsingular.

Under this assumption, we can carry out partial minimization in Θ in (4.14). Specifically, it is immediately seen that (4.14) is equivalent to the optimization problem

$$\min_{\lambda,\mu,\Theta}\Big\{\phi_{\mathcal{S}}(\lambda) + \phi_{\mathcal{T}}(\mu) + \sigma^2\chi^2\mathrm{Tr}(\Theta) : \lambda > 0,\ \mu \ge 0,\ \Theta \succeq 0 \qquad (4.15)$$
$$\Theta \succeq \underbrace{A^{-T}\left[\tfrac{1}{4}B^T\left[\textstyle\sum_\ell \lambda_\ell S_\ell\right]^{-1}B - \sum_k \mu_k T_k\right]A^{-1}}_{\mathfrak{T}(\lambda,\mu)}\Big\}$$

In the latter problem partial minimization in Θ is as follows: given $\lambda > 0$ and $\mu \ge 0$ we compute the eigenvalue decomposition

$$\mathfrak{T}(\lambda, \mu) = U\mathrm{Diag}\{\upsilon\}U^T$$

of $\mathfrak{T}(\lambda, \mu)$. The best in terms of the objective of (4.15) choice of Θ given λ, μ is

$$\Theta = U\mathrm{Diag}\{\upsilon^+\}U^T \qquad [\alpha^+ = \max[\alpha, 0]].$$

[3] Strictly speaking, this is so if we assume that when looking for a linear estimate, we are ready to tolerate a moderate constant factor (namely, 2) in the risk bound of the resulting estimate.

[4] We briefly describe the "conversion" of (4.14) into the form amenable for First Order algorithms in the case of singular A in Section 4.6.

Therefore, (4.15) reduces to

$$\min_{\lambda>0,\mu\geq 0}\left\{\Upsilon(\lambda,\mu):=\phi_{\mathcal{S}}(\lambda)+\phi_{\mathcal{T}}(\mu)+\sigma^2\chi^2\sum_i\lambda_i^+(\mathfrak{T}(\lambda,\mu))\right\} \tag{4.16}$$

$$\mathfrak{T}(\lambda,\mu)=A^{-T}\left[\tfrac{1}{4}B^T\Big[\sum_\ell\lambda_\ell S_\ell\Big]^{-1}B-\sum_k\mu_k T_k\right]A^{-1}$$

where $\lambda_1(Q)\geq\lambda_2(Q)\geq...\geq\lambda_p(Q)$ are the eigenvalues of symmetric $p\times p$ matrix Q. and $\lambda_i^+(Q)=\max[\lambda_i(Q),0]$.

4.3.1 Setting up Composite Truncated Level algorithm

We intend to solve the problem of interest (4.14) by applying to (4.16) a First Order algorithm—Composite Truncated Level algorithm (CTL). Detailed description of the method is presented in Section 4.4. CTL is aimed at solving convex optimization problems of the form

$$\mathrm{Opt}=\min_{x\in X}\{\phi(x):=\psi(x)+f(x)\} \tag{4.17}$$

where $X\subset\mathbf{R}^N$ is a nonempty bounded and closed convex set, and $\psi(\cdot)$ and $f(\cdot)$ are Lipschitz continuous convex functions on X with "simple" X and ψ (for details, see Section 4.4). Note that in order to reduce the problem of interest (4.16) to the form (4.17), it suffices to set

$$\begin{aligned}
&X=\left\{x=[\lambda;\mu]\in\mathbf{R}_+^{L+K},\ \lambda_\ell\geq\delta\,\forall\ell,\ \sum_\ell\lambda_\ell+\sum_k\mu_k\leq R\right\},\\
&\psi([\lambda;\mu])=\phi_{\mathcal{S}}(\lambda)+\phi_{\mathcal{T}}(\mu),\\
&f([\lambda;\mu])=\sigma^2\chi^2\sum_i\lambda_i^+(\mathfrak{T}(\lambda,\mu)).
\end{aligned} \tag{4.18}$$

When solving (4.17), CTL "learns" the difficult part $f(x)$ of the objective via *oracle* which, given on input a query point $\overline{x}\in X$, returns a "simple" Lipschitz continuous convex function (model) $f_{\overline{x}}(\cdot)$ such that

$$f_{\overline{x}}(\overline{x})=f(\overline{x})\ \ \&\ \ f_{\overline{x}}(y)\leq f(y)\quad\forall y\in X.$$

Oracle $\mathcal{O}_\varrho$. In the situation we are interested in with the data for (4.17) given by (4.18), the oracle may be built as follows:

1. Given query point $\overline{x}=[\overline{\lambda};\overline{\mu}]\in X$, we compute the matrices $\overline{\Lambda}=\sum_\ell\overline{\lambda}_\ell S_\ell$ and $\overline{\mathfrak{T}}=\mathfrak{T}(\overline{\lambda},\overline{\mu})$ along with the eigenvalue decomposition $\overline{\mathfrak{T}}=\overline{U}\mathrm{Diag}\{\overline{v}\}\overline{U}^T$, $\overline{v}_1\geq\overline{v}_2\geq...\geq\overline{v}_n$ of $\overline{\mathfrak{T}}$.

2. We put

$$T(\lambda,\mu)=A^{-T}\left[\tfrac{1}{4}B^T\overline{\Lambda}^{-1}\Big[2\overline{\Lambda}-\sum\nolimits_\ell\lambda_\ell S_\ell\Big]\overline{\Lambda}^{-1}B-\sum\nolimits_k\mu_k T_k\right]A^{-1}$$

   This function is obtained from $\mathfrak{T}(\lambda,\mu)$ by linearization in λ at $\lambda=\overline{\lambda}$ and clearly $\succeq$-underestimates $\mathfrak{T}(\lambda,\mu)$ on X, while $T(\overline{\lambda},\mu)\equiv\mathfrak{T}(\overline{\lambda},\mu)$. Consequently,

$$f([\lambda;\mu])=\sigma^2\chi^2\sum_i\lambda_i^+(\mathfrak{T}(\lambda,\mu))\geq f_{\overline{\lambda}}(\lambda,\mu):=\sigma^2\chi^2\sum_i\lambda_i^+(T(\lambda,\mu)),$$

   the inequality being equality when $\lambda=\overline{\lambda}$.

Recall that for every symmetric convex function g on $\mathbf{R}^n$ and every $n \times n$ symmetric matrix D one has [13, Proposition 4.2.1]

$$g([D_{11}; ...; D_{nn}]) \le g(\lambda(D)).$$

When specifying $g(s) = \sum_{i=1}^n s_i^+$ and denoting by $D_i(\lambda, \mu)$ the diagonal entries in the matrix $\overline{U}^T T(\lambda, \mu)\overline{U}$, $1 \le i \le n$, we get

$$[f([\lambda; \mu]) \ge] \, f_{\overline{\lambda}}(\lambda, \mu) \ge \sum_{i=1}^n D_i^+(\lambda, \mu) \quad \forall(\lambda > 0, \mu \ge 0)$$

with both inequalities becoming equalities for $\lambda = \overline{\lambda}$ and $\mu = \overline{\mu}$. Taking into account that functions $D_i(\lambda, \mu)$ are affine, we conclude that *piecewise linear function* $\sum_i D_i^+(\lambda, \mu)$ *underestimates* $f([\lambda; \mu])$ *in the domain* $\lambda > 0$ *and is equal to* $f([\lambda; \mu])$ *when* $(\lambda, \mu) = (\overline{\lambda}, \overline{\mu})$.

3. The above considerations justify the oracle $\mathcal{O}_\varrho$ defined as follows:

$$f_{\overline{x}=[\overline{\lambda};\overline{\mu}]}(x) = \sum_{\iota=1}^{\varrho} \max[\alpha_\iota(x), 0],$$

where ϱ, $1 \le \varrho \le n$, is "complexity parameter" of the oracle, and $\alpha_\iota(x)$ are affine functions of $x = [\lambda; \mu]$ specified according to

- for $\iota < \varrho$, $\alpha_\iota(x) = D_\iota(x)$;
- $\alpha_\varrho(x) = \sum_{\iota \ge \varrho} \overline{D}_\iota(x)$, $\overline{D}_\iota(x) = \begin{cases} D_\iota(x), & D_\iota(\overline{x}) \ge 0, \\ 0, & \text{otherwise.} \end{cases}$

By construction, $f_{\overline{x}}(x)$ is the sum of ϱ positive parts of linear forms, underestimates $f(x)$ everywhere, and coincides with $f(x)$ when $x = \overline{x}$.

4.3.2 Numerical illustration

Consider the situation in which $\|\cdot\|$ is $\|\cdot\|_2$. In this case problem (4.14) reads

$$\mathfrak{p}_\chi^* = \min_{\lambda,\mu,\Theta} \left\{ 2[\lambda + \phi_{\mathcal{T}}(\mu) + \sigma^2\chi^2 \mathrm{Tr}(\Theta)] : \begin{array}{c} \lambda \ge 0, \mu \ge 0, \Theta \succeq 0 \\ \left[\begin{array}{c|c} \lambda I_\nu & \frac{1}{2}B \\ \hline \frac{1}{2}B^T & A^T\Theta A + \sum_k \mu_k T_k \end{array}\right] \succeq 0 \end{array} \right\} \tag{4.19}$$

Observe that scaling a feasible solution (λ, μ, Θ) to the problem according to $(\lambda, \mu, \Theta) \mapsto (s\lambda, s^{-1}\mu, s^{-1}\Theta)$ with $s > 0$ preserves feasibility; the best in terms of the objective scaling of (λ, μ, Θ) corresponds to $s = \sqrt{[\phi_{\mathcal{T}}(\mu) + \sigma^2\chi^2\mathrm{Tr}(\Theta)]/\lambda}$ and results in the value of the objective $4\sqrt{\lambda[\phi_{\mathcal{T}}(\mu) + \sigma^2\chi^2\mathrm{Tr}(\Theta)]}$, As a result, we can eliminate the variable λ, thus arriving at the problem

$$\overline{\mathfrak{p}}_\chi^* = \min_{\overline{\mu},\overline{\Theta}} \left\{ F(\overline{\mu}, \overline{\Theta}) := \phi_{\mathcal{T}}(\overline{\mu}) + \sigma^2\chi^2\mathrm{Tr}(\overline{\Theta}) : \begin{array}{c} \overline{\mu} \ge 0, \overline{\Theta} \succeq 0 \\ \left[\begin{array}{c|c} I_\nu & \frac{1}{2}B \\ \hline \frac{1}{2}B^T & A^T\overline{\Theta}A + \sum_k \overline{\mu}_k T_k \end{array}\right] \succeq 0 \end{array} \right\} \tag{4.20}$$

A feasible (an optimal) solution $\overline{\mu}, \overline{\Theta}$ to (4.20) gives rise to feasible (resp., optimal) solution $\lambda = \sqrt{F(\overline{\mu}, \overline{\Theta})}$, $\mu = \overline{\mu}/\lambda$, $\Theta = \overline{\Theta}/\lambda$ to (4.19) with the value of the objective equal to $4\sqrt{F(\overline{\mu}, \overline{\Theta})}$; in particular,

$$\mathfrak{p}_\chi^* = 4\sqrt{\overline{\mathfrak{p}}_\chi^*}.$$

In our experiments, we used $B = I_n$, $A \in \mathbf{R}^{n\times n}$ with i.i.d. entries drawn from $\mathcal{N}(0,1)$, we put $\sigma = 0.1$, $\epsilon = 0.05$, and used ellitopic signal set

$$\mathcal{X} = \{x \in \mathbf{R}^n : \sum_{i\in I_k} i\, x_i^2 \le 1,\ 1 \le i \le K\}$$

where $I_1, ..., I_K$ are consecutive segments of the range $1 \le i \le n$ of cardinality n/K each .

Under the circumstances, problem (4.20) reads

$$\overline{\mathfrak{p}}^*_\chi = \min_{\overline{\mu},\overline{\Theta}} \left\{ \sum_{k=1}^K \overline{\mu}_k + \underbrace{\sigma^2\chi^2(\epsilon/n)}_{\gamma} \mathrm{Tr}(\overline{\Theta}) : \begin{array}{c} \overline{\mu} \ge 0, \overline{\Theta} \succeq 0 \\ \left[\begin{array}{c|c} I_\nu & \frac{1}{2}I_n \\ \hline \frac{1}{2}I_n & A^T\overline{\Theta}A + D[\overline{\mu}] \end{array}\right] \succeq 0 \end{array} \right\} \tag{4.21}$$

where $D[\overline{\mu}]$ is diagonal $n \times n$ matrix with the i-th diagonal entry equal to μ_k when $i \in I_k$.

After eliminating Θ by partial minimization (4.21) becomes

$$\min_{\overline{\mu}} \left\{ \sum_k \overline{\mu}_k + \gamma \sum_\iota \lambda^+_\iota(A^{-T}\left[I_n - D[\overline{\mu}]\right]A^{-1}) : 0 \le \overline{\mu}, \sum_k \overline{\mu}_k \le R \right\} \tag{4.22}$$

(we have imposed a large enough upper bound on $\sum_k \overline{\mu}_k$ to make the optimization domain bounded). The resulting problem was processed by the CTL algorithm utilizing oracle $\mathcal{O}_\varrho$. ϱ was the first of the two control parameters used the experiments; the second parameter was the maximum cardinality τ of bundle allowed for CTL.[5] We used "ℓ_1/ℓ_2 proximal setup," [85, Section 2.1], in which

$$\|\cdot\| = \|\cdot\|_1,\ \ \omega(\overline{\mu}) = \kappa_n\|\mu\|^2_{p_n},\ \ p_n = 1 + 1/\ln n,$$

where κ_n is an easy to compute constant ensuring strong convexity, modulus 1, of $\omega(\cdot)$ w.r.t. $\|\cdot\|_1$(see [85, Theorem 2.1]). The CTL parameters λ_ℓ, $\overline{\theta}$, $\underline{\theta}$ were set to 1/2, and the auxiliary problems (steps 4.2-4.3) were processed by Interior Point solver `Mosek` invoked via `CVX`, see [16]. When solving (4.22), computations were terminated when the best found so far value of the objective were within the factor 1.1 of the generated by the method lower bound on the optimal value.

We report on results of our experiments in Tables 4.1 and 4.2.[6] To put these results in proper perspective, note that solving (4.21) by state-of-the-art `Mosek` Interior Point solver takes 35 sec when $n = 64$, $K = 8$ and 1785 sec when $n = 128$, $K = 16$; as applied to the same problem with $n = 256$, $K = 64$, `Mosek` runs out of memory.

4.4 CTL—Composite Truncated Level algorithm

4.4.1 Situation and goal.

CTL is a First Order method for solving optimization problems

$$\mathrm{Opt} = \min_{x\in X} \phi(x) := \psi(x) + f(x), \tag{4.17}$$

where

[5] For description of CTL and related entities, see Section 4.4.

[6] MATLAB code for this experiment is available at GitHub repository `https://github.com/ai1-fr/Algorithms-for-linear-and-polyhedral-estimates/tree/main`.

n	64	128	256	512	1024	1024
K	8	16	32	64	128	1024
calls	23	8	10	28	24	26
phases	11	4	5	14	11	11
CPU, sec	7	2	4	20	96	781
0.05-risk	3.215	4.283	4.519	4.169	4.216	7.868
$\|\cdot\|_2$-risk	1.709	2.237	2.389	2.209	2.231	4.271

Table 4.1: Solving (4.22) by CTL with $\varrho = 10, \tau = 10$.

	ϱ=1	ϱ=10
τ=1	50/11/275	31/12/136
τ=10	26/11/109	24/11/96

Table 4.2: Performance of CTL vs. ϱ and τ, problem (4.22) with $n = 1024, K = 128$. Data in cells: # of calls/# of phases/CPU time, sec.

- $X \subset \mathbf{R}^n$ is nonempty, convex, closed, and bounded
- $\psi : X \to \mathbf{R}$ and $f : X \to \mathbf{R}$ are Lipschitz continuous convex functions.

Our assumptions are as follows:
Ass1: X is equipped with a *proximal setup* composed of a norm $\|\cdot\|$ on $\mathbf{R}^n$ and a *distance-generating function* $\omega : X \to \mathbf{R}$ which is continuously differentiable and strongly convex on X with convexity modulus 1 w.r.t. $\|\cdot\|$:

$$\langle \nabla\omega(x) - \nabla\omega(y), x - y\rangle \geq \|x - y\|^2 \quad \forall x, y \in X.$$

Proximal setup induces *Bregman distance* $V_x(y)$ on X and *Bregman diameter* Ω of X:

$$\begin{array}{rcl} V_x(y) & = & \omega(y) - [\omega(x) - \langle \nabla\omega(x), y - x\rangle] \geq \frac{1}{2}\|x - y\|^2,\ x, y \in X \\ \Omega & = & \left[2 \max\limits_{x,y\in X} V_x(y)\right]^{1/2} \geq \max\limits_{x,y\in X} \|x - y\|. \end{array}$$

Ass2: We have at our disposal *oracle* $\mathcal{O}$ which, given on input a query point $x \in X$, returns a *piece*—a Lipschitz continuous on X convex function

$$\phi_x(\cdot) = \psi(\cdot) + f_x(\cdot) : X \to \mathbf{R}$$

where $f_x(\cdot)$ belongs to some family $\mathcal{F}$ of "simple" Lipschitz continuous convex functions on X. We suppose that

$$\forall (x, y \in X) : \ f_x(y) \leq f(y) \ \ \& \ \ f_x(x) = f(x)$$

(in particular, $\phi(x) = \phi_x(x)$) and that functions $\phi_x(\cdot) : X \to \mathbf{R}$ are uniformly in $x \in X$ Lipschitz continuous on X:

$$|\phi_x(u) - \phi_x(v)| \leq L_\phi \|u - v\| \ \forall (u, v \in X) \tag{4.23}$$

for some $L_\phi < \infty$.

The simplest example of such oracle is that of family $\mathcal{F}$ comprised of affine functions of $\mathbf{R}^n$, and $f_x(y) = f(x) + \langle f'(x), y - x\rangle$ where $f'(x)$ is a subgradient of f at x ("first order oracle"). In a less trivial example, $f(x)$ is the largest eigenvalue of a symmetric matrix $S(x)$ which affinely depends on x, while $f_x(y) = \max_{i\le k} e_i^T S(y) e_i$, where $k \le n$ is fixed and $e_1, ..., e_k$ are the k leading eigenvectors of $S(x)$.

Ass3: We assume that for some positive integer τ we are able to solve efficiently problems of the form

$$\min_y \left\{ \psi(y) + \max_{\iota\le\tau} f_{x_\iota}(y) : y \in X, \alpha(y) \le 0 \right\}$$

and

$$\min_y \left\{ V_x(y) : y \in X, \psi(y) + \max_{\iota\le\tau} f_{x_\iota}(y) \le \ell, \alpha(y) \le 0 \right\}$$

where $\alpha(\cdot)$ is an affine function.

We remark that the algorithm to follow is the composite version of Non-Euclidean Restricted Memory Level method [86] operating with $\psi \equiv 0$ and the family of affine functions in the role of $\mathcal{F}$; the Euclidean version of the latter algorithm is the minimization version of the Proximal level bundle method from [87].

The algorithm

The description of CTL is as follows.

0. Control parameters of the algorithms are $\lambda_\ell \in (0,1), \overline{\theta} \in (0.1), \underline{\theta} \in (0,1)$.

1. At the beginning of an iteration of CTL, we have at our disposal

1.1) *upper bound* $\overline{\phi}$ on Opt—the best (the smallest) of the values of ϕ observed at the query points processed so far. These upper bounds do not increase as the iteration count grows. The query point $\widehat{y}$ with $\overline{\phi} = \phi(\widehat{y})$ is considered as the approximate solution generated so far.

1.2) *lower bound* $\underline{\phi}$ on Opt; these lower bounds do not decrease as the iteration number grows

1.3) *prox-center* $\overline{x} \in X$ and *level* $\ell \in (\underline{\phi}, \overline{\phi})$

1.4) *query point* $x \in X$

1.5) *bundle* – a nonempty collection $\mathcal{B}$ of $\tau_{\mathcal{B}} \le \tau$ pieces $\phi_\iota(\cdot) = \psi(\cdot) + f^\iota(\cdot)\}$, $1 \le \iota \le \tau_{\mathcal{B}}$, with $f^\iota \in \mathcal{F}$; positive integer τ is a parameter of the algorithm.

• The very first iteration is preceded by *initialization* where we call the oracle at a (whatever) point $x_{\rm ini} \in X$ and set

$$\overline{\phi} = \phi(x_{\rm ini}),\ \underline{\phi} = \min_{x\in X}[\psi(x) + f_{x_{\rm ini}}(x)],\ \mathcal{B} = \{\psi(x) + f_{x_{\rm ini}}(x)\}$$

Note that since the pieces reported by the oracle underestimate $\phi(\cdot)$, we do ensure $\underline{\phi} \le$ Opt.

2. Iterations are split into consecutive *phases*, with prox-center and level common to all iterations of a phase. For a particular phase,

2.1) the prox-center $\overline{x}$ is selected at the beginning of the first iteration of the phase and can be a whatever point of X;

2.2) the query point of the first iteration of the phase is $x = \overline{x}$,

2.3) the level ℓ is selected at the beginning of the very first iteration of the phase as

$$\ell = \lambda_\ell \underline{\phi} + (1 - \lambda_\ell)\overline{\phi}.$$

• At the beginning of the first iteration of a phase, we set

$$\overline{\Delta} = \overline{\phi} - \ell,\ \underline{\Delta} = \ell - \underline{\phi},\ \Delta = \overline{\Delta} + \underline{\Delta} = \overline{\phi} - \underline{\phi}.$$

Note that *the gap* Δ of a phase upper-bounds the inaccuracy in terms of the objective of the approximate solution available at the beginning of the phase.

3. At iterations of a phase, we maintain the relation

$$\phi(y) \geq \ell \ \text{ for all } \ y \in X \ \text{ such that } \ \langle \nabla\omega(x) - \nabla\omega(\overline{x}), y - x\rangle < 0 \tag{4.24}$$

where $\overline{x}, \ell$ are the prox-center and the level of the phase, and x is the query point of the iteration. Note that this relation takes place at the very first iteration of a phase, since for such an iteration $x = \overline{x}$.

4. An iteration of a phase is organized as follows:

4.1) We call the oracle at the query point x of the iteration, thus getting the value of the objective $\phi(x)$ and a piece $\phi_x(\cdot)$. After $\phi(x)$ is known, we

— update the upper bound $\overline{\phi}$ and the approximate solution $\widehat{x}$:

$$(\overline{\phi}, \widehat{x}) = \begin{cases} (\phi(x), x), & \text{if } \phi(x) < \overline{\phi} \\ (\overline{\phi}, \widehat{x}), & \text{otherwise.} \end{cases}$$

— update the bundle $\mathcal{B}$ by adding to it the piece $\phi_x(\cdot)$ and removing, if necessary, one of "old" pieces to keep the number of pieces $\phi_1, ..., \phi_{\tau_{\mathcal{B}}}$ in the resulting bundle to be at most τ.

4.2) If $\phi(x) - \ell \leq \overline{\theta}\,\overline{\Delta}$ ("essential progress in upper bound on Opt"), we terminate the phase and pass to the next one. Otherwise we solve the auxiliary problem

$$\widetilde{\phi} = \min_y \left\{ \widehat{\phi}(y) := \max_{1 \leq \iota \leq \tau_{\mathcal{B}}} \phi_\iota(y) : y \in X, \langle \nabla\omega(x) - \nabla\omega(\overline{x}), y - x\rangle \geq 0 \right\} \tag{4.25}$$

(as usual, $\widetilde{\phi} = +\infty$ when the right hand side problem is infeasible), and update the lower bound $\underline{\phi}$ on Opt according to

$$\underline{\phi} \mapsto \max\left[\underline{\phi}, \min[\widetilde{\phi}, \ell]\right]$$

Note: by (4.24), $\phi(y) \geq \ell$ at every point $y \in X$ which is not feasible for the optimization problem in (4.25). Besides this, the pieces $\phi_i(\cdot)$ in the bundle, and, consequently, the *model* $\widehat{\phi}(\cdot)$, underestimate $\phi(\cdot)$ on X. As a result, the quantity $\min[\widetilde{\phi}, \ell]$, and consequently the new value of $\underline{\phi}$, indeed is a lower bound on Opt.

If updated $\underline{\phi}$ satisfies

$$\ell - \underline{\phi} \leq \underline{\theta}\,\underline{\Delta}$$

("essential progress in lower bound on Opt"), we terminate the phase and pass to the next one.

4.3) If the iteration in question does not result in phase change, we solve the auxiliary problem

$$\min_y \left\{ V_{\overline{x}}(y) : \ y \in X, \ \widehat{\phi}(y) \leq \ell, \ \langle \nabla\omega(x) - \nabla\omega(\overline{x}), \ y - x\rangle \geq 0 \right\} \tag{4.26}$$

take its optimal solution, x_+, as the new query point, and pass to the next iteration of the phase. **Note:** when we need to solve (4.26), we have, by construction, $\widetilde{\phi} \leq \ell$, so that the problem in (4.26) is feasible, a feasible solution being a minimizer in (4.25). Thus, the new query point x_+ is well defined. Besides this, from the definition of x_+ it follows that

$$\langle \nabla\omega(x_+) - \nabla\omega(\overline{x}), y - x_+\rangle \geq 0$$

for every feasible solution y to (4.26). As a result, when $y \in X$ satisfies the relation

$$\langle \nabla\omega(x_+) - \nabla\omega(\overline{x}), y - x_+\rangle < 0,$$

y is infeasible for (4.26), meaning that either $\widehat{\phi}(y) > \ell$, and in such case $\phi(y) \geq \widehat{\phi}(y) \geq \ell$, or y satisfies the premise in (4.24), implying that $\phi(y) \geq \ell$ by (4.24). We see that (4.24) holds true when x is replaced with x_+, that is, (4.24) is maintained during the iterations.

Convergence analysis

Observe that

> (!) *If a phase is finite, then the gap Δ_+ of the subsequent phase does not exceed a fixed fraction $\theta\Delta$ of the gap Δ of the phase in question, where*
>
> $$\theta = \max\left[1 - \lambda_\ell\overline{\theta}, \underline{\theta} + \lambda_\ell(1 - \underline{\theta})\right] \in (0, 1);$$

Indeed, (!) is an immediate consequence of the phase termination rules in **4.2** combined with the facts that $\underline{\phi}$ does not decrease, and $\overline{\phi}$ does not increase as the iteration count grows.
The following observation is crucial:

> (!!) *The number of iterations at a phase with gap Δ does not exceed*
>
> $$\left\lceil \left(\frac{L_\phi\Omega}{\overline{\theta}\lambda_\ell\Delta}\right)^2 \right\rceil \tag{4.27}$$
>
> (here $\lceil a\rceil$ stands for the upper integer part of a—the smallest integer greater or equal to a).

Indeed, let ℓ be the level of the phase. Assume that the phase contains more that $T \geq 1$ iterations, so that the upper bound $\overline{\phi}$, the lower bound $\underline{\phi}$ on Opt, same as the model $\phi_t(\cdot)$ generated at iteration t of the phase are well defined for $t = 1, ..., T$, and the query points x_t are well defined for $t = 1, ..., T + 1$. By construction, for $1 \leq t \leq T$ we have

$$\phi_t(x_t) > \ell + \overline{\theta}\,\overline{\Delta}, \tag{4.28a}$$
$$\phi_t(x_{t+1}) \leq \ell, \tag{4.28b}$$
$$\langle \nabla V_{\overline{x}}(x_t), x_{t+1} - x_t\rangle \geq 0 \tag{4.28c}$$

where $\overline{x} = x_1$ is the prox-center of the phase. Indeed, when $t \leq T$,
— (4.28a) holds true since otherwise the phase would be terminated at its t-th iteration due to essential progress in upper bound on Opt, which is not the case when $t \leq T$;
— (4.28b) and (4.28c) hold because, by construction of x_{t+1} at a non-concluding iteration t of a phase, x_{t+1} minimizes continuously differentiable on X convex function $V_{\overline{x}}(\cdot)$ over the set

$$X \cap \{y \in X : \phi_t(y) \leq \ell\} \cap \{y \in Y : \langle \nabla V_{\overline{x}}(x_t), y - x_t\rangle \geq 0\}.$$

Now note that by construction of $\phi_t(\cdot)$ and due to Assumption Ass2, ϕ_t is Lipschitz continuous with constant L_ϕ w.r.t. $\|\cdot\|$, which combines with (4.28a) and (4.28b) to imply that

$$\|x_t - x_{t+1}\| > L_\phi^{-1}\overline{\theta}\cdot\overline{\Delta} = L_\phi^{-1}\overline{\theta}\lambda_\ell\Delta.$$

The latter relation, in turn, combines with (4.28c) and with inherited from $\omega(\cdot)$ strong convexity of $V_{\overline{x}}(\cdot)$ on X w.r.t. $\|\cdot\|$ to imply that

$$V_{\overline{x}}(x_{t+1}) > V_{\overline{x}}(x_t) + \frac{[\overline{\theta}\lambda_\ell]^2\Delta^2}{2L_\phi^2},\ 1 \le t \le T.$$

Taking into account that $V_{\overline{x}}(x_1) = 0$ and $V_{\overline{x}}(x) \le \frac{1}{2}\Omega^2$ for all $x \in X$, we arrive at (4.27).

As an immediate consequence of (!) and (!!), we get the following efficiency estimate:

Proposition 4.4.1 *For every $\epsilon \in (0, L_\phi\Omega)$, the overall number of CTL iterations until an ϵ-optimal, as certified by current gap, solution to the minimization problem is built does not exceed*

$$N(\epsilon) = C(L_\phi\Omega/\epsilon)^2$$

with C depending solely on the control parameters λ_ℓ, $\overline{\theta}$, and $\underline{\theta}$.

4.5 Proofs for Section 4.2

4.5.1 Proof of Proposition 4.2.1

To save notation, in this section we use shortcut notation $\widehat{w}_H$ for the linear estimate $\widehat{w}^H_{\rm lin}(Ax+\xi)$. Note that for all $x \in \mathcal{X}$, the loss $\|\widehat{w}_H - Bx\|$ of the estimate $\widehat{w}_H$ satisfies

$$\begin{aligned}
\|\widehat{w}_H - Bx\| &= \|(H^TA - B)x + \sigma H^T\xi\| = \max_{u\in\mathcal{B}_*}\left\{u^T[(H^TA-B)x + \sigma H^T\xi]\right\}\\
&= \max_{u\in\mathcal{B}_*}\left\{[u;x;\sigma\xi]^T\left[\begin{array}{c|c|c} & \frac{1}{2}(B-H^TA) & \frac{1}{2}H^T\\ \hline \frac{1}{2}(B-H^TA)^T & & \\ \hline \frac{1}{2}H & & \end{array}\right][u;x;\sigma\xi]\right\}\\
&\le \max_{u\in\mathcal{B}_*,x\in\mathcal{X}}\left\{u^T\left[\sum_\ell \lambda_\ell S_\ell\right]u + x^T\left[\sum_k \mu_k T_k\right]x + \sigma^2\xi^T\Theta\xi\right\}
\end{aligned}$$

where $\mu, \lambda \ge 0$ and $\Theta \in \mathbf{S}^m$ are such that

$$\left[\begin{array}{c|c|c} \sum_\ell \mu_\ell S_\ell & \frac{1}{2}(B-H^TA) & \frac{1}{2}H\\ \hline \frac{1}{2}(B-H^TA)^T & \sum_k \lambda_k T_k & \\ \hline \frac{1}{2}H^T & & \Theta \end{array}\right] \succeq 0.$$

We conclude that for all $x \in \mathcal{X}$,

$$\|\widehat{w}_H - Bx\| \le \max_{s\in\mathcal{S},t\in\mathcal{T}}\left\{\sum_\ell \lambda_\ell s_\ell + \sum_k \mu_k t_k + \sigma^2\xi^T\Theta\xi\right\} = \phi_{\mathcal{S}}(\lambda) + \phi_{\mathcal{T}}(\mu) + \sigma^2\xi^T\Theta\xi.$$

Due to (4.3) we have $\mathbf{E}\{\xi\xi^T\} \preceq \sigma^2 I$. We conclude that $\mathbf{E}\{\xi^T\Theta\xi\} \le \mathrm{Tr}(\Theta)$ what implies the first claim of the proposition. To complete the proof it remains to recall the bound

$$\mathrm{Prob}_\xi\left\{\xi^T\Theta\xi \ge \varkappa^2\mathrm{Tr}(\Theta))\right\} \le \epsilon$$

for deviations of the quadratic form of sub-Gaussian random vectors (cf., e.g., [78, 88, 89]). □

4.5.2 Proof of Corollary 4.2.1

Indeed, let $\varkappa = 1 + \sqrt{2\ln[\epsilon^{-1}]}$, and let λ^*, μ^* and Θ^* be components of an optimal solution to (4.8). Notice that λ^*, μ^* and Θ^* are feasible for (4.6). Moreover, $\overline{\lambda}, \overline{\mu}$ and $\overline{\Theta}$ where

$$\overline{\lambda} = \varkappa\lambda^*, \;\; \overline{\mu} = \varkappa^{-1}\mu^*, \;\; \overline{\Theta} = \varkappa^{-1}\Theta^*$$

are also feasible, so, by item (ii) of Proposition 4.2.1, the value

$$\phi_{\mathcal{T}}(\overline{\mu}) + \phi_{\mathcal{S}}(\overline{\lambda}) + \sigma^2\varkappa^2\mathrm{Tr}(\overline{\Theta}) \le \varkappa\left(\phi_{\mathcal{T}}(\mu^*) + \phi_{\mathcal{S}}(\lambda^*) + \sigma^2\mathrm{Tr}(\overline{\Theta})\right) \le \varkappa r^*$$

upper-bounds the ϵ-risk of $\widehat{w}^*_{\rm lin}$. □

4.5.3 Proof of Lemma 2

Let (λ, μ, Θ) be a feasible solution to (4.14) such that

$$\Lambda := \sum_\ell \lambda_\ell S_\ell \succ 0, \; \Xi := \sum_k \mu_k T_k \succ 0. \tag{4.29}$$

Note that every feasible solution to (4.14) remains feasible and satisfies (4.29) after replacing zero entries of λ_ℓ and μ_k, if any, with arbitrarily small positive entries.

Now, due to $\left[\begin{array}{c|c} \Lambda & \frac{1}{2}B \\ \hline \frac{1}{2}B^T & A^T\Theta A + \Xi \end{array}\right] \succeq 0$ we have

$$\left[\begin{array}{c|c} I_\nu & \frac{1}{2}\Lambda^{-1/2}B\Xi^{-1/2} \\ \hline \frac{1}{2}\Xi^{-1/2}B^T\Lambda^{-1/2} & \Xi^{-1/2}A^T\Theta A\Xi^{-1/2} + I_n \end{array}\right] \succeq 0.$$

When setting $\Theta^{1/2}A\Xi^{-1/2} = US$ with $UU^T = I_n$ and $S \succeq 0$, we get

$$\tfrac{1}{4}[\Lambda^{-1/2}B\Xi^{-1/2}]^T[\Lambda^{-1/2}B\Xi^{-1/2}] \preceq S^2 + I_n \preceq (S + I_n)^2$$

(recall that $S \succeq 0$). We conclude that there is $Q \in \mathbf{R}^{\nu\times n}$ of spectral norm $\|Q\|_{2,2} \le 1$ such that

$$\tfrac{1}{2}[\Lambda^{-1/2}B\Xi^{-1/2}] = Q(S + I_n)$$

and

$$B - 2\Lambda^{1/2}QS\Xi^{1/2} = 2\Lambda^{1/2}Q\Xi^{1/2}.$$

When recalling what S is and setting $H = 2\Theta^{1/2}UQ^T\Lambda^{1/2}$ we have

$$B - H^TA = 2\Lambda^{1/2}Q\Xi^{1/2}.$$

Due to $\|Q\|_{2,2} \le 1$, by Schur complement lemma, now it follows that

$$\left[\begin{array}{c|c} \Lambda & \frac{1}{2}(B - H^TA) \\ \hline \frac{1}{2}(B - H^TA)^T & \Xi \end{array}\right] \succeq 0. \tag{4.30}$$

Besides this, by construction,

$$\tfrac{1}{4}H\Lambda^{-1}H^T = \Theta^{1/2}\underbrace{UQ^TQU^T}_{\preceq I_n}\Theta^{1/2} \preceq \Theta,$$

that is,

$$\left[\begin{array}{c|c}\Lambda & \frac{1}{2}H^T\\ \hline \frac{1}{2}H & \Theta\end{array}\right]\succeq 0. \tag{4.31}$$

Finally, (4.30) together with (4.31) imply that matrix

$$\left[\begin{array}{c|c|c}2\Lambda & \frac{1}{2}(B-H^TA) & \frac{1}{2}H^T\\ \hline \frac{1}{2}(B-H^TA)^T & \Xi & \\ \hline \frac{1}{2}H & & \Theta\end{array}\right]$$

is positive semidefinite, meaning that $2\lambda,\mu,H$ and Θ form a feasible solution to (4.8). The corresponding objective value is

$$\phi_{\mathcal{S}}(2\lambda)+\phi_{\mathcal{T}}(\mu)+\sigma^2\varkappa^2\mathrm{Tr}(\Theta)=2\phi_{\mathcal{S}}(\lambda)+\phi_{\mathcal{T}}(\mu)+\sigma^2\varkappa^2\mathrm{Tr}(\Theta)\le\mathfrak{p}_{\varkappa}[\Theta]. \qquad \square$$

4.6 Contrast synthesis for the polyhedral estimate

Projecting, if necessary, the observations onto the image space of A, we reduce the situation to that in which this image space is the entire $\mathbf{R}^m$; that is, $m\le n$ with positive singular values $\sigma_1,...,\sigma_m$ of A. Let us assume that $n=m+d$ for some $d\ge 0$, and let

$$A=UDV^T,\quad D=\big[\underbrace{\mathrm{Diag}(\sigma_1,...,\sigma_m)}_{=:D_m},0_{m\times d}\big],\ U\in\mathbf{R}^{m\times m},\ V\in\mathbf{R}^{m\times n},$$

be the (full) singular-value decomposition of A. The starting point of the following computation is the bound (4.14) for the ϵ-risk of the polyhedral estimate: we have

$$\begin{aligned}
\mathfrak{p}^*_\chi &= 2\min_{\lambda,\mu,\Theta}\Big\{\phi_{\mathcal{S}}(\lambda)+\phi_{\mathcal{T}}(\mu)+\sigma^2\chi^2\mathrm{Tr}(\Theta):\ \lambda\ge 0,\mu\ge 0,\Theta\succeq 0,\\
&\qquad\qquad \left[\begin{array}{c|c}\sum_\ell\lambda_\ell S_\ell & \frac{1}{2}B\\ \hline \frac{1}{2}B^T & \sum_k\mu_kT_k+A^T\Theta A\end{array}\right]\succeq 0\Big\}\\
&= 2\min_{\lambda,\kappa,\Theta}\Big\{2\phi_{\mathcal{S}}(\lambda)+\phi_{\mathcal{T}}(\mu)+\sigma^2\chi^2\mathrm{Tr}(\Theta):\ \lambda>0,\mu\ge 0,\Theta\succeq 0,\\
&\qquad\qquad A^T\Theta A\succeq\tfrac{1}{4}B^T\Big[\sum_\ell\lambda_\ell S_\ell\Big]^{-1}B-\sum_k\mu_kT_k\Big\}\\
&= 2\min_{\lambda,\mu,\overline{\Theta}}\Big\{F(\lambda,\mu,\overline{\Theta}):=\phi_{\mathcal{S}}(\lambda)+\phi_{\mathcal{T}}(\mu)+\sigma^2\chi^2\mathrm{Tr}(\overline{\Theta}):\ \lambda>0,\mu\ge 0,\ \overline{\Theta}=U^T\Theta U\succeq 0,\\
&\qquad\qquad \left[\begin{array}{c|c}D_m\overline{\Theta}D_m & \\ \hline & \end{array}\right]=D^T\overline{\Theta}D\succeq V^T\left[\tfrac{1}{4}B^T\big[\sum_\ell\lambda_\ell S_\ell\big]^{-1}B-\sum_k\mu_kT_k\right]V\Big\}\\
&= 2\min_{\lambda,\mu,\overline{\Theta}}\Big\{F(\lambda,\mu,\overline{\Theta}):\ \lambda>0,\mu\ge 0,\ \overline{\Theta}\succeq 0,\\
&\left[\begin{array}{c|c}\overline{\Theta} & \\ \hline & \end{array}\right]\succeq\underbrace{\left[\begin{array}{c|c}D_m & \\ \hline & I\end{array}\right]^{-1}V^T\left[\tfrac{1}{4}B^T\big[\sum_\ell\lambda_\ell S_\ell\big]^{-1}B-\sum_k\mu_kT_k\right]V\left[\begin{array}{c|c}D_m & \\ \hline & I\end{array}\right]^{-1}}_{=:C(\lambda,\mu)}.\Big\}
\end{aligned}\tag{4.32}$$

Observe that the matrix-valued function $C(\lambda,\mu)$ is $\succeq$-convex for $\lambda>0$, and is negative definite for every fixed λ for all μ such that $\min_i\mu_i\ge\underline{\mu}$ large enough. On the other hand, when $Z(\lambda,\mu)\prec 0$

in the representation

$$C(\lambda,\mu)=\left[\begin{array}{c|c} X(\lambda,\mu) & Y(\lambda,\mu) \\ \hline Y^T(\lambda,\mu) & Z(\lambda,\mu) \end{array}\right]$$

the semidefinite constraint of (4.32) is satisfied if and only if

$$\overline{\Theta} \succeq W(\lambda,\mu) := X(\lambda,\mu)+Y^T(\lambda,\mu)Z(\lambda,\mu)^{-1}Y(\lambda,\mu).$$

As a result, when denoting $[M]_+$ the matrix obtained from a symmetric matrix M by replacing its eigenvalues with their positive parts in the eigenvalue decomposition of M, we conclude that

$$\mathfrak{p}^*_\chi = 2\min_{\lambda,\mu}\Big\{\phi_{\mathcal{S}}(\lambda)+\phi_{\mathcal{T}}(\mu)+\sigma^2\chi^2\mathrm{Tr}[W(\lambda,\mu)]_+ :\ \lambda>0,\mu\geq 0,\ Z(\lambda,\mu)\prec 0\Big\}.$$

The bottom line is that in the situation of this section, building Θ (and thus—the near-optimal polyhedral estimate) reduces to solving convex problem of design dimension $L+M$ with (relatively) easy-to-compute objective and constraints.

Chapter 5

On Robust Recovery of Signals from Indirect Observations

Abstract

Our focus is on robust recovery algorithms in statistical linear inverse problem. We consider two recovery routines—the much-studied linear estimate originating from Kuks and Olman [17] and polyhedral estimate introduced in [33]. It was shown in [32] that risk of these estimates can be tightly upper-bounded for a wide range of a priori information about the model through solving a convex optimization problem, leading to a computationally efficient implementation of nearly optimal estimates of these types. The subject of the present paper is design and analysis of linear and polyhedral estimates which are robust with respect to the uncertainty in the observation matrix. We evaluate performance of robust estimates under stochastic and deterministic matrix uncertainty and show how the estimation risk can be bounded by the optimal value of efficiently solvable convex optimization problem; "presumably good" estimates of both types are then obtained through optimization of the risk bounds with respect to estimate parameters.

5.1 Introduction

In this chapter we focus on the problem of recovering unknown signal x given noisy observation $\omega \in \mathbf{R}^m$,

$$\omega = Ax + \xi, \tag{5.1}$$

of the linear image Ax of x; here $\xi \in \mathbf{R}^m$ is observation noise. Our objective is to estimate the linear image $w = Bx \in \mathbf{R}^\nu$ of x known to belong to given convex and compact subset $\mathcal{X}$ of $\mathbf{R}^n$. The estimation problem above is a classical linear inverse problem. When statistically analysed, popular approaches to solving (5.1) (cf., e.g., [25–27, 74, 90–93]) usually assume a special structure of the problem, when matrix A and set $\mathcal{X}$ "fit each other," e.g., there exists a sparse approximation of the set $\mathcal{X}$ in a given basis/pair of bases, in which matrix A is "almost diagonal" (see, e.g. [34, 94] for detail). Under these assumptions, traditional results focus on estimation algorithms which are both numerically straightforward and statistically (asymptotically) optimal with closed form analytical description of estimates and corresponding risks. In this paper, A and B are "general" matrices of appropriate dimensions, and $\mathcal{X}$ is a rather general convex and compact set. Instead of deriving closed form expressions for estimates and risks (which under the circumstances seems to be impossible), we adopt an "operational" approach initiated in [34] and further developed in [30],

32, 33, 95, within which both the estimate and its risk are yielded by efficient computation, rather than by an explicit analytical description.

In particular, two classes of estimates were analyzed in [30, 32, 33] in the operational framework.

- *Linear estimates.* Since their introduction in [17], *linear estimates* are a standard part of the theoretical statistical toolkit. There is an extensive literature dealing with the design and performance analysis of linear estimates (see, e.g., [4, 23, 24, 28, 69–71]). When applied in the estimation problem we consider here, linear estimate $\widehat{w}^H_{\rm lin}(\omega)$ is of the form $\widehat{w}_H(\omega) = H^T\omega$ and is specified by a *contrast matrix* $H \in \mathbf{R}^{m\times\nu}$.

- *Polyhedral estimates*, of which the idea goes back to [72]. Authors show (see also [36, Chapter 2]) that it is near-optimal when recovering smooth multivariate regression function known to belong to a given Sobolev ball from noisy observations taken along a regular grid. It was recently reintroduced in [73] and [74]. In [33], these estimates are extended to the setting we are concerned with in the following way. In their paper, a Polyhedral estimate $\omega \mapsto \widehat{w}^H_{\rm poly}(\omega)$ is specified by a *contrast matrix* $H \in \mathbf{R}^{m\times M}$ according to

$$\omega \mapsto \widehat{x}^H(\omega) \in \operatorname*{Argmin}_{x\in\mathcal{X}} \|H^T(\omega - Ax)\|_\infty \mapsto \widehat{w}^H_{\rm poly}(\omega) := B\widehat{x}(\omega),$$

  and and shown to be nearly minimax in the same situations as Linear estimates, i.e., $\mathcal{X}$ and $\mathcal{B}_*$ being ellitopes, for a properly selected contrast matrix.

Our interest in these estimates stems from the results of [31–33] where it is shown that in the Gaussian case ($\xi \sim \mathcal{N}(0, \sigma^2 I_m)$), linear and polyhedral estimates with properly designed efficiently computable contrast matrices are near-minimax optimal in terms of their risks over a rather general class of loss functions and signal sets—ellitopes and spectratopes. [1]

In this paper we consider an estimation problem which is a generalization of that mentioned above in which observation matrix $A \in \mathbf{R}^{m\times n}$ is *uncertain*. Specifically, we assume that

$$\omega = A[\eta]x + \xi \tag{5.2}$$

where $\xi \in \mathbf{R}^m$ is zero mean random noise and

$$A[\eta] = A + \sum_{\alpha=1}^q \eta_\alpha A_\alpha \in \mathbf{R}^{m\times n} \tag{5.3}$$

where $A, A_1, ..., A_q$ are given matrices and $\eta \in \mathbf{R}^q$ is uncertain perturbation ("uncertainty" for short). We consider separately two situations: the first one in which the perturbation η is random ("random perturbation"), and the second one with η selected, perhaps in an adversarial fashion, from a given uncertainty set $\mathcal{U}$ ("uncertain-but-bounded perturbation"). Observation model (5.2) with random uncertainty is related to the linear regression problem with random errors in regressors [96–102] which is usually addressed through total least squares. It can also be seen as alternative modeling of the statistical inverse problem in which sensing matrix is recovered with stochastic error (see, e.g., [93, 103–107]). Estimation from observations (5.2) under uncertain-but-bounded perturbation of observation matrix can be seen as an extension of the problem of solving systems of equations affected by uncertainty which has received significant attention in the literature (cf.,

[1] Exact definitions of these sets are reproduced in the main body of the paper. For the time being, it suffices to point out two instructive examples: the bounded intersections of finitely many sets of the form $\{x : \|Px\|_p \le 1\}$, $p \ge 2$, is an ellitope (and a spectratope as well), and the unit ball of the spectral norm in the space of $m \times n$ matrices is a spectratope.

e.g., [108–114] and references therein). It is also tightly related to the system identification problem under uncertain-but-bounded perturbation of the observation of the state of the system [115–123].

In what follows, our goal is to extend the estimation constructions from [32] to the case of uncertain sensing matrix. Our strategy consists in constructing a tight efficiently computable convex in H upper bound on the risk of a candidate estimate, and then building a "presumably good" estimate by minimizing this bound in the estimate parameter H. Throughout the paper, we assume that the signal set $\mathcal{X}$ is an ellitope, and the norm $\|\cdot\|$ quantifying the recovery error is the maximum of a finite collection of Euclidean norms.

Our contributions can be summarized as follows.

A. In Section 5.2.1 we analyse the ϵ-risk (the maximum, over signals from $\mathcal{X}$, of the radii of $(1-\epsilon)$-confidence $\|\cdot\|$-balls) and the design of presumably good, in terms of this risk, linear estimates in the case of random uncertainty.

B. In Section 5.3.1, we build presumably good linear estimates in the case of *structured norm-bounded uncertainty* (cf. [124, Chapter 7] and references therein), thus extending the corresponding results of [118].

Developments in A and B lead to novel computationally efficient techniques for designing presumably good linear estimates for both random and uncertain-but-bounded perturbations.

Analysis and design of polyhedral estimates under uncertainty in sensing matrix form the subject of Sections 5.2.2 (random perturbations) and 5.3.2 (uncertain-but-bounded perturbations). The situation here is as follows:

C. The random perturbation case of the *Analysis problem*

> *Given contrast matrix H, find a provably tight efficiently computable upper bound on ϵ-risk of the associated estimate*

is the subject of Section 5.2.2, where it is solved "in the full range" of our assumptions (ellitopic $\mathcal{X}$, sub-Gaussian zero mean η and ξ). In contrast, the random perturbation case of the *Synthesis problem* in which we want to minimize the above bound w.r.t. H turns out to be more involving—the bound to be optimized happens to be nonconvex in H. When there is no uncertainty in sensing matrix, this difficulty can be somehow circumvented [32, Section 5.1]; however, when uncertainty in sensing matrix is present, the strategy developed in [32, Section 5.1] happens to work only when $\mathcal{X}$ is an ellipsoid rather than a general-type ellitope. The corresponding developments are the subject of Sections 5.2.2, 5.2.2, and 5.2.2.

D. In our context, analysis and design of polyhedral estimates under uncertain-but-bounded perturbations in the sensing matrix appears to be the most difficult; our very limited results on this subject form the subject of Section 5.3.2.

Notation and assumptions. We denote with $\|\cdot\|$ the norm on $\mathbf{R}^\nu$ used to measure the estimation error. In what follows, $\|\cdot\|$ is a maximum of Euclidean norms

$$\|u\| = \max_{\ell\le L} \sqrt{u^T R_\ell u} \tag{5.4}$$

where $R_\ell \in \mathbf{S}^\nu_+$, $\ell = 1, ..., L$, are given matrices with $\sum_\ell R_\ell \succ 0$.

Throughout the paper, unless otherwise is explicitly stated, we assume that observation noise ξ is zero-mean sub-Gaussian, $\xi \sim \mathcal{SG}(0, \sigma^2 I)$, i.e., for all $t \in \mathbf{R}^m$,

$$\mathbf{E}\left\{e^{t^T\xi}\right\} \le \exp\left(\tfrac{\sigma^2}{2}\|t\|_2^2\right). \tag{5.5}$$

5.2 Random perturbations

In this section we assume that uncertainty η is sub-Gaussian, with parameters $0, I$, i.e.,

$$\mathbf{E}\left\{e^{t^T\eta}\right\} \le \exp\left(\tfrac{1}{2}\|t\|_2^2\right) \quad \forall t \in \mathbf{R}^q. \tag{5.6}$$

In this situation, given $\epsilon \in (0,1)$, we quantify the quality of recovery $\widehat{w}(\cdot)$ of $w = Bx$ by its maximal over $x \in \mathcal{X}$ *ϵ-risk*

$$\mathrm{Risk}_\epsilon[\widehat{w}|\mathcal{X}] := \sup_{x\in\mathcal{X}} \inf\{\rho : \mathrm{Prob}_{\xi,\eta}\{\|Bx - \widehat{w}(A[\eta]x + \xi)\| > \rho\} \le \epsilon\} \tag{5.7}$$

(the radius of the smallest $\|\cdot\|$-ball centered at $\widehat{w}(\omega)$ which covers x, uniformly over $x \in \mathcal{X}$).

5.2.1 Design of presumably good linear estimate

Preliminaries: ellitopes

Throughout this section, we assume that the signal set $\mathcal{X}$ is *a basic ellitope*. Recall that, by definition [30, 32], a basic ellitope in $\mathbf{R}^n$ is a set of the form

$$\mathcal{X} = \{x \in \mathbf{R}^n : \exists t \in \mathcal{T} : z^T T_k z \le t_k,\ k \le K\}, \tag{5.8}$$

where $T_k \in \mathbf{S}^n_+$, $T_k \succeq 0$, $\sum_k T_k \succ 0$, and $\mathcal{T} \subset \mathbf{R}^K_+$ is a convex compact set with a nonempty interior which is monotone: whenever $0 \le t' \le t \in \mathcal{T}$ one has $t' \in \mathcal{T}$. We refer to K as *ellitopic dimension* of $\mathcal{X}$.

Clearly, every basic ellitope is a convex compact set with nonempty interior which is symmetric w.r.t. the origin. For instance,

A. Bounded intersection $\mathcal{X}$ of K centered at the origin ellipsoids/elliptic cylinders $\{x \in \mathbf{R}^n : x^T T_k x \le 1\}$ $[T_k \succeq 0]$ is a basic ellitope:

$$\mathcal{X} = \{x \in \mathbf{R}^n : \exists t \in \mathcal{T} := [0,1]^K : x^T T_k x \le t_k,\ k \le K\}$$

In particular, the unit box $\{x \in \mathbf{R}^n : \|x\|_\infty \le 1\}$ is a basic ellitope.

B. A $\|\cdot\|_p$-ball in $\mathbf{R}^n$ with $p \in [2, \infty]$ is a basic ellitope:

$$\{x \in \mathbf{R}^n : \|x\|_p \le 1\} = \big\{x : \exists t \in \mathcal{T} = \{t \in \mathbf{R}^n_+, \|t\|_{p/2} \le 1\} : \underbrace{x_k^2}_{x^T T_k x} \le t_k,\ k \le K\big\}.$$

In the present context, our interest for ellitopes is motivated by their special relationship with the optimization problem

$$\mathrm{Opt}_*(C) = \max_{x\in\mathcal{X}} x^T C x,\ C \in \mathbf{S}^n \tag{5.9}$$

of maximizing a homogeneous quadratic form over $\mathcal{X}$. As it is shown in [32], when $\mathcal{X}$ is an ellitope, (5.9) admits "reasonably tight" efficiently computable upper bound. Specifically,

Theorem 5.2.1 [32, Proposition 4.6] *Given ellitope (5.8) and matrix C, consider the quadratic maximization problem (5.9) along with its relaxation*[2]

$$\mathrm{Opt}(C) = \min_{\lambda}\left\{\phi_{\mathcal{T}}(\lambda) : \lambda \geq 0, \sum\nolimits_k \lambda_k T_k - C \succeq 0\right\} \tag{5.10}$$

The problem is computationally tractable and solvable, and $\mathrm{Opt}(C)$ is an efficiently computable upper bound on $\mathrm{Opt}_*(C)$. This upper bound is tight:

$$\mathrm{Opt}_*(C) \leq \mathrm{Opt}(C) \leq 3\ln(\sqrt{3}K)\mathrm{Opt}_*(C).$$

Tight upper bounding of the risk of linear estimates

Consider the linear estimate

$$\widehat{w}^H(\omega) = H^T\omega. \quad [H \in \mathbf{R}^{m\times\nu}]$$

Proposition 5.2.1 *In the setting of this section, synthesis of a presumably good linear estimate reduces to solving the convex optimization problem*

$$\min_{H\in\mathbf{R}^{m\times\nu}} \mathfrak{R}[H] \tag{5.11}$$

where

$$\begin{array}{rcl}\mathfrak{R}[H] & = & \min\limits_{\substack{\lambda_\ell,\mu^\ell,\kappa_\ell,\\ \varkappa^\ell,\rho,\varrho}} \left\{ \left[1+\sqrt{2\ln(2L/\epsilon)}\right]\left[\sigma\max\limits_{\ell\leq L}\|HR_\ell^{1/2}\|_{\mathrm{Fro}}+\rho\right]+\varrho : \right. \\ & & \mu^\ell\geq 0, \varkappa^\ell\geq 0,\ \lambda_\ell+\phi_{\mathcal{T}}(\mu_\ell)\leq\rho, \kappa_\ell+\phi_{\mathcal{T}}(\varkappa^\ell)\leq\varrho,\ \ell\leq L \\ & & \left[\begin{array}{c|c} \lambda_\ell I_{\nu q} & \frac{1}{2}\left[R_\ell^{1/2}H^TA_1;...;R_\ell^{1/2}H^TA_q\right] \\ \hline \frac{1}{2}\left[A_1^THR_\ell^{1/2},...,A_q^THR_\ell^{1/2}\right] & \sum_k\mu_k^\ell T_k\end{array}\right]\succeq 0,\ \ell\leq L \\ & & \left.\left[\begin{array}{c|c} \kappa^\ell I_\nu & \frac{1}{2}R_\ell^{1/2}[B-H^TA] \\ \hline \frac{1}{2}[B-H^TA]^TR_\ell^{1/2} & \sum_k\varkappa_k^\ell T_k\end{array}\right]\succeq 0,\ \ell\leq L\right\}\end{array} \tag{5.12}$$

For a candidate contrast matrix H, the ϵ-risk of the linear estimate $\widehat{w}^H_{\mathrm{lin}}(\omega) = H^T\omega$ is upper-bounded by $\mathfrak{R}[H]$.

A modification

Let us assume that a K-repeated version of observation (5.2) is available, i.e., we observe

$$\omega^K = \{\omega_k = A[\eta_k]x + \xi_k,\ k = 1,...,K\} \tag{5.13}$$

with independent across k pairs (ξ_k,η_k). In this situation, we can relax the assumption of sub-Gaussianity of ξ and η to the second moment boundedness condition

$$\mathbf{E}\{\xi\xi^T\}\preceq\sigma^2 I_m, \quad \mathbf{E}\left\{\eta\eta^T\right\}\preceq I_q. \tag{5.14}$$

[2] Here and below, we use notation $\phi_{\mathcal{S}}(\cdot)$ for the support function of a convex set $\mathcal{S}\subset\mathbf{R}^n$: for $y\in\mathbf{R}^n$,

$$\phi_{\mathcal{S}}(y) = \sup_{u\in\mathcal{S}} y^Ts.$$

Let us consider the following construction. For each $\ell \le L$, given $H \in \mathbf{R}^{m\times\nu}$ we denote

$$\begin{array}{rcl}\widetilde{\mathfrak{R}}_\ell[H] & = & \min_{\lambda,\mu,\kappa,\varkappa}\left\{\sigma\|HR_\ell^{1/2}\|_{\rm Fro}+\lambda+\phi_{\mathcal{T}}(\mu)+\kappa+\phi_{\mathcal{T}}(\varkappa):\right.\\ & & \mu\ge 0,\varkappa\ge 0,\left[\begin{array}{c|c}\kappa I_\nu & \frac{1}{2}R_\ell^{1/2}[B-H^TA]\\ \hline \frac{1}{2}[B-H^TA]^TR_\ell^{1/2} & \sum_k\varkappa_kT_k\end{array}\right]\succeq 0\\ & & \left.\left[\begin{array}{c|c}\lambda I_{\nu q} & \frac{1}{2}\left[R_\ell^{1/2}H^TA_1;...;R_\ell^{1/2}H^TA_q\right]\\ \hline \frac{1}{2}\left[A_1^THR_\ell^{1/2},...,A_q^THR_\ell^{1/2}\right] & \sum_k\mu_kT_k\end{array}\right]\succeq 0\right\}\end{array} \tag{5.15}$$

and consider the convex optimization problem

$$\widetilde{H}_\ell \in \operatorname*{Argmin}_H \widetilde{\mathfrak{R}}_\ell[H]. \tag{5.16}$$

We define the "reliable estimate" $\widehat{w}^{(r)}(\omega^K)$ of $w=Bx$ as follows.

1. Given $H_\ell \in \mathbf{R}^{m\times\nu}$ and observations ω_k we compute linear estimates $w_\ell(\omega_k)=H_\ell\omega_k$, $\ell=1,...,L$, $k=1,...,K$;

2. We define vectors $z_\ell\in\mathbf{R}^\nu$ as geometric medians of $w_\ell(\omega_k)$:

$$z_\ell(\omega^K)\in\operatorname*{Argmin}_z\sum_{k=1}^K\|R_\ell^{1/2}(w_\ell(\omega_k)-z)\|_2,\ \ell=1,...,L.$$

3. Finally, we select as $\widehat{w}^{(r)}(\omega^K)$ any point of the set

$$\mathcal{W}(\omega^K)=\bigcap_{\ell=1}^L\left\{w\in\mathbf{R}^\nu:\ \|R_\ell^{1/2}(z_\ell(\omega^K)-w)\|_2\le 4\widetilde{\mathfrak{R}}_\ell[H_\ell]\right\}.$$

or set $\widehat{w}^{(r)}(\omega^K)$ a once for ever fixed point, e.g., $\widehat{w}^{(r)}(\omega^K)=0$ if $\mathcal{W}(\omega^K)=\emptyset$.

We have the following analog of Proposition 5.2.1.

Proposition 5.2.2 *In the situation of this section, it holds*

$$\sup_{x\in\mathcal{X}}\mathbf{E}_{\eta_k,\xi_k}\left\{\|R_\ell^{1/2}(w_\ell(\omega_k)-Bx)\|_2^2\right\}\le\widetilde{\mathfrak{R}}_\ell^2[H_\ell],\ \ell\le L, \tag{5.17}$$

and

$$\mathrm{Prob}\left\{\|R_\ell^{1/2}(z_\ell(\omega^K)-Bx)\|_2\ge 4\widetilde{\mathfrak{R}}_\ell[H_\ell]\right\}\le e^{-0.1070K},\ \ell\le L. \tag{5.18}$$

As a consequence, whenever $K\ge\ln[L/\epsilon]/0.1070$, the ϵ-risk of the aggregated estimate $\widehat{w}^{(r)}(\omega^K)$ satisfies

$$\mathrm{Risk}_\epsilon[\widehat{w}^{(r)}(\omega^K)|\mathcal{X}]\le\overline{\mathfrak{R}},\ \ \overline{\mathfrak{R}}=8\max_{\ell\le L}\widetilde{\mathfrak{R}}_\ell[H_\ell].$$

Remark. Proposition 5.2.2 is motivated by the desire to capture situations in which sub-Gaussian assumption on η and ξ does not hold or is too restrictive. Consider, e.g., the case where the uncertainty in the sensing matrix reduces to zeroing out some randomly selected columns in the nominal matrix $\overline{A}$ (think of taking picture through the window with frost patterns). Denoting by γ the probability to zero out a particular column and assuming that columns are zeroed out independently, model (5.2) in this situation reads

$$\omega = A[\eta]x + \xi,\ A[\eta] = (1-\gamma)\overline{A} + \sum_{\alpha=1}^{n} \eta_\alpha A_\alpha$$

where $\eta_1, ..., \eta_n$ are i.i.d. zero mean random variables taking values $(\gamma-1)\rho$ and $\gamma\rho$ with probabilities γ and $1-\gamma$, and A_α, $1 \le \alpha \le n$, is an $m \times n$ matrix with all but the α-th column being zero and $\mathrm{Col}_\alpha[A_\alpha] = \rho^{-1}\mathrm{Col}_\alpha[\overline{A}]$. Scaling factor ρ is selected to yield the unit sub-Gaussianity parameter of η or $\mathbf{E}\{\eta_\alpha^2\} = 1$ depending on whether Proposition 5.2.1 or Proposition 5.2.2 is used. For small γ, the scaling factor ρ is essentially smaller in the first case, resulting in larger "disturbance matrices" A_α and therefore—in stricter constraints in the optimization problem (5.11), (5.12) responsible for the design of the linear estimate.

Numerical illustration

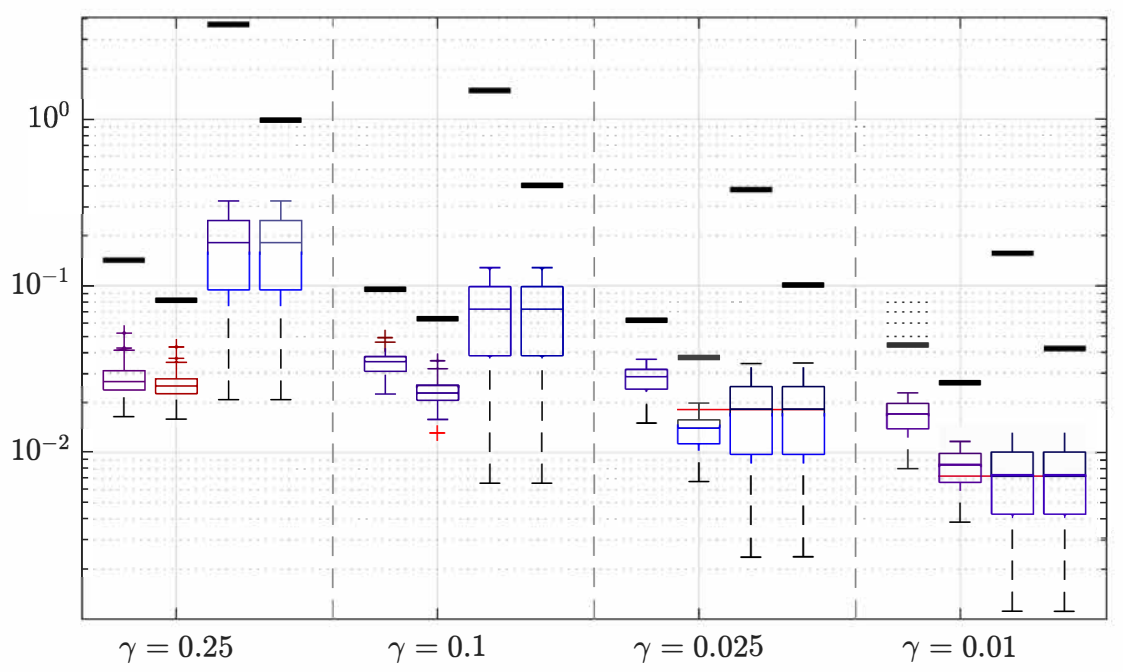


Figure 5.1: Distributions of ℓ_2-recovery errors and upper bounds of the robust and "nominal" estimates for different values of γ parameter.

In Figure 5.1 we present results of a toy experiment in which

- $n = 32, m = 32$, and $\nu = 16$, $\overline{A}x \in \mathbf{R}^m$ is the discrete time convolution of $x \in \mathbf{R}^n$ with a simple kernel $\varkappa$ of length 9 restricted onto the "time horizon" $\{1, ..., n\}$, and Bx cuts off x the first ν entries. We consider Gaussian perturbation $\eta \sim \mathcal{N}(0, \gamma^2 I_q)$, $q = 9$, and

$A[\eta]x = [A + \sum_{\alpha=1}^{q} \eta_\alpha A_\alpha]x$ which is the convolution of x with the kernel $\varkappa_\eta$ restricted onto the time horizon $\{1, ..., n\}$, γ being the control parameter.

- $L = 1$ and $\|\cdot\| = \|\cdot\|_2$,
- $\mathcal{X}$ is the ellipsoid $\{x : \sum_i i^2 [Dx]_i^2 \leq 1\}$, where D is the matrix of inverse Discrete Cosine Transform of size $n \times n$.
- $\xi \sim \mathcal{N}(0, \sigma^2 I_m)$, $\sigma = 10^{-4}$.

In each cell of the plot we represent error distributions and upper risk bounds (horizontal bar) of four estimates (from left to right) for different uncertainty levels γ: (1) robust estimate by Proposition 5.2.1 and upper bound $\mathfrak{R}$ on its 0.05-risk, (2) single-observation estimate $w_1(\omega_1) = H_1\omega_1$ yielded by the minimizer H_1 of $\widetilde{\mathfrak{R}}_1[H]$ over H, see (5.15), and upper bound $\widetilde{\mathfrak{R}}_1[H_1]$ on its *expected error risk*,[3] (3) "nominal" estimate—estimate by Proposition 5.2.1 as applied to the "no uncertainty" case where all A_α in (5.3) are set to 0 and upper bound $\mathfrak{R}$ from (5.12) on its 0.05-risk computed using actual uncertainty level, (4) "nominal" estimate $\widetilde{w}_1(\omega_1) = \widetilde{H}_1\omega_1$ yielded by the minimizer $\widetilde{H}_1$ of $\widetilde{\mathfrak{R}}_1[H]$ over H in the "no uncertainty" case and upper bound $\widetilde{\mathfrak{R}}_1[\widetilde{H}_1]$ on its "actual"—with uncertainty present—expected error risk.

5.2.2 Design of presumably good polyhedral estimate

Preliminaries on polyhedral estimates

Consider a slightly more general than (5.2), (5.3) observation scheme

$$\omega = Ax + \zeta \tag{5.19}$$

where $A \in \mathbf{R}^{m\times n}$ is given, unknown signal x is known to belong to a given signal set $\mathcal{X}$ given by (5.8), and ζ is observation noise with probability distribution P_x which can depend on x. For example, when observation ω is given by (5.2), (5.3), we have

$$\zeta = \sum_{\alpha=1}^{q} \eta_\alpha A_\alpha x + \xi \tag{5.20}$$

with zero mean sub-Gaussian η and ξ.

When building polyhedral estimate in the situation in question, one, given tolerance $\epsilon \in (0, 1)$ and a positive integer M, specifies a computationally tractable convex set $\mathcal{H}$, the larger the better, of vectors $h \in \mathbf{R}^m$ such that

$$\mathrm{Prob}_{\zeta\sim P_x}\{|h^T\zeta| > 1\} \leq \epsilon/M \quad \forall x \in \mathcal{X}. \tag{5.21}$$

A polyhedral estimate $\widehat{w}^H(\cdot)$ is specified by *contrast matrix* $H \in \mathbf{R}^{M\times n}$ restricted to have all columns in $\mathcal{H}$ according to

$$\omega \mapsto \widehat{x}^H(\omega) \in \operatorname*{Argmin}_{u\in\mathcal{X}} \left\{\|H^T[Au - \omega]\|_\infty\right\}, \ \widehat{w}^H_{\mathrm{poly}}(\omega) = B\widehat{x}^H(\omega). \tag{5.22}$$

It is easily seen (cf. [32, Proposition 5.1.1]) that the ϵ-risk (5.7) of the above estimate is upper-bounded by the quantity

$$\mathfrak{p}[H] = \sup_y \left\{\|By\| : y \in 2\mathcal{X}, \|H^T Ay\|_\infty \leq 2\right\}. \tag{5.23}$$

[3] We define expected error risk of a K-observation estimate $\widehat{x}(\omega^K)$ of Bx as $\sup_{x\in\mathcal{X}} \mathbf{E}_{\omega^K\sim P_x^K}\{\|\widehat{x}(\omega^K) - Bx\|\}$, where P_x^K is the distribution of ω^K stemming from x.

Indeed, let $h_1, ..., h_M$ be the columns of H. For $x \in \mathcal{X}$ fixed, the inclusions $h_j \in \mathcal{H}$ imply that the P_x-probability of the event $Z_x = \{\zeta : |\zeta^T h_j| \le 1 \, \forall j \le M\}$ is at least $1-\epsilon$. When this event takes place, we have $\|H^T[\omega - Ax]\|_\infty \le 1$, which combines with $x \in \mathcal{X}$ to imply that $\|H^T[\omega - A\widehat{x}^H(\omega)]\|_\infty \le 1$, so that $\|H^T A[x - \widehat{x}^H(\omega)]\|_\infty \le 2$, and besides this, $x - \widehat{x}^H(\omega) \in 2\mathcal{X}$, whence $\|Bx - \widehat{w}^H_{\rm poly}(\omega)\| \le \mathfrak{p}[H]$ by definition of $\mathfrak{p}[H]$. The bottom line is that whenever $x \in \mathcal{X}$ and $\zeta = \omega - Ax \in Z_x$, which happens with P_x-probability at least $1-\epsilon$, we have $\|Bx - \widehat{w}^H_{\rm poly}(\omega)\| \le \mathfrak{p}[H]$, whence the ϵ-risk of the estimate $\widehat{w}^H_{\rm poly}$ indeed is upper-bounded by $\mathfrak{p}[H]$.

To get a presumably good polyhedral estimate, one minimizes $\mathfrak{p}[H]$ over $M \times \nu$ matrices H with columns from $\mathcal{H}$. Precise minimization is problematic, because $\mathfrak{p}[\cdot]$, while being convex, is usually difficult to compute. Thus, the design routine proposed in [33] goes via minimizing an efficiently computable upper bound on $\mathfrak{p}[H]$. It is shown in [32, Section 5.1.5] that when $\mathcal{X}$ is ellitope (5.8) and $\|u\| = \|Ru\|_2$, a reasonably tight upper bound on $\mathfrak{p}[H]$ is given by the efficiently xcomputable function

$$\mathfrak{p}_+[H] = 2\min_{\lambda,\mu,\upsilon}\left\{\lambda + \phi_{\mathcal{T}}(\mu) + \sum\nolimits_i \upsilon_i : \begin{array}{c}\mu \ge 0, \upsilon \ge 0 \\ \left[\begin{array}{c|c} \lambda I_\nu & \frac{1}{2}RB \\ \hline \frac{1}{2}B^TR^T & A^TH\mathrm{Diag}\{\upsilon\}H^TA + \sum_k \mu_k T_k \end{array}\right] \succeq 0\end{array}\right\}.$$

Synthesis of a presumably good polyhedral estimate reduces to minimizing the latter function in H under the restriction $\mathrm{Col}_j[H] \in \mathcal{H}$. Note that the latter problem still is nontrivial because $\mathfrak{p}_+$ is nonconvex in H.

Our objective here is to implement the outlined strategy in the case of observation ω given by (5.2), (5.3).

Specifying $\mathcal{H}$

Our first goal is to specify, given tolerance $\delta \in (0,1)$, a set $\mathcal{H}_\delta \subset \mathbf{R}^m$, the larger the better, such that

$$h \in \mathcal{H}_\delta, x \in \mathcal{X} \Rightarrow \mathrm{Prob}_{\zeta \sim P_x}\{|h^T\zeta| > 1\} \le \delta. \tag{5.24}$$

Note that a "tight" sufficient condition for the validity of (5.24) is

$$\mathrm{Prob}_\xi\{|h^T\xi| > 1/2\} \le \delta/2, \tag{5.25a}$$

$$\mathrm{Prob}_\eta\left\{\left|\sum\nolimits_{\alpha=1}^q [h^TA_\alpha x]\eta_\alpha\right| > 1/2\right\} \le \delta/2, \, \forall x \in \mathcal{X}. \tag{5.25b}$$

Under the sub-Gaussian assumption (5.5), $h^T\xi$ is itself sub-Gaussian, $h^T\xi \sim \mathcal{SG}(0, \sigma^2\|h\|_2^2)$; thus, a tight sufficient condition for (5.25a) is

$$\|h\|_2 \le [\sigma\chi(\delta)]^{-1}, \; \chi(\delta) = 2\sqrt{2\ln(2/\delta)}. \tag{5.26}$$

Furthermore, by (5.6), r.v. $\sum_{\alpha=1}^q [h^TA_\alpha x]\eta_\alpha = h^T[A_1x, ..., A_qx]\eta$ is sub-Gaussian with parameters 0 and $\|[h^TA_1x; ...; h^TA_qx]\|_2^2$, implying the validity of (5.25b) for a given x whenever

$$\|[h^TA_1x; ...; h^TA_qx]\|_2 \le \chi^{-1}(\delta).$$

We want this relation to hold true for every $x \in \mathcal{X}$, that is, we want the operator norm $\|\cdot\|_{\mathcal{X},2}$ of the mapping

$$x \mapsto \mathcal{A}[h]x, \; \mathcal{A}[h] = [h^TA_1; h^TA_2; ...; h^TA_q] \tag{5.27}$$

induced by the norm $\|\cdot\|_{\mathcal{X}}$ on the argument and the norm $\|\cdot\|_2$ on the image space to be upper-bounded by $\chi(\delta)$:

$$\|\mathcal{A}[h]\|_{\chi,2} \le \chi^{-1}(\delta). \tag{5.28}$$

Invoking [118, Theorem 3.1] (cf. also the derivation in the proof of Proposition 5.2.1 in Section 5.4.2), a tight sufficient condition for the latter relation is

$$\mathrm{Opt}[h] := \min_{\lambda,\mu}\left\{\lambda + \phi_{\mathcal{T}}(\mu) : \mu \ge 0,\ \left[\begin{array}{c|c} \lambda I_q & \frac{1}{2}\mathcal{A}[h] \\ \hline \frac{1}{2}\mathcal{A}^T[h] & \sum_k \mu_k T_k \end{array}\right] \succ 0\right\} \le \chi^{-1}(\delta), \tag{5.29}$$

tightness meaning that $\mathrm{Opt}[h]$ is within factor $O(1)\sqrt{\ln(K+1)}$ of $\|\mathcal{A}[h]\|_{\mathcal{X},2}$.

The bottom line is that with $\mathcal{H}_\delta$ specified by constraints (5.26) and (5.28) (or by the latter replaced with its tight relaxation (5.29)) we do ensure (5.24).

Bounding the risk of the polyhedral estimate $\widehat{w}^H$

Proposition 5.2.3 *In the situation of this section, let $\epsilon \in (0,1)$, and let $H = [H_1, ..., H_L]$ be $m \times ML$ matrix with L blocks $H_\ell \in \mathbf{R}^{m\times M}$ such that $\mathrm{Col}_j[H] \in \mathcal{H}_\delta$ for all $j \le ML$ and $\delta = \epsilon/ML$. Consider optimization problem*

$$\mathfrak{p}_+[H] = 2\min_{\lambda_\ell,\mu^\ell,\upsilon^\ell,\rho}\left\{\rho : \mu^\ell \ge 0, \upsilon^\ell \ge 0,\ \lambda_\ell + \phi_{\mathcal{T}}(\mu^\ell) + \sum_{j=1}^M \upsilon_j^\ell \le \rho,\ \ell \le L\right.$$
$$\left.\left[\begin{array}{c|c} \lambda_\ell I_\nu & \frac{1}{2}R_\ell^{1/2}B \\ \hline \frac{1}{2}B^T R_\ell^{1/2} & A^T H_\ell \mathrm{Diag}\{\upsilon^\ell\} H_\ell^T A + \sum_k \mu_k^\ell T_k \end{array}\right] \succeq 0,\ \ell \le L\right\}. \tag{5.30}$$

Then

$$\mathrm{Risk}_\epsilon[\widehat{w}^H|\mathcal{X}] \le \mathfrak{p}_+[H].$$

Optimizing $\mathfrak{p}_+[H]$—the strategy

Proposition 5.2.3 resolves the *analysis* problem—it allows to efficiently upper-bound the ϵ-risk of a given polyhedral estimate $\widehat{w}^H_{\mathrm{poly}}$. At the same time, "as is," it does not allow to build the estimate itself (solve the "estimate synthesis" problem—compute a presumably good contrast matrix) because straightforward minimization of $\mathfrak{p}_+[H]$ (that is, adding H to decision variables of the right hand side of (5.30) results in a nonconvex problem. A remedy, as proposed in [32, Section 5.1], stems from the concept of a *cone compatible with a convex compact set* $\mathcal{H} \subset \mathbf{R}^m$ which is defined as follows:

Given positive integer J and real $\varkappa \ge 1$ we say that a closed convex cone $\mathbf{K} \subset \mathbf{S}^m_+ \times \mathbf{R}_+$ is $(J,\varkappa)$-compatible with $\mathcal{H}$ if

(i) whenever $h_1, ..., h_J \in \mathcal{H}$ and $\upsilon \in \mathbf{R}^J_+$, the pair $\left(\sum_{j=1}^J \upsilon_j h_j h_j^T, \sum_j \upsilon_j\right)$ belongs to $\mathbf{K}$, and "nearly vice versa":

(ii) given $(\Theta, \varrho) \in \mathbf{K}$ and $\varkappa \ge 1$, we can efficiently build collections of vectors $h_j \in \mathcal{H}$, and reals $\upsilon_j \ge 0$, $j \le J$, such that $\Theta = \sum_{j=1}^J \upsilon_j h_j h_j^T$ and $\sum_j \upsilon_j \le \varkappa\varrho$.

Example. Let $\mathcal{H}$ be a centered at the origin Euclidean ball of radius $R > 0$ in $\mathbf{R}^J$. When setting

$$\mathbf{K} = \{(\Theta, \varrho) : \Theta \succeq 0, \mathrm{Tr}(\Theta) \le R^2 \varrho \mathrm{Tr}(\Theta)\},$$

we obtain a cone $(M, 1)$-compatible with $\mathcal{H}$. Indeed, for $h_j \in \mathcal{H}$ and $\upsilon_j \ge 0$ we have

$$\mathrm{Tr}\left(\sum_j \upsilon_j h_j h_j^T\right) \le R^2 \sum_j \upsilon_j,$$

that is $\left(\Theta := \sum_j \upsilon_j h_j h_j^T, \varrho := \sum_j \upsilon_j\right) \in \mathbf{K}$. Vice versa, given $(\Theta, \varrho) \in \mathbf{K}$, i.e., $\Theta \succeq 0$ and $\varrho \ge \mathrm{Tr}(\Theta)/R^2$ and specifying $f_1, ..., f_m$ as the orthonormal system of eigenvectors of Θ, and λ_j as the corresponding eigenvalues and setting $h_j = Rf_j$, $\upsilon_j = R^{-2}\lambda_j$), we get $h_j \in \mathcal{H}$, $\Theta = \sum_j \upsilon_j h_j h_j^T$ and $\sum_j \upsilon_j = \mathrm{Tr}(\Theta)/R^2 \le \varrho$.

Coming back to the problem of minimizing $\mathfrak{p}_+[H]$ in H, assume that we have at our disposal a cone $\mathbf{K}$ which is $(M, \varkappa)$-compatible with $\mathcal{H}_\delta$. In this situation, we can replace the nonconvex problem

$$\min_{H=[H^1,...,H^L]} \{\mathfrak{p}_+[H] : \mathrm{Col}_j[H^\ell]_j \in \mathcal{H}_\delta\} \tag{5.31}$$

with the problem

$$\min_{\substack{\bar\lambda_\ell, \bar\mu^\ell, \\ \Theta_\ell, \varrho_\ell, \bar\rho}} \left\{\bar\rho : (\Theta_\ell, \varrho_\ell) \in \mathbf{K}, \bar\mu^\ell \ge 0, \bar\lambda_\ell + \phi_{\mathcal{T}}(\bar\mu^\ell) + \varrho_\ell \le \bar\rho, \ell \le L, \right.$$
$$\left. \left[\begin{array}{c|c} \bar\lambda_\ell I_\nu & \frac{1}{2} R_\ell^{1/2} B \\ \hline \frac{1}{2} B^T R_\ell^{1/2} & A^T \Theta_\ell A + \sum_k \bar\mu_k^\ell T_k \end{array}\right] \succeq 0, \ell \le L \right\}. \tag{5.32}$$

Unlike (5.31), the latter problem is convex and efficiently solvable provided that $\mathbf{K}$ is computationally tractable, and can be considered as "tractable $\sqrt{\varkappa}$-tight" relaxation of the problem of interest (5.31). Namely,

- Given a feasible solution $H_\ell, \lambda_\ell, \mu^\ell, \upsilon^\ell, \rho$ to the problem of interest (5.31), we can set

  $$\Theta_\ell = \sum_{j=1}^M \upsilon_j^\ell \mathrm{Col}_j[H_\ell] \mathrm{Col}_j^T[H_\ell], \quad \varrho_\ell = \sum_j \upsilon_j^\ell,$$

  thus getting $(\Theta_\ell, \varrho_\ell) \in \mathbf{K}$. By (i) in the definition of compatibility, $\Theta_\ell, \varrho_\ell, \bar\lambda_\ell = \lambda_\ell, \bar\mu^\ell = \mu^\ell, \bar\rho = \rho$ is a feasible solution to (5.32), and this transformation preserves the value of the objective

- Vice versa, given a feasible solution $\Theta_\ell, \varrho_\ell, \bar\lambda_\ell, \bar\mu^\ell, \bar\rho$ to (5.32) and invoking (ii) of the definition of compatibility, we can convert, in a computationally efficient way, the pairs $(\Theta_\ell, \rho_\ell) \in \mathbf{K}$ into the pairs $H_\ell \in \mathbf{R}^{m\times M}$, $\bar\upsilon^\ell \in \mathbf{R}_+^m$ in such a way that the columns of H_ℓ belong to $\mathcal{H}_\delta$, $\Theta_\ell = H_\ell \mathrm{Diag}\{\bar\upsilon^\ell\} H_\ell^T$, $\sum_j \bar\upsilon_j^\ell \le \varkappa\varrho_\ell$. Assuming w.l.o.g. that all matrices $R_\ell^{1/2} B$ are nonzero, we obtain $\phi_{\mathcal{T}}(\bar\mu^\ell) + \varrho_\ell > 0$ and $\bar\lambda_\ell > 0$ for all ℓ. We claim that setting

  $$\gamma_\ell = \sqrt{[\phi_{\mathcal{T}}(\bar\mu^\ell) + \varkappa\varrho_\ell]/\bar\lambda_\ell}, \ \lambda_\ell = \gamma_\ell \bar\lambda_\ell, \ \mu_\ell = \gamma_\ell^{-1}\bar\mu_\ell, \upsilon^\ell = \gamma_\ell^{-1}\bar\upsilon^\ell, \ \rho = \sqrt{\varkappa}\bar\rho$$

  we get a feasible solution to (5.31). Indeed, all we need is to verify that this solution satisfies, for every $\ell \le L$, constraints of (5.30). To check the semidefinite constraint, note that

  $$\left[\begin{array}{c|c} \lambda_\ell I_\nu & \frac{1}{2} R_\ell^{1/2} B \\ \hline \frac{1}{2} B^T R_\ell^{1/2} & A^T H_\ell \mathrm{Diag}\{\upsilon^\ell\} H_\ell^T A + \sum_k \mu_k^\ell T_k \end{array}\right] = \left[\begin{array}{c|c} \gamma_\ell\bar\lambda_\ell I_\nu & \frac{1}{2} R_\ell^{1/2} B \\ \hline \frac{1}{2} B^T R_\ell^{1/2} & \gamma_\ell^{-1}\left[A^T H_\ell \mathrm{Diag}\{\bar\upsilon^\ell\} H_\ell^T A + \sum_k \bar\mu_k^\ell T_k\right] \end{array}\right]$$

and the matrix in the right-hand side is $\succeq 0$ by the semidefinite constraint of (5.32) combined with $\Theta_\ell = \sum_j \bar{v}_j^\ell \mathrm{Col}_j[H_\ell]\mathrm{Col}_j^T[H_\ell]$. Furthermore, note that by construction $\sum_j \bar{v}_j^\ell \le \varkappa\varrho_\ell$, whence

$$\begin{aligned}\lambda_\ell + \phi_{\mathcal{T}}(\mu^\ell) + \sum\nolimits_j v_j^\ell &= \gamma_\ell\bar{\lambda}_\ell + \gamma_\ell^{-1}[\phi_{\mathcal{T}}(\bar{\mu}^\ell) + \varkappa\varrho_\ell] = 2\sqrt{\bar{\lambda}_\ell[\phi_{\mathcal{T}}(\bar{\mu}^\ell) + \varkappa\varrho_\ell]} \\ &\le 2\sqrt{\varkappa}\sqrt{\bar{\lambda}_\ell[\phi_{\mathcal{T}}(\bar{\mu}^\ell) + \varrho_\ell]} \le \sqrt{\varkappa}\left[\bar{\lambda}_\ell + \phi_{\mathcal{T}}(\bar{\mu}^\ell) + \varrho_\ell\right] \le \sqrt{\varkappa}\bar{\rho} = \rho\end{aligned}$$

(we have taken into account that $\varkappa \ge 1$).

We conclude that the (efficiently computable) optimal solution to the relaxed problem (5.32) can be efficiently converted to a feasible solution to problem (5.31) which is within the factor at most $\sqrt{\varkappa}$ from optimality in terms of the objective. Thus,

(!) *Given a $\varkappa$-compatible with $\mathcal{H}_\delta$ cone $\mathbf{K}$, we can find, in a computationally efficient fashion, a feasible solution to the problem of interest (5.31) with the value of the objective by at most the factor $\sqrt{\varkappa}$ greater than the optimal value of the problem.*

What we propose is to build a presumably good polyhedral estimate by applying the just outlined strategy to the instance of (5.31) associated with $\mathcal{H} = \mathcal{H}_\delta$ given by (5.26) and (5.29). The still missing—and crucial—element in this strategy is a computationally tractable cone $\mathbf{K}$ which is $(M, \varkappa)$-compatible, for some "moderate" $\varkappa$, with our $\mathcal{H}_\delta$. For the time being, we have at our disposal such a cone only for the "no uncertainty in sensing matrix" case (that is, in the case where all A_α are zero matrices), and it is shown in [32, Chapter 5] that in this case the polyhedral estimate stemming from the just outlined strategy is near minimax-optimal, provided that $\xi \sim \mathcal{N}(0, \sigma^2 I_m)$.

When "tight compatibility"—with $\varkappa$ logarithmic in the dimension of $\mathcal{H}$—is sought, the task of building a cone $(M, \varkappa)$-compatible with a given convex compact set $\mathcal{H}$ reveals to be highly nontrivial. To the best of our knowledge, for the time being, the widest family of sets $\mathcal{H}$ for which tight compatibility has been achieved is the family of ellitopes [125]. Unfortunately, this family seems to be too narrow to capture the sets $\mathcal{H}_\delta$ we are interested in now. At present, the only known to us "tractable case" here is the ball case $K = 1$, and even handling this case requires extending compatibility results of [125] from ellitopes to *spectratopes*.

Estimate synthesis utilizing cones compatible with spectratopes

Let for $S^{ij} \in \mathbf{S}^{d_i}$, $1 \le i \le I$, $1 \le j \le N$, and let for $g \in \mathbf{R}^N$, $S_i[g] = \sum_{j=1}^N g_j S^{ij}$. A *basic spectratope in* $\mathbf{R}^N$ is a set $\mathcal{H} \subset \mathbf{R}^N$ represented as

$$\mathcal{H} = \{g \in \mathbf{R}^N : \exists r \in \mathcal{R} : S_i^2[g] \preceq r_i I_{d_i}, i \le I\}; \tag{5.33}$$

here $\mathcal{R}$ is a compact convex *monotone* subset of $\mathbf{R}^I_+$ with nonempty interior, and $\sum_i S_i^2[g] \succ 0$ for all $g \ne 0$. We refer to $d = \sum_i d_i$ as *spectratopic dimension* of $\mathcal{H}$. A spectratope, by definition, is a linear image of a basic spectratope.

As shown in [32], where the notion of a spectratope was introduced, spectratopes are convex compact sets symmetric w.r.t. the origin, and basic spectratopes have nonempty interiors. The family of spectratopes is rather rich—finite intersections, direct products, linear images, and arithmetic sums of spectratopes, same as inverse images of spectratopes under linear embeddings, are spectratopes, with spectratopic representations of the results readily given by spectratopic representations of the operands.

Every ellitope is a spectratope. An example of spectratope which is important to us is the set $\mathcal{H}_\delta$ given by (5.26) and (5.28) in the "ball case" where $\mathcal{X}$ is an ellipsoid (case of $K = 1$). In this case, by one-to-one linear parameterization of signals x, accompanied for the corresponding updates in A, A_α, and B, we can assume that $T_1 = I_n$ in (5.8), so that $\mathcal{X}$ is the unit Euclidean ball,

$$\mathcal{X} = \{x \in \mathbf{R}^n : x^T x \le 1\}.$$

In this situation, denoting by $\|\cdot\|_{2,2}$ the spectral norm of a matrix, constraints (5.26) and (5.28) specify the set

$$\begin{array}{rcl} \mathcal{H}_\delta & = & \left\{h \in \mathbf{R}^m : \|h\|_2 \le (\sigma\chi(\delta))^{-1}, \|\mathcal{A}[h]\|_{2,2} \le \chi^{-1}(\delta)\right\} \\ & = & \left\{h \in \mathbf{R}^m : \exists r \in \mathcal{R} : S_j^2[h] \preceq r_j I_{d_j}, j \le 2\right\} \end{array} \tag{5.34}$$

where $\mathcal{R} = \{[r_1; r_2] : 0 \le r_1, r_2 \le 1\}$,

$$S_1[h] = \sigma\chi(\delta)\left[\begin{array}{c|c} & h \\ \hline h^T & \end{array}\right] \in \mathbf{S}^{m+1},\ S_2[h] = \chi(\delta)\left[\begin{array}{c|c} & \mathcal{A}[h] \\ \hline \mathcal{A}[h]^T & \end{array}\right] \in \mathbf{S}^{m+q}$$

with $d_1 = m + 1$, $d_2 = m + q$. We see that in the ball case $\mathcal{H}_\delta$ is a basic spectratope.

We associate with a spectratope $\mathcal{H}$, as defined in (5.33), linear mappings

$$\mathcal{S}_i[G] = \sum\nolimits_{p,q} G_{pq} S^{ip} S^{iq} : \mathbf{S}^N \to \mathbf{S}^{d_i}.$$

Note that

$$\mathcal{S}_i\left[\sum\nolimits_j g_j g_j^T\right] = \sum\nolimits_j S_i^2[g_j],\ g_j \in \mathbf{R}^N,$$

and

$$G \preceq G' \Rightarrow \mathcal{S}_i[G] \preceq \mathcal{S}_i[G'], \tag{5.35a}$$

$$\{G \succeq 0\ \&\ \mathcal{S}_i[G] = 0\ \forall \ell\} \Rightarrow G = 0. \tag{5.35b}$$

A cone "tightly compatible" with a basic spectratope is given by the following

Proposition 5.2.4 *Let $\mathcal{H} \subset \mathbf{R}^N$ be a basic spectratope*

$$\mathcal{H} = \{g \in \mathbf{R}^N : \exists r \in \mathcal{R} :\ S_i^2[g] \preceq r_i I_{d_i},\ i \le I\}$$

with "spectratopic data" $\mathcal{R}$ and $S_i[\cdot]$, $i \le I$, satisfying the requirements in the above definition.

Let us specify the closed convex cone $\mathbf{K} \subset \mathbf{S}_+^N \times \mathbf{R}_+$ as

$$\mathbf{K} = \left\{(\Sigma, \rho) \in \mathbf{S}_+^N \times \mathbf{R}_+ : \exists r \in \mathcal{R} :\ \mathcal{S}_i[\Sigma] \preceq \rho r_i I_{d_i}, i \le I\right\}.$$

Then

(i) *whenever $\Sigma = \sum_j \lambda_j g_j g_j^T$ with $\lambda_j \ge 0$ and $g_j \in \mathcal{H}$ $\forall j$, we have*

$$\left(\Sigma, \sum\nolimits_j \lambda_j\right) \in \mathbf{K},$$

(ii) *and "nearly" vice versa: when $(\Sigma, \rho) \in \mathbf{K}$, there exist (and can be found efficiently by a randomized algorithm) $\lambda_j \ge 0$ and g_j, $j \le N$, such that*

$$\Sigma = \sum\nolimits_j \lambda_j g_j g_j^T \ \text{ with } \ \sum\nolimits_j \lambda_j \le \varkappa\rho \ \text{ and } \ g_j \in \mathcal{H},\ j \le N.$$

where

$$\varkappa = 4\ln(4DN),\ D = \sum\nolimits_i d_i.$$

For the proof and for the sketch of the randomized algorithm mentioned in (ii), see Section 5.5.2 of the appendix.

Implementing the strategy

We may now summarize our approach to the design of a presumably good polyhedral estimate. By reasons outlined at the end of Section 5.2.2, the only case where the components we have developed so far admit "smooth assembling" is the one where $\mathcal{X}$ is ellipsoid which in our context w.l.o.g. can be assumed to be the unit Euclidean ball. Thus, *in the rest of this Section it is assumed that $\mathcal{X}$ is the unit Euclidean ball in $\mathbf{R}^n$*. Under this assumption the recipe, suggested by the preceding analysis, for designing presumably good polyhedral estimate is as follows. Given $\epsilon \in (0,1)$, we

• set $\delta = \epsilon/Lm$ and solve the convex optimization problem

$$\mathrm{Opt} = \min_{\substack{\Theta_\ell \in \mathbf{S}^m, \\ \varrho_\ell, \bar\lambda_\ell, \bar\mu_\ell}} \left\{ \bar\rho : \begin{array}{l} \bar\mu_\ell \geq 0,\ \Theta_\ell \succeq 0,\ \sigma^2\chi^2(\delta)\mathrm{Tr}(\Theta_\ell) \leq \varrho_\ell,\ \bar\lambda_\ell + \bar\mu_\ell + \varrho_\ell \leq \bar\rho,\ \ell \leq L, \\ \left[\begin{array}{c|c} \left[\mathrm{Tr}(A_\alpha^T \Theta_\ell A_\beta)\right]_{\alpha,\beta=1}^q & \\ \hline & \sum\limits_{\alpha,\beta} A_\alpha^T \Theta_\ell A_\beta \end{array}\right] \preceq \chi^{-2}(\delta)\varrho_\ell I_{q+n},\ \ell \leq L, \\ \left[\begin{array}{c|c} \bar\lambda_\ell I_\nu & \frac{1}{2}R_\ell^{1/2} B \\ \hline \frac{1}{2}B^T R_\ell^{1/2} & A^T\Theta_\ell A + \bar\mu_\ell I_n \end{array}\right] \succeq 0, \ell \leq L \end{array} \right\} \tag{5.36}$$

—this is what under the circumstances becomes problem (5.32) with the cone $\mathbf{K}$ given by Proposition 5.2.4 as applied to the spectratope $\mathcal{H}_\delta$ given by (5.34). Note that by Proposition 5.2.4, $\mathbf{K}$ is $\varkappa$-compatible with $\mathcal{H}_\delta$, with

$$\varkappa = 4\ln(4m(m+n+q+1)). \tag{5.37}$$

For instance, in the case of rank 1 matrices $A_\alpha = f_\alpha g_\alpha^T$ and $\|\cdot\| = \|\cdot\|_2$ (5.36) becomes

$$\mathrm{Opt} = \min_{\substack{\Theta \in \mathbf{S}^m, \\ \varrho, \bar\lambda, \bar\mu}} \left\{ \bar\rho : \begin{array}{l} \bar\mu \geq 0,\ \Theta \succeq 0,\ \sigma^2\chi^2(\delta)\mathrm{Tr}(\Theta) \leq \varrho,\ \bar\lambda + \bar\mu + \varrho \leq \bar\rho \\ \left[\begin{array}{c|c} \left[(f_\alpha^T \Theta f_\beta) g_\alpha^T g_\beta\right]_{\alpha,\beta=1}^q & \\ \hline & \left[\sum_{\alpha,\beta=1}^q [f_\alpha^T \Theta f_\beta]\, g_\alpha g_\beta^T\right] \end{array}\right] \preceq \chi^{-2}(\delta)\varrho I_{q+n} \\ \left[\begin{array}{c|c} \bar\lambda_\ell I_\nu & \frac{1}{2}B \\ \hline \frac{1}{2}B^T & A^T\Theta A + \bar\mu I_n \end{array}\right] \succeq 0 \end{array} \right\}; \tag{5.38}$$

• use the randomized algorithm described in the proof of Proposition 5.2.4 to convert the Θ_ℓ-components of the optimal solution to (5.36) into a contrast matrix. Specifically,

1. for $\ell = 1, 2, ..., L$ we generate matrices $G_\varsigma^k = \Theta_\ell^{1/2}\mathrm{Diag}\{\varsigma^k\}O$, $k = 1, ..., K$, where O is the orthonormal matrix of $m \times m$ Discrete Cosine Transform, and ς^k are i.i.d. realizations of m-dimensional Rademacher random vector;

2. for every $k \leq K$, we compute the maximum $\theta(G_\ell^k)$ of values of the Minkowski function of $\mathcal{H}_\delta$ as evaluated at the columns of G_ℓ^k, with $\mathcal{H}_\delta$ given by (5.26), (5.28), and select among G_ℓ^k matrix G_ℓ with the smallest value of $\theta(G_\ell^k)$.
   Then the ℓ-th block of the contrast matrix we are generating is $H_\ell = G_\ell\theta^{-1}(G_\ell)$.

With reliability $1 - 2^{-K}L$ the resulting contrast matrix H (which definitely has all columns in $\mathcal{H}_\delta$) is, by (!), near-optimal, within factor $\sqrt{\varkappa}$ in terms of the objective, solution to (5.31), and the ϵ-risk of the associated polyhedral estimate is upper-bounded by $2\sqrt{\varkappa}\mathrm{Opt}$ with Opt given by (5.36).

In Figure 5.2 we present error distributions and upper risk bounds (horizontal bar) of linear and polyhedral estimates in the numerical experiment with the model described in Section 5.2.1. In the

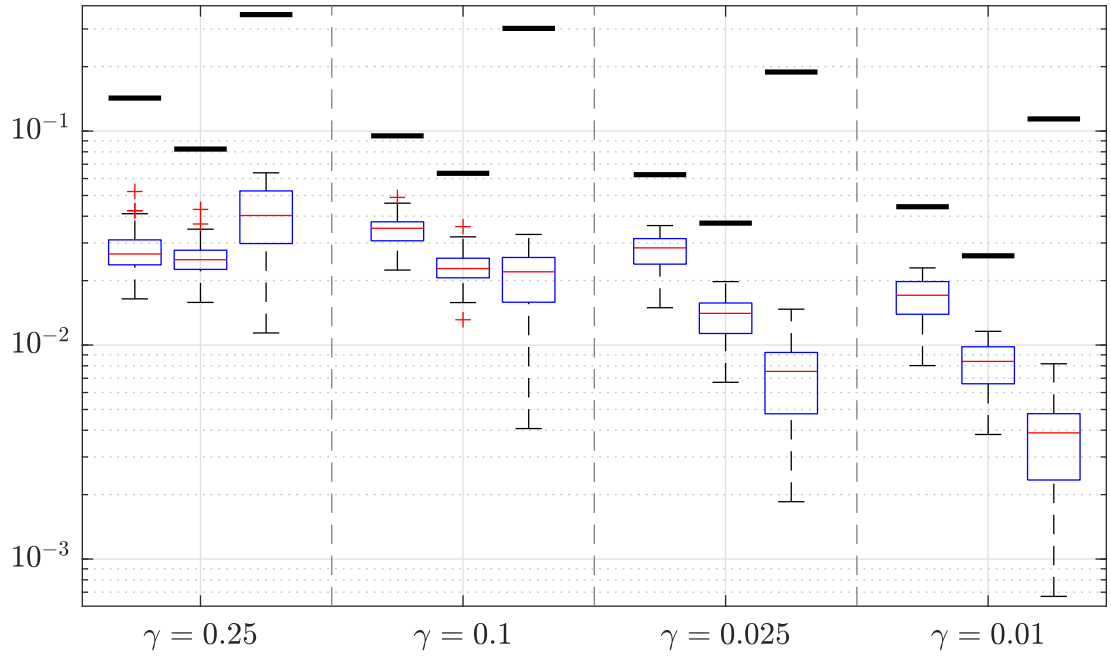


Figure 5.2: Distributions of ℓ_2-recovery errors and upper bounds of the robust linear and robust polyhedral estimates for different values of γ parameter.

plot cells, from left to right: (1) robust linear estimate by Proposition 5.2.1 and upper bound $\mathfrak{R}$ on its 0.05-risk, (2) robust linear estimate $w_1(\omega_1)$ yielded by Proposition 5.2.2 and upper bound $\widetilde{\mathfrak{R}}_1$ on its expected error risk, (3) robust polyhedral estimate by Proposition 5.2.4 and upper bound on its 0.05-risk.

A modification

So far, our considerations related to polyhedral estimates were restricted to the case of sub-Gaussian η and ξ. Similarly to what was done in Section 5.2.1, we are about to show that passing from observation (5.2) to its K-repeated, with "moderate" K, version (cf. (5.13))

$$\omega^K = \{\omega_k = A[\eta_k]x + \xi_k, \quad k = 1, ..., K\}$$

with pairs (η_k, ξ_k) independent across k, we can relax the sub-Gaussianity assumption replacing it with moment condition (5.14). Specifically, let us set

$$\mathcal{H} = \left\{h \in \mathbf{R}^m : \sigma\|h\|_2 \leq \tfrac{1}{8}, \|\mathcal{A}[h]\|_{\mathcal{X},2} \leq \tfrac{1}{8}\right\}, \ \mathcal{A}[h] = [h^T A_1; ...h^T A_q]$$

(cf. (5.26) and (5.28)).

Given tolerance ϵ an $m \times M$ contrast matrix H with columns $h_j \in \mathcal{H}$, and observation (5.13), we build the polyhedral estimate as follows.[4]

[4]Readers acquainted with the literature on robust estimation will immediately recognize that the proposed construction is nothing but a reformulation of the celebrated "median-of-means" estimate of [56] (see also [126–129]) for our purposes.

1. For $j = 1, ..., M$ we compute empirical medians y_j of the data $h_j^T \omega_k$, $k = 1, ..., K$,

$$y_j = \text{median}\{h_j^T \omega_k,\ 1 \le k \le K\}.$$

2. We specify $\widehat{x}^H(\omega^K)$ as a point from $\text{Argmin}_{u \in \mathcal{X}} \|y - H^T A u\|_\infty$ and use, as the estimate of Bx, the vector $\widehat{w}^H_{\text{poly}}(\omega^K) = B\widehat{x}^H(\omega^K)$.

Lemma 3 *In the situation of this section, let ξ_k and η_k satisfy moment constraint of (5.14), and let $K \ge \overline{\kappa} = 2.5 \ln[M/\epsilon]$. Then estimate $\widehat{w}^H_{\text{poly}}(\omega^K)$ satisfies*

$$\text{Risk}_\epsilon[\widehat{w}^H_{\text{poly}}(\omega^K)|\mathcal{X}] \le \mathfrak{p}[H]$$

(cf. (5.23)).

As an immediate consequence of the result of Lemma 3, the constructions and results of Sections 5.2.2–5.2.2 apply, with $\chi(\delta) = 8$ and $\mathcal{H}$ in the role of $\mathcal{H}_\delta$, to our present situation in which the sub-Gaussianity of ξ, η is relaxed to the second moment condition (5.14) and instead of single observation ω, we have access to a "short"—with K logarithmic in M/ϵ—sample of K independent realizations of ω.

5.3 Uncertain-but-bounded perturbations

In this section we assume that perturbation vector η in (5.2) is deterministic and runs through a given *uncertainty set* $\mathcal{U}$, so that (5.2) becomes

$$\omega = A[\eta]x + \xi,\ \ A[\eta] = A + D[\eta], \tag{5.39}$$

where $D[\eta]$ is (homogeneous) linear matrix-valued function of perturbation η running through $\mathcal{U}$. As about observation noise ξ, we still assume that its distribution P_x (which may depend on x) satisfies (5.5), i.e., is sub-Gaussian with zero mean and sub-Gaussian matrix parameter $\sigma^2 I_m$ for every $x \in \mathcal{X}$.

In our present situation it is natural to redefine the notion of the ϵ-risk of an estimate $\omega \mapsto \widehat{x}(\omega)$: here we consider uniform over $x \in \mathcal{X}$ and $\eta \in \mathcal{U}$ ϵ-risk

$$\text{Risk}_\epsilon[\widehat{w}|\mathcal{X}] = \sup_{x \in \mathcal{X}, \eta \in \mathcal{U}} \inf \Big\{\rho : \text{Prob}_{\xi \sim P_x}\{\|\widehat{w}(A[\eta]x + \xi) - Bx\| > \rho\} \le \epsilon\Big\}.$$

Besides this, we, as before, assume that

$$\|y\| = \max_{\ell \le L} \sqrt{y^T R_\ell y} \qquad [R_\ell \succeq 0, \textstyle\sum_\ell R_\ell \succ 0]$$

5.3.1 Design of presumably good linear estimate

Observe that the error of the linear estimate $\widehat{w}^H(\omega) = H^T\omega$ satisfies

$$\|\widehat{w}(A[\eta]x + \xi) - Bx\| \le \|H^T\xi\| + \max_{x \in \mathcal{X}, \eta \in \mathcal{U}} \big\|H^T D[\eta]x\big\| + \max_{x \in \mathcal{X}} \|[B - H^T A]x\| \tag{5.40}$$

Similarly to what was done in Section 5.2.1, design of a presumably good linear estimate $\widehat{x}_H(\omega)$ consists in minimizing over H the sum of tight efficiently computable upper bounds on the terms

in the right-hand side of (5.40). Recall that bounds on the first and the last term were already established in Section 5.2.1 (cf. (5.58) and (5.59) in the proof of Proposition 5.2.1). What is missing is a tight upper bound on

$$\mathfrak{s}(H) = \max_{x\in\mathcal{X},\eta\in\mathcal{U}} \left\|H^T D[\eta]x\right\|.$$

In the rest of this section we focus on building efficiently computable upper bound on $\mathfrak{s}(H)$ which is convex in H; the synthesis of the contrast H is then conducted by minimizing with respect to H the resulting upper bound on estimation risk.

We assume from now on that $\mathcal{U}$ is a convex compact set in certain $\mathbf{R}^q$. In this case $\mathfrak{s}(H)$ is what in [118] was called the *robust norm*

$$\|\mathcal{Z}[H]\|_{\mathcal{X}} = \max_{Z\in\mathcal{Z}[H]} \|Z\|_{\mathcal{X}},\ \|Z\|_{\mathcal{X}} = \max_{x\in\mathcal{X}} \|Zx\|$$

of the *uncertain $\nu\times n$ matrix*

$$\mathcal{Z}[H] = \{Z = H^T D[\eta] : \eta\in\mathcal{U}\},$$

i.e., the maximum, over *instances* $Z\in\mathcal{Z}[H]$, of operator norms of the linear mappings $x\mapsto Zx$ induced by the norm with the unit ball $\mathcal{X}$ on the argument space and the norm $\|\cdot\|$ on the image space.

It is well known that aside of a very restricted family of special cases, robust norms do not allow for efficient computation. We are about to list known to us generic cases when these norms admit efficiently computable upper bounds which are tight within logarithmic factors.

Scenario uncertainty

This is the case where the nuisance set $\mathcal{U} = \mathrm{Conv}\{\eta^1,...,\eta^S\}$ is given as a convex hull of moderate number of scenarios η^s. In this case, $\mathfrak{s}(H)$ the maximum of operator norms:

$$\mathfrak{s}(H) = \max_{s\le S}\max_{x\in\mathcal{X}} \|H^T D[\eta^s]x\| = \max_{s\le S,\ell\le L} \|\mathcal{M}_{s\ell}[H]\|_{\mathcal{X},2},\quad \mathcal{M}_{s\ell}[H] = R_\ell^{1/2} H^T D[\eta^s],$$

where, for $Q\in\mathbf{R}^{\nu\times n}$, $\|Q\|_{\mathcal{X},2} = \max_{x\in\mathcal{X}}\|Qx\|_2$ is the operator norm of the linear mapping $x\mapsto Qx : \mathbf{R}^n\to\mathbf{R}^\nu$ induced by the norm $\|\cdot\|_{\mathcal{X}}$ with the unit ball $\mathcal{X}$ on the argument space, and the Euclidean norm $\|\cdot\|_2$ on the image space. Note that this norm is efficiently computable in the *ellipsoid case* where $\mathcal{X} = \{x\in\mathbf{R}^n : x^T Tx\le 1\}$ with $T\succ 0$ (that is, for $K=1$, $T_1 = T$, $\mathcal{T} = [0,1]$ in (5.8))—one has $\|Q\|_{\mathcal{X},2} = \|QT^{-1/2}\|_{2,2}$. When $\mathcal{X}$ is a general ellitope, norm $\|\cdot\|_{\mathcal{X},2}$ is difficult to compute. However, it admits a tight efficiently computable convex in Q upper bound:[5] it is shown in [118, Theorem 3.1] that function

$$\mathrm{Opt}[Q] = \min_{\lambda,\mu}\left\{\lambda + \phi_{\mathcal{T}}(\mu) : \mu\ge 0,\ \left[\begin{array}{c|c}\lambda I_\nu & \frac{1}{2}Q\\ \hline \frac{1}{2}Q^T & \sum_k \mu_k T_k\end{array}\right]\succeq 0\right\}$$

satisfies $\|Q\|_{\mathcal{X},2}\le \mathrm{Opt}[Q]\le 2.4\sqrt{\ln(4K)}\|Q\|_{\mathcal{X},2}$. As a result, under the circumstances,

$$\begin{aligned}\overline{\mathfrak{s}}(H) &= \max_{s\le S,\ell\le L}\mathrm{Opt}_{s\ell}[H],\\ \mathrm{Opt}_{s\ell}[H] &= \min_{\lambda_\ell,\mu^\ell}\left\{\lambda_\ell + \phi_{\mathcal{T}}(\mu^\ell) : \mu^\ell\ge 0,\ \left[\begin{array}{c|c}\lambda_\ell I_\nu & \frac{1}{2}R_\ell^{1/2}H^T D[\eta^s]\\ \hline \frac{1}{2}D^T[\eta^s]HR_\ell^{1/2} & \sum_k \mu^\ell{}_k T_k\end{array}\right]\succeq 0\right\},\end{aligned}$$

is a tight within the factor $2.4\sqrt{\ln(4K)}$ efficiently computable convex in H upper bound on $\mathfrak{s}(H)$.

[5] We have already used it in the proof of Proposition 5.2.1 when upper-bounding the corresponding terms $s_\ell(H)$ in the case of random uncertainty.

Box and structured norm-bounded uncertainty

In the case of *structured norm-bounded uncertainty* function $D[\eta]$ in the model (5.39) is of the form

$$\begin{array}{rl} D[\eta] = \sum_{\alpha=1}^{q} P_\alpha^T \eta_\alpha Q_\alpha & [P_\alpha \in \mathbf{R}^{p_\alpha \times m}, Q_\alpha \in \mathbf{R}^{q_\alpha \times n}], \\ \mathcal{U} = \{\eta = (\eta_1, ..., \eta_q)\} = \mathcal{U}_1 \times ... \times \mathcal{U}_q, & \\ \mathcal{U}_\alpha = \left\{ \begin{array}{ll} \{\eta_\alpha = \delta I_{p_\alpha} : |\delta| \le 1\} \subset \mathbf{R}^{p_\alpha \times p_\alpha}, q_\alpha = p_\alpha & , \alpha \le q_{\rm s}, \quad [\text{"scalar perturbation blocks"}] \\ \{\eta_\alpha \in \mathbf{R}^{p_\alpha \times q_\alpha} : \|\eta_\alpha\|_{2,2} \le 1\} & , q_{\rm s} < \alpha \le q. \quad [\text{"general perturbation blocks"}] \end{array} \right. \end{array} \tag{5.41}$$

The special case of (5.41) where $q_{\rm s} = q$, that is,

$$\mathcal{U} = \{\eta \in \mathbf{R}^q : \|\eta\|_\infty \le 1\} \ \&\ A[\eta] = A + D[\eta] = A + \sum_{\alpha=1}^{q} \eta_\alpha A_\alpha$$

is referred to as *box uncertainty*. In this section we operate with structured norm-bounded uncertainty (5.41), assuming w.l.o.g. that all P_α are nonzero. The main result here (for underlying rationale and proof, see Section 5.6.2) is as follows:

Proposition 5.3.1 *Let $\mathcal{X} \subset \mathbf{R}^n$ be an ellitope: $\mathcal{X} = P\mathcal{Y}$, where*

$$\mathcal{Y} = \{y \in \mathbf{R}^n : \exists t \in \mathcal{T} : y^T T_k y \le t_k, k \le K\}$$

is a basic ellitope. Given the data of structured norm-bounded uncertainty (5.41), consider the efficiently computable convex function

$$\begin{array}{rl} \overline{\mathfrak{s}}(H) = \max\limits_{\ell \le L} \mathrm{Opt}_\ell(H), & \\ \mathrm{Opt}_\ell(H) = \min\limits_{\mu,\upsilon,\lambda,U_s,V_s,U^t,V^t} & \left\{ \frac{1}{2}[\mu + \phi_{\mathcal{T}}(\upsilon)] : \ \mu \ge 0, \upsilon \ge 0, \lambda \ge 0 \right. \\ & \left[\begin{array}{c|c} U_s & -A_{s\ell}[H]P \\ \hline -P^T A_{s\ell}^T[H] & V_s \end{array} \right] \succeq 0, \ s \le q_{\rm s}, \ \left[\begin{array}{c|c} U^t & -L_{t\ell}^T[H] \\ \hline -L_{t\ell}[H] & \lambda_t I_{p_{q_{\rm s}+t}} \end{array} \right] \succeq 0, \ t \le q - q_{\rm s} \\ & V^t - \lambda_t P^T R_t^T R_t P \succeq 0, \ t \le q - q_{\rm s} \\ & \left. \mu I_\nu - \sum_s U_s - \sum_t U^t \succeq 0, \ \ \sum_k \upsilon_k T_k - \sum_s V_s - \sum_t V^t \succeq 0 \right\} \end{array}$$

where

$$\begin{array}{l} A_{s\ell}[H] = R_\ell^{1/2} H^T P_s^T Q_s, \ 1 \le s \le q_{\rm s} \\ L_{t\ell}[H] = P_{q_{\rm s}+t} H R_\ell^{1/2}, \ R_t = Q_{q_{\rm s}+t}, \ 1 \le t \le q - q_{\rm s}. \end{array}$$

Then

$$\mathfrak{s}(H) \le \overline{\mathfrak{s}}(H) \le \varkappa(K) \max[\vartheta(2\kappa), \pi/2] \mathfrak{s}(H),$$

where $\kappa = \max\limits_{\alpha \le q_{\rm s}} \min[p_\alpha, q_\alpha]$ ($\kappa = 0$ when $q_{\rm s} = 0$),

$$\varkappa(K) = \left\{ \begin{array}{ll} 1, & K = 1, \\ \frac{5}{2}\sqrt{\ln(2K)}, & K > 1, \end{array} \right.$$

and $\vartheta(k)$ is a universal function of integer $k \ge 0$ specified in (5.75) such that

$$\vartheta(0) = 0, \ \vartheta(1) = 1, \ \vartheta(2) = \pi/2, \ \vartheta(3) = 1.7348..., \ \vartheta(4) = 2, \ \ \vartheta(k) \le \tfrac{1}{2}\pi\sqrt{k}, \ \ k \ge 1.$$

Note that the "box uncertainty" version of Proposition 5.3.1 was derived in [118].

Robust estimation of linear forms

Until now, we imposed no restrictions on the matrix B. We are about to demonstrate that when aiming to recover the value of a given linear form b^Tx of signal $x \in \mathcal{X}$, i.e., when B is a row vector:

$$Bx = b^Tx \qquad [b \in \mathbf{R}^n], \tag{5.42}$$

we can handle much wider family of uncertainty sets $\mathcal{U}$ than those considered so far. Specifically, assume on the top of (5.42) that $\mathcal{U}$ is a spectratope:

$$\begin{array}{c}\mathcal{U} = \{\eta = Qv, v \in \mathcal{V}\},\ \mathcal{V} = \{v \in \mathbf{R}^M : \exists s \in \mathcal{S} : S_\ell^2[v] \preceq s_\ell I_{d_\ell},\ \ell \le L\},\\ S_\ell[v] = \sum_{i=1}^M v_i S^{i\ell},\ S^{i\ell} \in \mathbf{S}^{d_\ell}\end{array} \tag{5.43}$$

(as is the case, e.g., with structured norm-bounded uncertainty) and let $\mathcal{X}$ be a spectratope as well:

$$\begin{array}{c}\mathcal{X} = \{x = Py, y \in \mathcal{Y}\},\ \mathcal{Y} = \{y \in \mathbf{R}^N : \exists t \in \mathcal{T} : T_k^2[y] \preceq t_k I_{f_k},\ k \le K\},\\ T_k[y] = \sum_{j=1}^N y_j T^{jk},\ T^{jk} \in \mathbf{S}^{f_k}.\end{array} \tag{5.44}$$

The contrast matrix H underlying a candidate linear estimate becomes a vector $h \in \mathbf{R}^m$, the associated linear estimate being $\widehat{w}_h(\omega) = h^T\omega$. In our present situation $\nu = 1$ we lose nothing when setting $\|\cdot\| = |\cdot|$. Representing $D[\eta]$ as $\sum_{\alpha=1}^q \eta_\alpha A_\alpha$, we get

$$\mathfrak{r}_b(h) = \max_{x\in\mathcal{X},\eta\in\mathcal{U}} \left|h^T\sum\nolimits_\alpha \eta_\alpha A_\alpha x\right| = \max_{\eta\in\mathcal{U},x\in\mathcal{X}} \eta^T A[h]x, \quad A[h] = [h^TA_1; ...; h^TA_q].$$

In other words, $\mathfrak{r}_b(h)$ is the operator norm $\|A[h]\|_{\mathcal{X},\mathcal{U}_*}$ of the linear mapping $x \mapsto A[h]x$ induced by the norm $\|\cdot\|_{\mathcal{X}}$ with the unit ball $\mathcal{X}$ on the argument space and the norm with the unit ball $\mathcal{U}_*$—the polar of the spectratope $\mathcal{U}$—on the image space. Denote

$$\begin{array}{l}\lambda[\Lambda] = [\mathrm{Tr}(\Lambda_1); ...; \mathrm{Tr}(\Lambda_K)],\ \ \Lambda_k \in \mathbf{S}^{f_k},\\ \lambda[\Upsilon] = [\mathrm{Tr}(\Upsilon_1); ...; \mathrm{Tr}(\Upsilon_L)],\ \ \Upsilon_\ell \in \mathbf{S}^{d_\ell},\end{array}$$

and for $Y \in \mathbf{S}^{d_\ell}$ and $X \in \mathbf{S}^{f_k}$

$$R_\ell^{+,*}[Y] = \left[\mathrm{Tr}(YR^{i\ell}R^{j\ell})\right]_{i,j\le M}, \quad T_k^{+,*}[X] = \left[\mathrm{Tr}(XT^{ik}T^{jk})\right]_{i,j\le N}.$$

Invoking [118, Theorem 7], we arrive at

Proposition 5.3.2 *In the case of (5.43) and (5.44), efficiently computable convex function*

$$\overline{\mathfrak{r}}_b(h) = \min_{\Lambda,\Upsilon}\left\{\tfrac{1}{2}[\phi_{\mathcal{T}}(\lambda[\Lambda]) + \phi_{\mathcal{S}}(\lambda[\Upsilon]) : \begin{array}{c}\Lambda = \{\Lambda_k \in \mathbf{S}_+^{f_k}, k \le K\}, \Upsilon = \{\Upsilon_\ell \in \mathbf{S}_+^{d_\ell}, \ell \le L\}\\ \left[\begin{array}{c|c}\sum_\ell R_\ell^{+,*}[\Upsilon_\ell] & Q^TA[h]P\\ \hline P^TA^T[h]Q & \sum_k T_k^{+,*}[\Lambda_k]\end{array}\right] \succeq 0\end{array}\right\} \tag{5.45}$$

is a reasonably tight upper bound on $\mathfrak{r}_b(h)$*:*

$$\mathfrak{r}_b(h) \le \overline{\mathfrak{r}}_b(h) \le \overline{\varsigma}\left(\sum\nolimits_{k=1}^K f_k\right)\ \ \overline{\varsigma}\left(\sum\nolimits_{\ell=1}^L d_\ell\right)\mathfrak{r}_b(h)$$

where $\overline{\varsigma}(J) = \sqrt{2\ln(5J)}$.

5.3.2 Design of the robust polyhedral estimate

On a close inspection, the strategy for designing a presumably good polyhedral estimate developed in Section 5.2.2 for the case of random uncertainty works in the case of uncertain-but-bounded perturbations $A[\eta] = A + \underbrace{\sum_\alpha \eta_\alpha A_\alpha}_{D[\eta]}$, $\eta \in \mathcal{U}$, provided that the constraints (5.25) on the allowed columns h of the contrast matrices are replaced with the constraint

$$\mathrm{Prob}_\xi\{|h^T\xi| > 1/2\} \le \delta/2, \tag{5.46a}$$

$$\left|\sum_{\alpha=1}^q [h^T A_\alpha x]\eta_\alpha\right| \le 1/2 \ \forall (x \in \mathcal{X}, \eta \in \mathcal{U}). \tag{5.46b}$$

Assuming that $\mathcal{U}$ and $\mathcal{X}$ are the spectratopes (5.43), (5.44) and invoking Proposition 5.3.2, an efficiently verifiable sufficient condition for h to satisfy the constraints (5.46) is

$$\|h\|_2 \le 2\sigma\sqrt{2\ln(2/\delta)} \ \text{ and } \ \bar{\mathfrak{r}}_b(h) \le 1/2 \tag{5.47}$$

(see (5.26), (5.45)). It follows that in order to build an efficiently computable upper bound for the ϵ-risk of a polyhedral estimate associated with a given $m \times ML$ contrast matrix $H = [H_1, .., H_L]$, $H_\ell \in \mathbf{R}^{m\times M}$, it suffices to check whether the columns of H satisfy constraints (5.47) with $\delta = \epsilon/ML$. If the answer is positive, one can upper-bound the risk utilizing the following spectratopic version of Proposition 5.2.3:

Proposition 5.3.3 *In the situation of this section, let $\epsilon \in (0,1)$, and let $H = [H_1, ..., H_L]$ be $m \times ML$ matrix with L blocks $H_\ell \in \mathbf{R}^{m\times M}$ such that all columns of H satisfy (5.47) with $\delta = \epsilon/ML$. Consider optimization problem*

$$\mathfrak{p}_+[H] = 2 \min_{\lambda_\ell, \Upsilon^\ell, \upsilon^\ell, \rho} \left\{ \rho : \ \upsilon^\ell \ge 0,\ \Upsilon^\ell = \{\Upsilon_k^\ell \in \mathbf{S}_+^{f_k}, k \le K\},\ \ell \le L \right. \tag{5.48}$$

$$\lambda_\ell + \phi_{\mathcal{T}}(\lambda[\Upsilon^\ell]) + \textstyle\sum_{j=1}^M \upsilon_j^\ell \le \rho,\ \ell \le L$$

$$\left. \left[\begin{array}{c|c} \lambda_\ell I_\nu & \frac{1}{2}R_\ell^{1/2} BP \\ \hline \frac{1}{2}P^T B^T R_\ell^{1/2} & P^T A^T H_\ell \mathrm{Diag}\{\upsilon^\ell\} H_\ell^T AP + \sum_k T_k^{+,*}[\Upsilon_k^\ell] \end{array}\right] \succeq 0,\ \ell \le L \right\}$$

where

$$\lambda[\Upsilon^\ell] = [\mathrm{Tr}(\Upsilon_1^\ell); ...; \mathrm{Tr}(\Upsilon_K^\ell)], \ \ \textit{and} \ \ T_k^{+,*}(V) = \left[\mathrm{Tr}(VT^{ik}T^{jk})\right]_{1\le i,j\le N} \ \textit{for } V \in \mathbf{S}^{f_k}.$$

Then

$$\mathrm{Risk}_\epsilon[\widehat{w}^H|\mathcal{X}] \le \mathfrak{p}_+[H].$$

Remarks. As it was already explained, when taken together, Propositions 5.3.2 and 5.3.3 allow to compute efficiently an upper bound on the ϵ-risk of the polyhedral estimate associated with a given $m \times ML$ contrast matrix H: when the columns of H satisfy (5.47) with $\delta = \epsilon/ML$, this bound is $\mathfrak{p}_+[H]$, otherwise it is, say, $+\infty$. The outlined methodology can be applied to *any* pair of spectratopes $\mathcal{X}$, $\mathcal{Y}$. However, to design a presumably good polyhedral estimate, we need to optimize the risk bound obtained in H, and this seems to be difficult because the bound, same as its "random perturbation" counterpart, is nonconvex in H. At present, we know only one generic situation where the synthesis problem admits "presumably good" solution—the case where both

$\mathcal{X}$ and $\mathcal{U}$ are ellipsoids. Applying appropriate one-to-one linear transformations to perturbation η and signal x, the latter situation can be reduced to that with

$$\mathcal{X} = \{x \in \mathbf{R}^n : \|x\|_2 \le 1\}, \;\; \mathcal{U} = \{\eta \in \mathbf{R}^q : \|\eta\|_2 \le 1\}, \tag{5.49}$$

which we assume till the end of this section. In this case (5.47) reduces to

$$\|h\|_2 \le [2\sigma\sqrt{2\ln(2/\delta)}]^{-1} \;\; \text{and} \;\; \|\mathcal{A}[h]\|_{2,2} \le 1/2 \tag{5.50}$$

where the matrix $\mathcal{A}[h]$ is given by (5.27). Note that (5.50) is nothing but the constraint (5.29) where the ellitope $\mathcal{X}$ is set to be the unit Euclidean ball (that is, when $K = 1$, $T_1 = I_n$, and $\mathcal{T} = [0;1]$ in (5.8)) and the right hand side $\chi^{-1}(\delta)$ in the constraint is replaced with $1/2$. As a result, (5.46) can be processed in the same fashion as constraints (5.26) and (the single-ellipsoid case of) (5.29) were processed in Sections 5.2.2 and 5.2.2 to yield a computationally efficient scheme for building a presumably good, in the case of (5.49), polyhedral estimate. This scheme is the same as that described at the end of Section 5.2.2 with just one difference: the quantity $\chi(\delta)$ in the first semidefinite constraint of (5.36) and (5.38) should now be replaced with constant 2. Denoting by Opt the optimal value of the modified in the way just explained problem (5.36), the ϵ-risk of the polyhedral estimate yielded by an optimal solution to the problem is upper-bounded by $2\sqrt{\varkappa}\mathrm{Opt}$, with $\varkappa$ given by (5.37).

5.4 Proofs for Section 5.2.1

5.4.1 Preliminaries: concentration of quadratic forms of sub-Gaussian vectors

For the reader's convenience, we recall in this section some essentially known bounds for deviations of quadratic forms of sub-Gaussian random vectors (cf., e.g., [78, 88, 89]).

1^o. Let ξ be a d-dimensional normal vector, $\xi \sim \mathcal{N}(\mu, \Sigma)$. For all $h \in \mathbf{R}^d$ and $G \in \mathbf{S}^d$ such that $G \prec \Sigma^{-1}$ we have the well known relationship:

$$\begin{aligned}\ln\left(\mathbf{E}_\xi\left\{e^{h^T\xi+\frac{1}{2}\xi^TG\xi}\right\}\right) &= -\tfrac{1}{2}\ln\mathrm{Det}(I-\Sigma^{1/2}G\Sigma^{1/2}) \\ &\quad + h^T\mu + \tfrac{1}{2}\mu^TG\mu + \tfrac{1}{2}[G\mu+h]^T\Sigma^{1/2}(I-\Sigma^{1/2}G\Sigma^{1/2})^{-1}\Sigma^{1/2}[G\mu-h].\end{aligned} \tag{5.51}$$

Now, suppose that $\eta \sim \mathcal{SG}(0, \Sigma)$ where $\Sigma \in \mathbf{S}^d_+$, let also $h \in \mathbf{R}^d$ and $S \in \mathbf{R}^{d\times d}$ such that $S\Sigma S^T \prec I$. Then for $\xi \sim \mathcal{N}(h, S^TS)$ one has

$$\mathbf{E}_\eta\left\{e^{h^T\eta+\frac{1}{2}\eta^TS^TS\eta}\right\} = \mathbf{E}_\eta\left\{\mathbf{E}_\xi\left\{e^{\eta^T\xi}\right\}\right\} = \mathbf{E}_\xi\left\{\mathbf{E}_\eta\left\{e^{\eta^T\xi}\right\}\right\} \le \mathbf{E}_\xi\left\{e^{\frac{1}{2}\xi^T\Sigma\xi}\right\},$$

so that

$$\begin{aligned}\ln\left(\mathbf{E}_\eta\left\{e^{h^T\eta+\frac{1}{2}\eta^TS^TS\eta}\right\}\right) &\le \ln\left(\mathbf{E}_\xi\left\{e^{\frac{1}{2}\xi^T\Sigma\xi}\right\}\right) \\ &= -\tfrac{1}{2}\ln\mathrm{Det}(I-S\Sigma S^T) + \tfrac{1}{2}h^T\Sigma h + \tfrac{1}{2}h^T\Sigma S^T(I-S\Sigma S^T)^{-1}S\Sigma h \\ &= -\tfrac{1}{2}\ln\mathrm{Det}(I-S\Sigma S^T) + \tfrac{1}{2}h^T\Sigma^{1/2}(I-S\Sigma S^T)^{-1}\Sigma^{1/2}h.\end{aligned}$$

In particular, when $\zeta \sim \mathcal{SG}(0, I)$, one has

$$\ln\left(\mathbf{E}_\zeta\left\{e^{h^T\zeta+\frac{1}{2}\zeta^TG\zeta}\right\}\right) \le -\tfrac{1}{2}\ln\mathrm{Det}(I-G) + \tfrac{1}{2}h^T(I-G)^{-1}h =: \Phi(h,G).$$

Observe that $\Phi(h,G)$ is convex and continuous in $h \in \mathbf{R}^d$ and $0 \preceq G \prec I$ on its domain. Using the inequality (cf. [65, Lemma 1])

$$\forall v \in [0,1[\quad -\ln(1-v) \le v + \frac{v^2}{2(1-v)}, \tag{5.52}$$

we get

$$\Phi(h,G) \le \tfrac{1}{2}\mathrm{Tr}[G] + \tfrac{1}{4}\mathrm{Tr}[G(I-G)^{-1}G] + \tfrac{1}{2}h^T(I-G)^{-1}h =: \widetilde{\Phi}(h,G).$$

Finally, using

$$\mathrm{Tr}[G(I-G)^{-1}G] \le (1-\lambda_{\max}(G))^{-1}\mathrm{Tr}[G^2], \quad h^T(I-G)^{-1}h \le (1-\lambda_{\max}(G))^{-1}h^Th,$$

we arrive at

$$\widetilde{\Phi}(h,G) \le \tfrac{1}{2}\mathrm{Tr}[G] + \tfrac{1}{4}(1-\lambda_{\max}(G))^{-1}(\mathrm{Tr}[G^2] + 2\|h\|_2^2) =: \overline{\Phi}(h,G).$$

2^o. In the above setting, let $Q \in \mathbf{S}^d_+$, $\alpha > 2\lambda_{\max}(Q)$, $G = 2Q/\alpha$, and let $h = 0$. By the Cramer argument we conclude that

$$\mathrm{Prob}\left\{\zeta^TQ\zeta \ge \alpha[\Phi(2Q/\alpha) + \ln\epsilon^{-1}]\right\} \le \epsilon \tag{5.53}$$

where $\Phi(\cdot) = \Phi(0,\cdot)$. In particular,

$$\mathrm{Prob}\left\{\zeta^TQ\zeta \ge \min_{\alpha > 2\lambda_{\max}(Q)} \alpha[\Phi(2Q/\alpha) + \ln\epsilon^{-1}]\right\} \le \epsilon \tag{5.54}$$

Clearly, similar bounds hold with Φ replaced with $\widetilde{\Phi}$ and $\overline{\Phi}$. For instance,

$$\mathrm{Prob}\left\{\zeta^TQ\zeta \ge \alpha[\overline{\Phi}(2Q/\alpha) + \ln\epsilon^{-1}]\right\} \le \epsilon,$$

so, when choosing $\alpha = 2\lambda_{\max}(Q) + \sqrt{\frac{\mathrm{Tr}(Q^2)}{\ln\epsilon^{-1}}}$ we arrive at the "standard bound"

$$\mathrm{Prob}\left\{\zeta^TQ\zeta \ge \mathrm{Tr}(Q) + 2\|Q\|_{\mathrm{Fro}}\sqrt{\ln\epsilon^{-1}} + 2\lambda_{\max}(Q)\ln\epsilon^{-1}\right\} \le \epsilon. \tag{5.55}$$

Corollary 5.4.1 *Let $\epsilon \in (0,1)$, $W_1,...,W_L$ be matrices from $\mathbf{S}^d_+$, and let $v \sim \mathcal{SG}(0,V)$ be a d-dimensional sub-Gaussian random vector. Then*

$$\mathrm{Prob}\left\{\max_{\ell\le L} v^TW_\ell v \ge \left[1+\sqrt{2\ln(L/\epsilon)}\right]^2 \max_{\ell\le L}\mathrm{Tr}(W_\ell V)\right\} \le \epsilon.$$

Proof. Let $R^2 = \max\limits_{\ell\le L}\mathrm{Tr}(W_\ell V)$. W.l.o.g. we may assume that $v = V^{1/2}\zeta$ where $\zeta \sim \mathcal{SG}(0,I)$. Let us fix $\ell \le L$. Applying (5.55) with $Q = V^{1/2}W_\ell V^{1/2}$ and ϵ replaced with ϵ/L, when taking into account that $v^TW_\ell v = \zeta^TQ\zeta$ with

$$\lambda_{\max}(Q) \le \|Q\|_{\mathrm{Fro}} \le \mathrm{Tr}(Q) \le R^2,$$

we get

$$\mathrm{Prob}\left\{v^TW_\ell v \ge \left[1+\sqrt{2\ln(L/\epsilon)}\right]^2 R^2\right\} \le \frac{\epsilon}{L},$$

and the claim of the corollary follows. □

5.4.2 Proof of Proposition 5.2.1

Let H be a candidate contrast matrix.

1^o. Observe that

$$\|\widehat{w}^H(\omega) - Bx\| \le \|H^T\xi\| + \left\|H^T\sum\nolimits_{\alpha=1}^{q}\eta_\alpha A_\alpha x\right\| + \|[B - H^TA]x\|. \tag{5.56}$$

Clearly,

$$\|[B - H^TA]x\| \le \max_{\ell\le L}\left\{\max_{x\in\mathcal{X}} x^T[B - H^TA]^TR_\ell[B - H^TA]x\right\}^{1/2},$$

so that by Theorem 5.2.1,

$$\forall x \in \mathcal{X} \quad \|[B - H^TA]x\| \le \max_{\ell\le L}\mathfrak{r}_\ell(H) \tag{5.57}$$

where

$$\mathfrak{r}_\ell^2(H) = \min_{\upsilon}\left\{\phi_{\mathcal{T}}(\upsilon) : \upsilon \ge 0, \left[\begin{array}{c|c} I_\nu & R_\ell^{1/2}[B - H^TA] \\ \hline [B - H^TA]^TR_\ell^{1/2} & \sum_k \upsilon_k T_k\end{array}\right] \succeq 0\right\}.$$

Taking into account that $\sqrt{u} = \min_{\lambda\ge 0}\{\frac{u}{4\lambda} + \lambda\}$ for $u > 0$, we get

$$\mathfrak{r}_\ell(H) = \min_{\upsilon,\lambda}\left\{\lambda + \frac{\phi_{\mathcal{T}}(\upsilon)}{4\lambda} : \upsilon \ge 0, \lambda \ge 0, \left[\begin{array}{c|c} I_\nu & R_\ell^{1/2}[B - H^TA] \\ \hline [B - H^TA]^TR_\ell^{1/2} & \sum_k \upsilon_k T_k\end{array}\right] \succeq 0\right\}.$$

Setting $\mu = \upsilon/(4\lambda)$, by the homogeneity of $\phi_{\mathcal{T}}(\cdot)$ we obtain

$$\mathfrak{r}_\ell(H) = \min_{\mu,\lambda}\left\{\lambda + \phi_{\mathcal{T}}(\mu) : \mu \ge 0, \left[\begin{array}{c|c} \lambda I_\nu & \frac{1}{2}R_\ell^{1/2}[B - H^TA] \\ \hline \frac{1}{2}[B - H^TA]^TR_\ell^{1/2} & \sum_k \mu_k T_k\end{array}\right] \succeq 0\right\}. \tag{5.58}$$

2^o. Next, by Corollary 5.4.1 of the appendix,

$$\mathrm{Prob}\left\{\|H^T\xi\| \ge [1 + \sqrt{2\ln(2L/\epsilon)}]\sigma\max_{\ell\le L}\sqrt{\mathrm{Tr}(HR_\ell H^T)}\right\} \le \epsilon/2. \tag{5.59}$$

Similarly, because

$$\left\|H^T\sum\nolimits_{\alpha=1}^{q}\eta_\alpha A_\alpha x\right\| = \max_{\ell\le L}\left\|R_\ell^{1/2}H^T[A_1x, ..., A_qx]\eta\right\|_2,$$

we conclude that for any $x \in \mathcal{X}$

$$\mathrm{Prob}\left\{\left\|H^T\sum\nolimits_{\alpha=1}^{q}\eta_\alpha A_\alpha x\right\| \ge [1 + \sqrt{2\ln(2L/\epsilon)}]\max_{\ell\le L}s_\ell(H)\right\} \le \epsilon/2$$

where $s_\ell(H) = \left\{\max_{x\in\mathcal{X}} x^T\left[\sum_\alpha A_\alpha^THR_\ell H^TA_\alpha\right]x\right\}^{1/2}$. Again, by Theorem 5.2.1, $s_\ell(H)$ may be tightly upper-bounded by the quantity $\bar{\mathfrak{s}}_\ell(H)$ such that

$$\bar{\mathfrak{s}}_\ell^2(H) = \min_{\upsilon}\left\{\phi_{\mathcal{T}}(\upsilon) : \upsilon \ge 0, \left[\begin{array}{c|c} I_{\nu q} & [R_\ell^{1/2}H^TA_1; ...; R_\ell^{1/2}H^TA_q] \\ \hline [A_1^THR_\ell^{1/2}, ..., A_q^THR_\ell^{1/2}] & \sum_k \upsilon_k T_k\end{array}\right] \succeq 0\right\}.$$

Now, repeating the steps which led to (5.58) above, we conclude that

$$\bar{\mathfrak{s}}_\ell(H) = \min_{\mu',\lambda'}\left\{\lambda' + \phi_{\mathcal{T}}(\mu') : \mu' \ge 0, \left[\begin{array}{c|c} \lambda' I_{\nu q} & \frac{1}{2}[R_\ell^{1/2}H^TA_1; ...; R_\ell^{1/2}H^TA_q] \\ \hline \frac{1}{2}[A_1^THR_\ell^{1/2}, ..., A_q^THR_\ell^{1/2}] & \sum_k \mu'_k T_k\end{array}\right] \succeq 0\right\}. \tag{5.60}$$

3^o. When substituting the above bounds into (5.56), we conclude that for every feasible solution $\lambda_\ell, \mu^\ell, \kappa^\ell, \varkappa^\ell, \rho, \varrho$ to problem (5.12) associated with H, the ϵ-risk of the linear estimate $\widehat{w}^H_{\rm lin}(\cdot)$ may be upper-bounded by the quantity

$$[1+\sqrt{2\ln(2L/\epsilon)}]\left[\sigma\max_{\ell\le L}\|HR_\ell^{1/2}\|_{\rm Fro}+\rho\right]+\varrho. \qquad \square$$

5.5 Proofs for Section 5.2.2

5.5.1 Proof of Proposition 5.2.3

All we need to prove is that if $\lambda_\ell, \mu^\ell, \upsilon^\ell, \rho$ is a feasible solution to the optimization problem (5.30), then the inequality

$$\mathrm{Risk}_\epsilon[\widehat{w}^H_{\rm poly}|\mathcal{X}]\le 2\rho \tag{5.61}$$

holds. Indeed, let us fix $x\in\mathcal{X}$. Since the columns of H belong to $\mathcal{H}_\delta$, the P_x-probability of the event

$$\mathcal{Z}^c=\{\zeta:\|H^T\zeta\|_\infty>1\} \qquad [\zeta=\textstyle\sum_\alpha\eta_\alpha A_\alpha x+\xi]$$

is at most $ML\delta=\epsilon$. Let us fix observation $\omega=Ax+\zeta$ with ζ belonging to the complement $\mathcal{Z}$ of $\mathcal{Z}^c$. Then

$$\|H^T[\omega-Ax]\|_\infty=\|H^T\zeta\|_\infty\le 1,$$

implying that the optimal value in the optimization problem $\min_{u\in\mathcal{X}}\|H^T[Au-\omega]\|_\infty$ is at most 1. Consequently, setting $\widehat{x}=\widehat{x}^H(\omega)$, we have $\widehat{x}\in\mathcal{X}$ and $\|H^T[A\widehat{x}-\omega]\|_\infty\le 1$, see (5.22). We conclude that setting $z=\frac{1}{2}[x-\overline{x}]$, we have

$$\|H_\ell^TAz\|_\infty\le 1, \ell\le L$$

with $z\in\mathcal{X}$, implying that $z^TT_kz\le t_k$, $k\le K$, for some $t\in\mathcal{T}$. Now let $u\in\mathbf{R}^\nu$ with $\|u\|_2\le 1$. Semidefinite constraints in (5.30) imply that

$$\begin{aligned}u^TR_\ell^{1/2}Bz&\le u^T\lambda_\ell I_\nu u+z^T\left[A^TH_\ell\mathrm{Diag}\{\upsilon^\ell\}H_\ell^TA+\textstyle\sum_k\mu_k^\ell T_k\right]z\\&\le\lambda_\ell u^Tu+\textstyle\sum_j\upsilon_j^\ell[H^TAz]_j^2+\sum_k\mu_k^\ell t_k\\&\le\lambda_\ell+\textstyle\sum_j\upsilon_j^\ell+\phi_{\mathcal{T}}(\mu^\ell)\le\rho\end{aligned}$$

(recall that $\|u\|_2\le 1$, $\lambda_\ell\ge 0, \mu^\ell\ge 0, \upsilon^\ell\ge 0$, $t\in\mathcal{T}$, and $\|H_\ell^TAz\|_\infty\le 1$). We conclude that $u^TR_\ell^{1/2}Bz\le\rho$, $\ell\le L$, whenever $\|u\|_2\le 1$, i.e., $\|R_\ell^{1/2}[Bz]\|_2\le\rho^2$. The latter relation holds true for all $\ell\le L$, implying that $\|Bz\|\le\rho$, that is, $\|Bx-\widehat{x}(\omega)\|=2\|Bz\|\le 2\rho$ whenever $\zeta\in\mathcal{Z}$. $\square$

5.5.2 Proof of Proposition 5.2.4

0^o. We need the following technical result.

Theorem 5.5.1 [130, Theorem 4.6.1] *Let $Q_i\in\mathbf{S}^n$, $1\le i\le I$, and let ξ_i, $i=1,...,I$, be independent Rademacher (±1 with probabilities $1/2$) or $\mathcal{N}(0,1)$ random variables. Then for all $t\ge 0$ one has*

$$\mathrm{Prob}\left\{\left\|\textstyle\sum_{i=1}^I\xi_iQ_i\right\|\ge t\right\}\le 2n\exp\left\{-\dfrac{t^2}{2v_Q}\right\}$$

where $\|\cdot\|$ is the spectral norm, and $v_Q=\left\|\sum_{i=1}^IQ_i^2\right\|$.

1^o. Proof of (i). Let $\lambda_j \geq 0$, $g_j \in \mathcal{H}$, $j \leq M$, and $\Sigma = \sum_j \lambda_j g_j g_j^T$. Then for every j there exists $r^j \in \mathcal{R}$ such that $S_i^2[g_j] \preceq [r^j]_i I_{d_i}$, $i \leq I$. Assuming $\sum_j \lambda_j > 0$ and setting $\kappa_j = [\sum_j \lambda_j]^{-1}\lambda_j$ and $r = \sum_j \kappa_j r^j \in \mathcal{R}$, we have

$$\mathcal{S}_i\left[\sum_j \lambda_j g_j g_j^T\right] = \sum_j \lambda_j S_i^2[g_j] \preceq \sum_j \lambda_j [r^j]_i I_{d_i} = \left[\sum_j \lambda_j\right] r_i I_{d_i},$$

implying that $(\Sigma, \sum_j \lambda_j) \in \mathbf{K}$. The latter inclusion is true as well when $\lambda = 0$.

2^o. Proof of (ii). Let $(\Sigma, \rho) \in \mathbf{M}$, and let us prove that $\Sigma = \sum_{j=1}^N \lambda_j g_j g_j^T$ with $g_j \in \mathcal{H}$, $\lambda_j \geq 0$, and $\sum_j \lambda_j \leq \varkappa\rho$. There is nothing to prove when $\rho = 0$, since in this case $\Sigma = 0$ due to $(\Sigma, 0) \in \mathbf{K}$ combined with (5.35b). Now let $\rho > 0$, so that for some $r \in \mathcal{R}$ we have

$$\mathcal{S}_i[\Sigma] \preceq \rho r_i I_{d_i},\ i \leq I, \tag{5.62}$$

let $Z = \Sigma^{1/2}$, and let O be the orthonormal $N \times N$ matrix of N-point Discrete Cosine Transform, so that all entries in O are in magnitude $\leq \sqrt{2/N}$. For a Rademacher random vector $\varsigma = [\varsigma_1; ...; \varsigma_M]$ (i.e., with entries ς_i which are independent Rademacher random variables), let

$$Z^\varsigma = Z\mathrm{Diag}\{\varsigma\}O.$$

In this case, one has $Z^\varsigma [Z^\varsigma]^T \equiv \Sigma$, that is,

$$\sum_{p=1}^N \mathrm{Col}_p[Z^\varsigma]\mathrm{Col}_p^T[Z^\varsigma] \equiv \Sigma.$$

Recall that

$$\mathrm{Col}_j[Z^\varsigma] = \sum_p \varsigma_p O_{pj}\mathrm{Col}_p[Z],$$

and thus

$$S_i[\mathrm{Col}_j[Z^\varsigma]] = \sum_p \varsigma_p O_{pj} S_i[\mathrm{Col}_p[Z]].$$

Now observe that

$$\begin{aligned}\sum_p \big(O_{pj} S_i[\mathrm{Col}_p[Z]]\big)^2 &= \sum_p O_{pj}^2 S_i^2[\mathrm{Col}_p[Z]] = \sum_p O_{pj}^2 \mathcal{S}_i[\mathrm{Col}_p[Z]\mathrm{Col}_p^T[Z]] \\ \text{[see (5.35a)]} \quad &\preceq \frac{2}{N}\sum_p \mathcal{S}_i[\mathrm{Col}_p[Z]\mathrm{Col}_p^T[Z]] \\ &= \frac{2}{N}\mathcal{S}_i[\sum_p \mathrm{Col}_p[Z]\mathrm{Col}_p^T[Z]] = \frac{2}{N}\mathcal{S}_i[\Sigma] \preceq \frac{2}{N}\rho s_i I_{d_i}\end{aligned}$$

due to (5.62). By the Noncommutative Khintchine Inequality we have

$$\forall \gamma > 0 : \mathrm{Prob}\left\{S_i^2[\mathrm{Col}_j[Z^\varsigma]] \preceq \gamma\frac{2}{N}\rho s_i I_{d_i}\right\} \geq 1 - 2d_i\exp\{-\gamma/2\} \tag{5.63}$$

Setting

$$\gamma = 2\ln(4DN),\ D = \sum_i d_i, \quad g_j^\varsigma = \sqrt{\frac{N}{2\gamma\rho}}\mathrm{Col}_j[Z^\varsigma], \quad \lambda_j = \frac{2\gamma\rho}{N},\ 1 \leq j \leq N,$$

we conclude that event

$$\Xi = \left\{\varsigma : S_i^2[g_j^\varsigma] \preceq s_i I_{d_i}, i \leq I, j \leq N\right\} \subset \left\{g_j^\varsigma \in \mathcal{H},\ j \leq N\right\}$$

satisfies $\mathrm{Prob}(\Xi) \geq \frac{1}{2}$, while

$$\sum_j \lambda_j g_j^\varsigma [g_j^\varsigma]^T = \sum_j \mathrm{Col}_j[Z^\varsigma]\mathrm{Col}_j^T[Z^\varsigma] \equiv \Sigma \ \text{ and } \ \sum_j \lambda_j = \gamma\kappa\rho = 2\gamma\rho = \varkappa\rho.$$

Thus, with probability $\geq 1/2$ (whenever $\varsigma \in \Xi$), vectors $g_j = g_j^\varsigma$ and λ_j meet the requirements in (ii). □

Note that the proof of the proposition suggests an efficient randomized algorithm for generating the required g_j and λ_j: we generate realizations of ς of a Rademacher random vector, compute the corresponding vectors g_j^ς, and terminate when all of them happen to belong to $\mathcal{H}$. The corresponding probability not to terminate in course of the first k rounds of randomization is then $\leq 2^{-k}$.

5.5.3 Proof of Lemma 3

The proof of the lemma is given by the standard argument underlying median-of-means construction (cf. [56, Section 6.5.3.4]). For the sake of completeness, we reproduce it here.

1^o. Observe that when (5.14) holds, $h \in \mathcal{H}$, $x \in \mathcal{X}$ and $\zeta = \xi + \sum_\alpha \eta_\alpha A_\alpha x$, the probability of the event

$$\{|h^T\zeta| > 1\}$$

is at most $1/8$. Indeed, when $|h^T\zeta| > 1$ implies that either $|h^T\xi| > 1/2$ or $|\eta^T\mathcal{A}[h]x| > 1/2$. By the Chebyshev inequality, the probability of the first of these events is at most $4\mathbf{E}\{(h^T\xi)^2\} \leq 4\sigma^2\|h\|_2^2 \leq \frac{1}{16}$ (we have used the first relation in (5.14) and took into account that $h \in \mathcal{H}$). By similar argument, the probability of the second event is at most $4\mathbf{E}\{(\eta^T\mathcal{A}[h]x)^2\} \leq 4\|\mathcal{A}[h]x\|_2^2 \leq \frac{1}{16}$.

2^o. Let $\zeta_k = \omega_k - Ax$. By construction, $z_j = y_j - h_j^T Ax$ is the median of the i.i.d. sequence $h_j^T\zeta_k$, $k = 1, ..., K$. When $|z_j| > 1$, at least $K/2$ of the events $\{|h_j^T\zeta_k| > 1\}$, $k \leq K$, take place. Because the probability of each of K independent events is $\leq 1/8$, it is easily seen[6] that the probability that at least $K/2$ of them happen is bounded with

$$\pi(K) := \sum_{k \geq K/2} \binom{K}{k} (1/8)^k (7/8)^{K-k} \leq \sum_{k \geq K/2} \binom{K}{k} 2^{-K}[(1/4)^k(7/4)^{K-k}] \leq (\sqrt{7}/4)^K \leq e^{-0.4K}.$$

In other words, the probability of each event $E_j = \{\omega^K : |y_j - h_j^T Ax| > 1\}$, $j = 1, ..., M$, is bounded with $\pi(K)$. Thus, none of the events $E_1, ..., E_M$ takes place with probability at least $1 - M\pi(K)$, and in such case we have $\|y - H^T Ax\|_\infty \leq 1$, and so $\|y - H^T A\widehat{x}^H(\omega^K)\|_\infty \leq 1$ as well. We conclude that for every $x \in \mathcal{X}$, the probability of the event

$$\big\{x - \widehat{x}^H(\omega^K) \in 2\mathcal{X},\ \|H^T A[x - \widehat{x}^H(\omega^K)]\|_\infty \leq 2\big\}$$

is at least $1 - M\pi(K) \geq 1 - \epsilon$ when $K \geq 2.5\ln[M/\epsilon]$, and when it happens, one has $\|Bx - \widehat{w}^H_{\mathrm{poly}}(\omega^K)\| \leq \mathfrak{p}[H]$. □

[6] We refer to, e.g., [131, Section 2.3.2] for the precise justification of this obvious claim.

5.5.4 Proof of Proposition 5.2.2

1^o. Let $\ell \le L$ and $k \le K$ be fixed, let $H = H_\ell \in \mathbf{R}^{m\times\nu}$ be a candidate contrast matrix, and let $\lambda, \mu, \kappa, \varkappa$ be a feasible solution to (5.15). One has

$$\mathbf{E}_{\xi_k}\left\{\|R_\ell^{1/2}H^T\xi_k\|_2^2\right\} = \mathrm{Tr}\left(\mathbf{E}_{\xi_k}\left\{R_\ell^{1/2}H^T\xi_k\xi_k^T HR_\ell^{1/2}\right\}\right) \le \sigma^2\mathrm{Tr}(HR_\ell H^T) = \sigma^2\|HR_\ell^{1/2}\|_{\mathrm{Fro}}^2. \tag{5.64}$$

Next, for any $x \in \mathcal{X}$ fixed we have

$$\begin{aligned}\mathbf{E}_{\eta_k}\left\{\left\|R_\ell^{1/2}H^T[\textstyle\sum_\alpha[\eta_k]_\alpha A_\alpha]x\right\|_2^2\right\} &= \mathbf{E}_{\eta_k}\left\{\left\|R_\ell^{1/2}H^T[A_1x,...,A_qx]\eta_k\right\|_2^2\right\} = x^T\left[\textstyle\sum_\alpha A_\alpha^T HR_\ell H^T A_\alpha\right]x\\ &= \|[R_\ell^{1/2}H^TA_1;...;R_\ell^{1/2}H^TA_q]x\|_2^2 \le (\lambda+\phi_{\mathcal{T}}(\mu))^2\end{aligned} \tag{5.65}$$

where the concluding inequality follows from the constraints in (5.15) (cf. item 2^o of the proof of Proposition 5.2.1). Next, similarly to item 1^o of the proof of Proposition 5.2.1 we have

$$\|R_\ell^{1/2}(B - H^TA)x\|_2^2 \le (\kappa + \phi_{\mathcal{T}}(\varkappa))^2.$$

Put together, the latter bound along with (5.64) and (5.65) imply (5.17).

2^o. By the Chebyshev inequality,

$$\forall \ell, k \quad \mathrm{Prob}\left\{\|R_\ell^{1/2}(w_\ell(\omega_k) - Bx)\|_2 \ge 2\widetilde{\mathfrak{R}}_\ell[H_\ell]\right\} \le \tfrac{1}{4};$$

applying [128, Theorem 3.1] we conclude that

$$\forall \ell \quad \mathrm{Prob}\left\{\|R_\ell^{1/2}(z_\ell(\omega^K) - Bx)\|_2 \ge 2C_\alpha\widetilde{\mathfrak{R}}_\ell[H_\ell]\right\} \le e^{-K\psi(\alpha,\frac{1}{4})}$$

where

$$\psi(\alpha,\beta) = (1-\alpha)\ln\frac{1-\alpha}{1-\beta} + \alpha\ln\frac{\alpha}{\beta} \tag{5.66}$$

and $C_\alpha = \frac{1-\alpha}{\sqrt{1-2\alpha}}$. When choosing $\alpha = \frac{\sqrt{3}}{2+\sqrt{3}}$ which corresponds to $C_\alpha = 2$ we obtain $\psi(\alpha,\frac{1}{4}) = 0.1070...$ so that for $\ell \le L$

$$\mathrm{Prob}\left\{\|R_\ell^{1/2}(z_\ell(\omega^K) - Bx)\|_2 \ge 4\widetilde{\mathfrak{R}}_\ell[H_\ell]\right\} \le e^{-0.1070K}$$

what is (5.18).

3^o. Now, let $K \ge \ln(L/\epsilon)/0.1070$. In this case, for all $\ell \le L$

$$\mathrm{Prob}\left\{\|R_\ell^{1/2}(z_\ell(\omega^K) - Bx)\|_2 \ge 4\widetilde{\mathfrak{R}}_\ell[H_\ell]\right\} \le \epsilon/L,$$

so that with probability $\ge 1-\epsilon$ the set $\mathcal{W}(\omega^K)$ is not empty (it contains Bx), and for all $v \in \mathcal{W}(\omega^K)$ one has

$$\|R_\ell^{1/2}(v - Bx)\|_2 \le \|R_\ell^{1/2}(z_\ell(\omega^K) - v)\|_2 + \|R_\ell^{1/2}(z_\ell(\omega^K) - Bx)\|_2 \ge 8\widetilde{\mathfrak{R}}_\ell[H_\ell]. \qquad \square$$

5.6 Proofs for Section 5.3

5.6.1 Proof of Proposition 5.3.3

The proof follows that of Proposition 5.2.3. All we need to prove is that if H satisfies the premise of the proposition and $\lambda_\ell, \Upsilon^\ell, \upsilon^\ell, \rho$ is a feasible solution to (5.48), then the inequality

$$\mathrm{Risk}_\epsilon[\widehat{w}^H_{\mathrm{poly}}|\mathcal{X}] \le 2\rho \tag{5.67}$$

holds. Indeed, let us fix $x \in \mathcal{X}$ and $\eta \in \mathcal{U}$. Since the columns of H satisfy (5.47), the P_x-probability of the event

$$\mathcal{Z}_{x,\eta} = \{\xi : \|H^T[D[\eta]x + \xi]\|_\infty \le 1\}$$

is at least $1 - ML\delta = 1 - \epsilon$. Let us fix observation $\omega = Ax + D[\eta]x + \xi$ with $\xi \in \mathcal{Z}_{x,\eta}$. Then

$$\|H^T[\omega - Ax]\|_\infty = \|H^T[D[\eta]x + \xi]\|_\infty \le 1, \tag{5.68}$$

implying that the optimal value in the optimization problem $\min_{u\in\mathcal{X}} \|H^T[Au - \omega]\|_\infty$ is at most 1. Consequently, setting $\widehat{x} = \widehat{x}^H(\omega)$, we have $\widehat{x} \in \mathcal{X}$ and $\|H^T[A\widehat{x} - \omega]\|_\infty \le 1$, see (5.22). These observations combine with (5.68) and the inclusion $x \in \mathcal{X}$ to imply that for $z = \frac{1}{2}[x - \widehat{x}]$ we have $z \in \mathcal{X}$ and $\|H^T z\|_\infty \le 1$. Recalling what $\mathcal{X}$ is we conclude that $z = Py$ with $T^2_k[y] \preceq t_k I_{f_k}, k \le K$ for some $t \in \mathcal{T}$ and

$$\|H^T_\ell APy\|_\infty = \|H^T_\ell Az\|_\infty \le 1,\ \ell \le L. \tag{5.69}$$

Now let $u \in \mathbf{R}^\nu$ with $\|u\|_2 \le 1$. Semidefinite constraints in (5.48) imply that

$$\begin{aligned}
u^T R^{1/2}_\ell Bz = u^T R^{1/2}_\ell BPy &\le u^T \lambda_\ell I_\nu u + y^T \left[PA^T H_\ell \mathrm{Diag}\{\upsilon^\ell\} H^T_\ell AP + \sum\nolimits_k T^{+,*}_k[\Upsilon^\ell_k]\right] y \\
&= \lambda_\ell u^T u + \sum\nolimits_j \upsilon^\ell_j \underbrace{[H^T_\ell APy]^2_j}_{\le 1 \text{ by } (5.69)} + \sum\nolimits_k y^T T^{+,*}_k[\Upsilon^\ell_k] y \\
&\le \lambda_\ell + \sum\nolimits_j \upsilon^\ell_j + \sum_k \sum_{i,j\le N} y_i y_j \mathrm{Tr}(\Upsilon^\ell_k T^{ik} T^{jk}) \\
&= \lambda_\ell + \sum\nolimits_j \upsilon^\ell_j + \sum_k \mathrm{Tr}(\Upsilon^\ell_k T^2_k[y]) \\
&\le \lambda_\ell + \sum\nolimits_j \upsilon^\ell_j + \sum_k t_k \mathrm{Tr}(\Upsilon^\ell_k) \text{ [due to } \Upsilon^\ell \succeq 0 \text{ and } T^2_k[y] \preceq t_k I_{f_k}] \\
&\le \lambda_\ell + \sum\nolimits_j \upsilon^\ell_j + \phi_{\mathcal{T}}(\lambda[\Upsilon^\ell]) \le \rho
\end{aligned} \tag{5.70}$$

where the concluding inequality follows from the constraints of (5.48). (5.70) holds true for all u with $\|u\|_2 \le 1$, and we conclude that for $x \in \mathcal{X}$ and $\eta \in \mathcal{U}$ and $\xi \in \mathcal{Z}_{x,\eta}$ (recall that the latter inclusion takes place with P_x-probability $\ge 1 - \epsilon$) we have

$$\|R^{1/2}_\ell B[\widehat{x}^H(Ax + D[\eta]x + \xi) - x]\|_2 \le 2\rho,\ \ell \le L.$$

Recalling what $\|\cdot\|$ is, we get

$$\forall(x \in \mathcal{X}, \eta \in \mathcal{U}) : \mathrm{Prob}_{\xi\sim P_x}\{\|B[x - \widehat{x}^H(Ax + D[\eta]x + \xi)]\| > 2\rho\| \le \epsilon,$$

that is, $\mathrm{Risk}_\epsilon[\widehat{w}^H_{\mathrm{poly}}|\mathcal{X}] \le 2\rho$. The latter relation holds true whenever ρ can be extended to a feasible solution to (5.48), and (5.67) follows. □

5.6.2 Robust norm of uncertain matrix with structured norm-bounded uncertainty

Situation and goal

Let matrices $A_s \in \mathbf{R}^{m\times n}$, $s \le S$, and $L_t \in \mathbf{R}^{p_t\times m}$, $R_t \in \mathbf{R}^{q_t\times n}$, $t \le T$, be given. These data specify *uncertain $m \times n$ matrix*

$$\mathcal{A} = \{A = \sum\nolimits_s \delta_s A_s + \sum\nolimits_t L_t^T \Delta_t R_t : |\delta_s| \le 1\, \forall s \le S, \|\Delta_t\|_{2,2} \le 1\, \forall t \le T\}. \tag{5.71}$$

Given ellitopes

$$\begin{array}{rcl} \mathcal{X} & = & \{Py : y \in \mathcal{Y}\} \subset \mathbf{R}^n,\ \mathcal{Y} = \{y \in \mathbf{R}^N\ \&\ \exists t \in \mathcal{T} : y^T T_k y \le t_k, k \le K\}, \\ \mathcal{B}_* & = & \{Qz : z \in \mathcal{Z}\} \subset \mathbf{R}^m,\ \mathcal{Z} = \{z \in \mathbf{R}^M : \exists s \in \mathcal{S} : z^T S_\ell z \le s_\ell,\ \ell \le L\}, \end{array} \tag{5.72}$$

we want to upper-bound the robust norm

$$\|\mathcal{A}\|_{\mathcal{X},\mathcal{B}} = \max_{A\in\mathcal{A}} \|A\|_{\mathcal{X},\mathcal{B}},$$

of uncertain matrix $\mathcal{A}$ induced by the norm $\|\cdot\|_{\mathcal{X}}$ with the unit ball $\mathcal{X}$ in the argument space and the norm $\|\cdot\|_{\mathcal{B}}$ with the unit ball $\mathcal{B}$ which is the polar of $\mathcal{B}_*$ in the image space.

Main result

Proposition 5.6.1 *Given uncertain matrix (5.71) and ellitopes (5.72), consider convex optimization problem*

$$\begin{array}{rl} \mathrm{Opt} = \min\limits_{\substack{\mu,\upsilon,\lambda,\\ U_s,V_s,U^t,V^t}} & \frac{1}{2}[\phi_{\mathcal{S}}(\mu) + \phi_{\mathcal{T}}(\upsilon)] \\ \text{subject to} & \\ & \mu \ge 0,\ \upsilon \ge 0,\ \lambda \ge 0 \\ & \left[\begin{array}{c|c} U_s & -Q^T A_s P \\ \hline -P^T A_s^T Q & V_s \end{array}\right] \succeq 0 \qquad (5.73a) \\ & \left[\begin{array}{c|c} U^t & -Q^T L_t^T \\ \hline -L_t Q & \lambda_t I_{p_t} \end{array}\right] \succeq 0,\ V^t - \lambda_t P^T R_t^T R_t P \succeq 0 \qquad (5.73b) \\ & \sum\nolimits_\ell \mu_\ell S_\ell - \sum\nolimits_s U_s - \sum\nolimits_t U^t \succeq 0 \qquad (5.73c) \\ & \sum\nolimits_k \upsilon_k T_k - \sum\nolimits_s V_s - \sum\nolimits_t V^t \succeq 0 \qquad (5.73d) \end{array}$$

The problem is strictly feasible and solvable, and

$$\|\mathcal{A}\|_{\mathcal{X},\mathcal{B}} \le \mathrm{Opt} \le \varkappa(K)\varkappa(L)\max[\vartheta(2\kappa), \pi/2]\, \|\mathcal{A}\|_{\mathcal{X},\mathcal{B}} \tag{5.74}$$

where

- *the function $\vartheta(k)$ of nonnegative integer k is given by $\vartheta(0) = 0$ and*

$$\vartheta(k) = \left[\min_\alpha \left\{(2\pi)^{-k/2} \int |\alpha_1 u_1^2 + ... + \alpha_k u_k^2| \mathrm{e}^{-u^T u/2} du,\ \alpha \in \mathbf{R}^k, \|\alpha\|_1 = 1\right\}\right]^{-1},\ \ k \ge 1; \tag{5.75}$$

- $\kappa = \max\limits_{s\le S} \mathrm{Rank}(A_s)$ *when $S \ge 1$, otherwise $\kappa = 0$;*
- $\varkappa(\cdot)$ *is given by*

$$\varkappa(J) = \begin{cases} 1, & J = 1, \\ \frac{5}{2}\sqrt{\ln(2J)}, & J > 1. \end{cases} \tag{5.76}$$

Remarks. The rationale behind (5.73) is as follows. Checking that the $\mathcal{X},\mathcal{B}$-norm of uncertain $m\times n$ matrix (5.71) is $\le a\in\mathbf{R}$ is the same as to verify that for all $\delta_s\in[-1,1]$, $\Delta_t:\|\Delta_t\|_{2,2}\le 1$

$$\sum\nolimits_s \delta_s u^T A_s v + \sum\nolimits_t u^T L_t^T \Delta_t R_t v \le a\|u\|_{\mathcal{B}_*}\|v\|_{\mathcal{X}} \quad \forall(u\in\mathbf{R}^m, v\in\mathbf{R}^n),$$

or, which is the same due to what $\mathcal{B}_*$ and $\mathcal{X}$ are, that for all $\delta_s\in[-1,1], \Delta_t:\|\Delta_t\|_{2,2}\le 1$

$$\sum\nolimits_s \delta_s z^T Q^T A_s P y + \sum\nolimits_t z^T Q^T L_t^T \Delta_t R_t P y \le a\|z\|_{\mathcal{Z}}\|y\|_{\mathcal{Y}} \quad \forall(z\in\mathbf{R}^M, y\in\mathbf{R}^N). \tag{5.77}$$

A simple certificate for (5.77) is a collection of positive semidefinite matrices U_s, V_s, U^t, V^t, U, V such that for all $z\in\mathbf{R}^M$, $y\in\mathbf{R}^N$ and all $s\le S$, $t\le T$ it holds

$$2z^T[Q^T A_s P]y \le z^T U_s z + y^T V_s y, \tag{5.78a}$$
$$2z^T Q^T L_t^T \Delta_t R_t P y \le z^T U^t z + y^T V^t y \ \ \forall(\Delta_t:\|\Delta_t\|_{2,2}\le 1), \tag{5.78b}$$
$$\sum\nolimits_s U_s + \sum\nolimits_t U^t \preceq U, \tag{5.78c}$$
$$\sum\nolimits_s V_s + \sum\nolimits_t V^t \preceq V, \tag{5.78d}$$
$$\max_{z\in\mathcal{Z}} z^T U z + \max_{y\in\mathcal{Y}} y^T V y \le 2a. \tag{5.78e}$$

Now, (5.78a) clearly is the same as (5.73a). It is known (this fact originates from [132]) that (5.78b) is the same as existence of $\lambda_t\ge 0$ such that (5.73b) holds. Finally, existence of $\mu\ge 0$ such that $\sum_\ell \mu_\ell S_\ell \succeq U$ and $\upsilon\ge 0$ such that $\sum_k \upsilon_k T_k \succeq V$ (see (5.73c) and (5.73d)) implies due to the structure of $\mathcal{Z}$ and $\mathcal{Y}$ that $\max_{z\in\mathcal{Z}} z^T U z \le \phi_{\mathcal{S}}(\mu)$ and $\max_{y\in\mathcal{Y}} y^T V y \le \phi_{\mathcal{T}}(\upsilon)$. The bottom line is that a feasible solution to (5.73) implies the existence of a certificate

$$\left\{U_s, U^t, V_s, V^t, s\le S, t\le T, U=\sum\nolimits_\ell \mu_\ell S_\ell, V=\sum\nolimits_k \upsilon_k T_k\right\}$$

for relation (5.77) with $a=\frac{1}{2}[\phi_{\mathcal{S}}(\mu)+\phi_{\mathcal{T}}(\upsilon)]$.

Proof of Proposition 5.6.1. 1^o. Strict feasibility and solvability of the problem are immediate consequences of $\sum_\ell S_\ell \succ 0$ and $\sum_k T_k \succ 0$.

Let us prove the first inequality in (5.74). All we need to show is that if

[a] $\mu, \upsilon, \lambda, U_s, V_s, U^t, V^t$ is feasible for (5.73),

[b] $x=Py$ with $y^T T_k y\le \tau_k$, $k\le K$, for some $\tau\in\mathcal{T}$ and $u=Qz$ for some z such that $z^T S_\ell z\le \varsigma_\ell$, $\ell\le L$, for some $\varsigma\in\mathcal{S}$, and

[c] δ_s, Δ_t satisfy $|\delta_s|\le 1$, $\|\Delta_t\|_{2,2}\le 1$,

then $\gamma := u^T[\sum_s \delta_s A_s + \sum_t L_t^T \Delta_t R_t]x \le \frac{1}{2}[\phi_{\mathcal{S}}(\mu)+\phi_{\mathcal{T}}(\upsilon)]$. Assuming [a–c], we have

$$\begin{aligned}
\gamma &= \sum\nolimits_s \delta_s z^T Q^T A_s P y + \sum\nolimits_t z^T Q^T L_t^T \underbrace{\Delta_t R_t P y}_{\zeta_t} \\
&\le \tfrac{1}{2} z^T \left[\sum\nolimits_s U_s\right] z + \tfrac{1}{2} y^T \left[\sum\nolimits_s V_s\right] y + \sum\nolimits_t \|L_t Q z\|_2 \|\zeta_t\|_2 \quad \text{[by (5.73a) and due to } |\delta_s|\le 1] \\
&\le \tfrac{1}{2} z^T \left[\sum_s U_s\right] z + \tfrac{1}{2} y^T \left[\sum_s V_s\right] y + \sum_t \sqrt{(\lambda_t z^T U^t z)(y^T P^T R_t^T R_t P y)} \\
&\quad \text{[due to (5.73b) and } \|\Delta_t\|_{2,2}\le 1] \\
&= \tfrac{1}{2} z^T \left[\sum\nolimits_s U_s\right] z + \tfrac{1}{2} y^T \left[\sum\nolimits_s V_s\right] y + \sum\nolimits_t \sqrt{(z^T U^t z)(\lambda_t y^T P^T R_t^T R_t P y)}.
\end{aligned}$$

Thus, by the second inequality of (5.73b),

$$\begin{aligned}
\gamma &\le \tfrac{1}{2} z^T \left[\sum\nolimits_s U_s\right] z + \tfrac{1}{2} y^T \left[\sum\nolimits_s V_s\right] y + \sum\nolimits_t \sqrt{(z^T U^t z)(y^T V^t y)} \\
&\le \tfrac{1}{2} z^T \left[\sum\nolimits_s U_s\right] z + \tfrac{1}{2} y^T \left[\sum\nolimits_s V_s\right] y + \tfrac{1}{2} \sum\nolimits_t [z^T U^t z + y^T V^t y] \\
&= \tfrac{1}{2} \left[z^T \left[\sum\nolimits_s U_s + \sum\nolimits_t U^t\right] z + y^T \left[\sum\nolimits_s V_s + \sum\nolimits_t V^t\right] y \right] \\
&\le \tfrac{1}{2} \left[\sum\nolimits_\ell \mu_\ell z^T S_\ell z + \sum\nolimits_k y^T \upsilon_k T_k y \right] \quad \text{[by (5.73c) and (5.73d)]} \\
&\le \tfrac{1}{2} \left[\sum\nolimits_\ell \mu_\ell \varsigma_\ell + \sum\nolimits_k \upsilon_k \tau_k \right] \quad \text{[due to } z^T S_\ell z \le \varsigma_\ell, y^T T_k y \le \tau_k\text{]} \\
&\le \tfrac{1}{2} [\phi_{\mathcal{S}}(\mu) + \phi_{\mathcal{T}}(\upsilon)] \quad \text{[since } \varsigma \in \mathcal{S}, \tau \in \mathcal{T}\text{]}
\end{aligned} \tag{5.79}$$

as claimed.

2^o. Now, let us prove the second inequality in (5.74). Observe that

$$\mathbf{S} = \{0\} \cup \{[s;\sigma] : \sigma > 0, s/\sigma \in \mathcal{S}\}, \quad \mathbf{T} = \{0\} \cup \{[t;\tau] : \tau > 0, t/\tau \in \mathcal{T}\},$$

are regular cones with the duals

$$\mathbf{S}_* = \{[g;\sigma] : \sigma \ge \phi_{\mathcal{S}}(-g)\}, \quad \mathbf{T}_* = \{[h;\tau] : \tau \ge \phi_{\mathcal{T}}(-h)\},$$

and (5.73) can be rewritten as the conic problem

$$\begin{aligned}
2\mathrm{Opt} = \min_{\substack{\alpha,\beta,\mu,\upsilon,\\ \lambda,U_s,V_s,U^t,V^t}} \quad & \alpha + \beta \qquad \text{(P)} \\
\text{subject to} \quad & \\
& [-\mu;\alpha]^{[\overline{g},\overline{\alpha}]} \in \mathbf{S}_*,\ [-\upsilon;\beta]^{[\overline{h},\overline{\beta}]} \in \mathbf{T}_*,\ \mu^{\overline{\mu}} \ge 0,\ \upsilon^{\overline{\upsilon}} \ge 0,\ \lambda^{\overline{\lambda}} \ge 0 \\
& \left[\begin{array}{c|c} U_s & -Q^T A_s P \\ \hline -P^T A_s^T Q & V_s \end{array}\right] \left[\begin{array}{c|c} \overline{U}_s & \overline{A}_s \\ \hline \overline{A}_s^T & \overline{V}_s \end{array}\right] \succeq 0,\ s \le S \\
& \left[\begin{array}{c|c} U^t & -Q^T L_t^T \\ \hline -L_t Q & \lambda_t I_{p_t} \end{array}\right] \left[\begin{array}{c|c} \overline{U}^t & \overline{L}_t^T \\ \hline \overline{L}_t & \overline{\Lambda}_t \end{array}\right] \succeq 0,\ [V^t - \lambda_t P^T R_t^T R_t P]^{\overline{V}^t} \succeq 0,\ t \le T \\
& [\textstyle\sum_\ell \mu_\ell S_\ell - \sum_s U_s - \sum_t U^t]^{\overline{S}} \succeq 0,\ [\sum_k \upsilon_k T_k - \sum_s V_s - \sum_t V^t]^{\overline{T}} \succeq 0
\end{aligned} \tag{5.80}$$

(superscripts are the Lagrange multipliers for the corresponding constraints). (P) clearly is solvable and strictly feasible, so that 2Opt is the optimal value of the (solvable!) conic dual of (P):

$$2\mathrm{Opt} = \max_{\substack{\overline{\alpha},\overline{\beta},\overline{g},\overline{h},\overline{\mu},\overline{\upsilon},\overline{\lambda},\overline{S},\overline{T},\\ \overline{U}_s,\overline{V}_s,\overline{A}_s,\overline{U}^t,\overline{L}_t,\overline{\Lambda}_t,\overline{V}^t}} 2\sum\nolimits_s \mathrm{Tr}(Q^T A_s P \overline{A}_s^T) + 2\sum\nolimits_t \mathrm{Tr}(Q^T L_t^T \overline{L}_t) \tag{D}$$

subject to

$$[\overline{g};\overline{\alpha}] \in \mathbf{T},\ [\overline{h};\overline{\beta}] \in \mathbf{S},\ \overline{\mu} \geq 0, \overline{\upsilon} \geq 0,\ \overline{\lambda} \geq 0,\ \overline{V}^t \succeq 0, \overline{S} \succeq 0, \overline{T} \succeq 0$$

$$\left[\begin{array}{c|c} \overline{U}_s & \overline{A}_s \\ \hline \overline{A}_s^T & \overline{V}_s \end{array}\right] \succeq 0,\ \left[\begin{array}{c|c} \overline{U}^t & \overline{L}_t^T \\ \hline \overline{L}_t & \overline{\Lambda}_t \end{array}\right] \succeq 0$$

$$\overline{\alpha} = 1,\ [\overline{g};\overline{\alpha}] \in \mathbf{S},\ \overline{\beta} = 1,\ [\overline{h};\overline{\beta}] \in \mathbf{T},\ -\overline{g}_\ell + \mathrm{Tr}(\overline{S} S_\ell) + \overline{\mu}_\ell = 0,\ -\overline{h}_k + \mathrm{Tr}(\overline{T} T_k) + \overline{\upsilon}_k = 0$$

$$\mathrm{Tr}(\overline{\Lambda}_t) - \mathrm{Tr}(\overline{V}_t P^T R^T R_t P) + \overline{\lambda}_t = 0$$

$$\overline{U}_s = \overline{S},\ \overline{U}_t = \overline{S},\ \overline{V}_s = \overline{T},\ \overline{V}^t = \overline{T}$$

(here and in what follows the constraints should be satisfied for all values of "free indexes" $s \leq S$, $t \leq T$, $\ell \leq L$, $k \leq K$). Taking into account that relation $\left[\begin{array}{c|c} X & Y \\ \hline Y^T & Z \end{array}\right] \succeq 0$ is equivalent to $X \succeq 0, Z \succeq 0$, and $Y = X^{1/2}\Delta Z^{1/2}$ with $\|\Delta\|_{2,2} \leq 1$, and that $[\overline{g};1] \in \mathbf{S}$, $[\overline{h};1] \in \mathbf{T}$ is the same as $\overline{g} \in \mathcal{S}$, $\overline{h} \in \mathcal{T}$, (D) boils down to

$$\mathrm{Opt} = \max_{\substack{\overline{g},\overline{h},\overline{S},\overline{T},\\ \overline{\Delta}_s,\overline{\delta}_t,\overline{\Lambda}_t}} \left\{ \sum\nolimits_s \mathrm{Tr}(Q^T A_s P \overline{A}_s^T) + \sum\nolimits_t \mathrm{Tr}(Q^T L_t^T \overline{L}_t) : \right.$$
$$\left. \begin{array}{l} \overline{g} \in \mathcal{T},\ \overline{h} \in \mathcal{S},\ \overline{S} \succeq 0,\ \overline{T} \succeq 0,\ \mathrm{Tr}(\overline{S} S_\ell) \leq \overline{g}_\ell,\ \mathrm{Tr}(\overline{T} T_k) \leq \overline{h}_k \\ \overline{A}_s = \overline{S}^{1/2} \overline{\Delta}_s \overline{T}^{1/2},\ \|\overline{\Delta}_s\|_{2,2} \leq 1, \overline{L}_t^T = \overline{S}^{1/2} \overline{\delta}_t \overline{\Lambda}_t^{1/2},\ \|\overline{\delta}_t\|_{2,2} \leq 1 \\ \mathrm{Tr}(\overline{\Lambda}_t) \leq \mathrm{Tr}(\overline{S}^{1/2} P^T R^T R_t P \overline{S}^{1/2}) \end{array} \right\}$$

or, which is the same,

$$\mathrm{Opt} = \max_{\substack{\overline{g},\overline{h},\overline{S},\overline{T}\\ \overline{\Delta}_s,\overline{\delta}_t,\overline{\Lambda}_t,\overline{L}_t}} \left\{ \sum\nolimits_s \mathrm{Tr}(\overline{S}^{1/2} Q^T A_s P \overline{T}^{1/2} \overline{\Delta}_s^T) + 2\sum\nolimits_t \mathrm{Tr}(\overline{S}^{1/2} Q^T L_t^T \overline{\Lambda}_t^{1/2} \overline{\delta}_t^T) : \right. \tag{D$'$}$$
$$\left. \begin{array}{l} \overline{g} \in \mathcal{T},\ \overline{h} \in \mathcal{S},\ \overline{S} \succeq 0,\ \overline{T} \succeq 0,\ \mathrm{Tr}(\overline{S} S_\ell) \leq \overline{g}_\ell,\ \mathrm{Tr}(\overline{T} T_k) \leq \overline{h}_k \\ \|\overline{\Delta}_s\|_{2,2} \leq 1,\ \|\overline{\delta}_t\|_{2,2} \leq 1 \\ \mathrm{Tr}(\overline{\Lambda}_t) \leq \mathrm{Tr}(\overline{T}^{1/2} P^T R_t^T R_t P \overline{T}^{1/2}),\ \overline{\Lambda}_t \succeq 0 \end{array} \right\}$$

Note that for Δ and δ such that $\|\Delta\|_{2,2} \leq 1$ and $\|\delta\|_{2,2} \leq 1$ one has

$$\mathrm{Tr}(A\Delta) \leq \|A\|_{\mathrm{nuc}} = \|\lambda(\mathcal{L}[A])\|_1,\ \mathcal{L}[A] = \left[\begin{array}{c|c} & \frac{1}{2}A \\ \hline \frac{1}{2}A^T & \end{array}\right]$$

and

$$Tr(AB^T\delta) = \langle A, \delta^T B\rangle_{\mathrm{Fro}} \leq \|A\|_{\mathrm{Fro}} \|\delta^T B\|_{\mathrm{Fro}} \leq \|A\|_{\mathrm{Fro}} \|B\|_{\mathrm{Fro}}$$

(here $\|A\|_{\mathrm{nuc}}$ stands for the nuclear norm and $\lambda(A)$ for the vector of eigenvalues of a symmetric matrix A). Consequently, for a feasible solution to (D$'$) it holds

$$\mathrm{Tr}(\overline{S}^{1/2} Q^T A_s P \overline{T}^{1/2} \overline{\Delta}_s^T) \leq \|\lambda(\mathcal{L}[\overline{S}^{1/2} Q^T A_s P \overline{T}^{1/2}])\|_1,$$

and

$$\mathrm{Tr}(\overline{S}^{1/2}Q^TL_t^T\overline{\Lambda}_t^{1/2}\overline{\delta}_t^T)\le\|\overline{S}^{1/2}Q^TL_t^T\|_{\mathrm{Fro}}\|\overline{\Lambda}_t^{1/2}\|_{\mathrm{Fro}}.$$

The latter bound combines with the last constraint in (D′) to imply that

$$\mathrm{Tr}(\overline{S}^{1/2}Q^TL_t^T\overline{\Lambda}_t^{1/2}\overline{\delta}_t^T)\le\|\overline{S}^{1/2}Q^TL_t^T\|_{\mathrm{Fro}}\|\overline{T}^{1/2}P^TR_t^T\|_{\mathrm{Fro}},$$

and we conclude that

$$\mathrm{Opt}\le\max_{\overline{S},\overline{g},\overline{T},\overline{h}}\left\{\sum\nolimits_s\left\|\lambda(\mathcal{L}[\overline{S}^{1/2}Q^TA_sP\overline{T}^{1/2}])\right\|_1+\sum\nolimits_t\left\|\overline{S}^{1/2}Q^TL_t^T\|_{\mathrm{Fro}}\|\overline{T}^{1/2}P^TR_t^T\right\|_{\mathrm{Fro}}:\right.\quad(5.81)$$
$$\left.\begin{array}{c}\overline{S}\succeq0,\overline{g}\in\mathcal{S},\ \mathrm{Tr}(\overline{S}S_\ell)\le\overline{g}_\ell,\ \ell\le L\\ \overline{T}\succeq0,\overline{h}\in\mathcal{T},\ \mathrm{Tr}(\overline{T}T_k)\le\overline{h}_k,\ k\le K\end{array}\right\}$$

4^o. We need the following result:

Lemma 4 *[133, Lemma 2.3] (cf. also [13, Lemma 3.4.3]) If the ranks of all matrices A_s (and thus—matrices $\overline{S}^{1/2}Q^TA_sP\overline{T}^{1/2}$) do not exceed a given $\kappa\ge1$, then for $\omega\sim\mathcal{N}(0,I_{M+N})$ one has*

$$\mathbf{E}\left\{|\omega^T\mathcal{L}[\overline{S}^{1/2}Q^TA_sP\overline{T}^{1/2}]\omega|\right\}\ge\|\lambda(\mathcal{L}[\overline{S}^{1/2}Q^TA_sP\overline{T}^{1/2}])\|_1/\vartheta(2\kappa),$$

with $\vartheta(\cdot)$ as described in Proposition 5.6.1.

Our next result is as follows (cf. [124, Proposition B.4.12])

Lemma 5 *Let $\in\mathbf{R}^{p\times q}$, $B\in\mathbf{R}^{r\times q}$ and $\xi\sim\mathcal{N}(Q,I_q)$. Then*

$$\mathbf{E}_\xi\left\{\|A\xi\|_2\|B\xi\|_2\right\}\ge\frac{2}{\pi}\|A\|_{\mathrm{Fro}}\|\|B\|_{\mathrm{Fro}}.$$

Proof. Setting $A^TA=U\mathrm{Diag}\{\lambda\}U^T$ with orthogonal U and $\zeta=U^T\xi$, we have

$$\mathbf{E}\left\{\|A\xi\|_2\|B\xi\|_2\right\}=\mathbf{E}\left\{\sqrt{\textstyle\sum_{i=1}^q\lambda_i[U^T\xi]_i^2}\|B\xi\|_2\right\}.$$

The right hand side is concave in λ, so that the infimum of this function in λ varying in the simplex $\sum_i\lambda_i=\mathrm{Tr}(A^TA)$ is attained at an extreme point. In other words, there exists vector $a\in\mathbf{R}^q$ with $a^Ta=\|A\|_{\mathrm{Fro}}^2$ such that

$$\mathbf{E}\left\{\|A\xi\|_2\|B\xi\|_2\right\}\ge\mathbf{E}_\xi\left\{|a^T\xi|\,\|B\xi\|_2\right\}.$$

Applying the same argument to $\|B\xi\|_2$-factor, we can now find a vector $b\in\mathbf{R}^q$, $b^Tb=\|B\|_{\mathrm{Fro}}^2$, such that

$$\mathbf{E}_\xi\left\{|a^T\xi|\,\|B\xi\|_2\right\}\ge\mathbf{E}_\xi\left\{|a^T\xi|\,|b^T\xi|\right\}.$$

It suffices to prove that the concluding quantity is $\ge2\|a\|_2\|b\|_2/\pi$. By homogeneity, this is the same as to prove that if $[s;t]\sim\mathcal{N}(0,I_2)$, then $\mathbf{E}\{|t|\,|\cos(\phi)t+\sin(\phi)s|\}\ge\frac{2}{\pi}$ for all $\phi\in[0,2\pi)$, which is straightforward (for the justification, see the proof of Proposition 2.3 of [134]). □

The last building block is the following

Lemma 6 *[118, Lemma 6] Let*

$$\mathcal{V} = \{v \in \mathbf{R}^d : \exists r \in \mathcal{R} : v^T R_j v \le r_j, 1 \le j \le J\} \subset \mathbf{R}^d$$

be a basic ellitope, $W \succeq 0$ be symmetric $d \times d$ matrix such that

$$\exists r \in \mathcal{R} : \mathrm{Tr}(W R_j) \le r_j, j \le J,$$

and $\omega \sim \mathcal{N}(0, W)$. Denoting by $\rho(\cdot)$ the norm on $\mathbf{R}^d$ with the unit ball $\mathcal{V}$, we have

$$\mathbf{E}\{\rho(\omega)\} \le \varkappa(J).$$

with $\varkappa(\cdot)$ given by (5.76).

4° Now we can complete the proof of the second inequality in (5.74). Let $\kappa \ge 1$, and let $\overline{g}, \overline{S}, \overline{h}, \overline{T}$ be feasible for the optimization problem in (5.81). Denoting by $\|\cdot\|_{\mathcal{Q}}$ the norm with the unit ball $\mathcal{Q}$, for all $A \in \mathbf{R}^{m\times n}$, $u \in \mathbf{R}^m$, and $v \in \mathbf{R}^n$ we have

$$u^T A v \le \|u\|_{\mathcal{B}_*}\|Av\|_{\mathcal{B}} \le \|u\|_{\mathcal{B}_*}\|A\|_{\mathcal{X},\mathcal{B}}\|v\|_{\mathcal{X}},$$

so that for all $u \in \mathbf{R}^m$ and $v \in \mathbf{R}^n$

$$\begin{aligned}\|u\|_{\mathcal{B}_*}\|v\|_{\mathcal{X}}\|\mathcal{A}\|_{\mathcal{X},\mathcal{B}} &\ge \max_{\substack{\epsilon_s,|\epsilon_s|\le 1,\\ \delta_t,\|\delta_t\|_{2,2}\le 1}} \left[\sum\nolimits_s \epsilon_s u^T A_s v + \sum\nolimits_t u^T L_t^T \delta_t R_t v\right] \\ &= \sum\nolimits_s |u^T A_s v| + \sum\nolimits_t \|L_t u\|_2 \|R_t v\|_2.\end{aligned}$$

Thus, for all $\overline{g}, \overline{S}, \overline{h}, \overline{T}$ which are feasible for (5.81) and $\xi \in \mathbf{R}^M$, $\eta \in \mathbf{R}^N$,

$$\begin{aligned}\|\overline{S}^{1/2}\xi\|_{\mathcal{Z}}\|\overline{T}^{1/2}\eta\|_{\mathcal{Y}}\|\mathcal{A}\|_{\mathcal{X},\mathcal{B}} &\ge \|Q\overline{S}^{1/2}\xi\|_{\mathcal{B}_*}\|P\overline{T}^{1/2}\eta\|_{\mathcal{X}}\|\mathcal{A}\|_{\mathcal{X},\mathcal{B}} \text{ [due to } \mathcal{B}_* = Q\mathcal{Z}, \mathcal{X} = P\mathcal{Y}] \\ &\ge \sum\nolimits_s |\xi^T \overline{S}^{1/2} Q^T A_s P \overline{T}^{1/2}\eta| + \sum\nolimits_t \|L_t Q \overline{S}^{1/2}\xi\|_2 \|R_t P \overline{T}^{1/2}\eta\|_2 \\ &= \sum\nolimits_s |[\xi;\eta]^T \mathcal{L}[\overline{S}^{1/2} Q^T A_s P \overline{T}^{1/2}][\xi;\eta]| \\ &\quad + \sum\nolimits_t \|[L_t Q \overline{S}^{1/2}, 0_{p_t\times N}][\xi;\eta]\|_2 \|[0_{q_t\times M}, R_t P \overline{T}^{1/2}][\xi;\eta]\|_2. \qquad (5.82)\end{aligned}$$

As a result, for $[\xi;\eta] \sim \mathcal{N}(0, I_{M+N})$, applying the bounds of Lemmas 4 and 5,

$$\begin{aligned}&\mathbf{E}\left\{\left\|\overline{S}^{1/2}\xi\right\|_{\mathcal{Z}}\right\}\mathbf{E}\left\{\left\|\overline{T}^{1/2}\eta\right\|_{\mathcal{Y}}\right\}\|\mathcal{A}\|_{\mathcal{X},\mathcal{B}} = \mathbf{E}\left\{\left\|\overline{S}^{1/2}\xi\right\|_{\mathcal{Z}}\|\overline{T}^{1/2}\eta\|_{\mathcal{Y}}\|\mathcal{A}\|_{\mathcal{X},\mathcal{B}}\right\} \\ &\ge \sum\nolimits_s \mathbf{E}\left\{\left|[\xi;\eta]^T \mathcal{L}[\overline{S}^{1/2} Q^T A_s P \overline{T}^{1/2}][\xi;\eta]\right|\right\} \\ &\quad + \sum\nolimits_t \mathbf{E}\left\{\left\|[L_t Q \overline{S}^{1/2}, 0_{p_t\times N}][\xi;\eta]\right\|_2 \left\|[0_{q_t\times M}, R_t P \overline{T}^{1/2}][\xi;\eta]\right\|_2\right\} \\ &\ge \vartheta(2\kappa)^{-1}\sum_s \left\|\lambda\left(\mathcal{L}[\overline{S}^{1/2} Q^T A_s P \overline{T}^{1/2}]\right)\right\|_1 + \tfrac{2}{\pi}\sum_t \|L_t Q \overline{S}^{1/2}\|_{\mathrm{Fro}}\|R_t P \overline{T}^{1/2}\|_{\mathrm{Fro}}.\end{aligned}$$

Besides this, by Lemma 6 we have

$$\mathbf{E}\left\{\|\overline{S}^{1/2}\xi\|_{\mathcal{Z}}\right\} \le \varkappa(L), \quad \mathbf{E}\left\{\|\overline{T}^{1/2}\eta\|_{\mathcal{Y}}\right\} \le \varkappa(K)$$

due to the fact that $\overline{g}, \overline{S}, \overline{h}$ and $\overline{T}$ are feasible for (5.81). This combines with (5.82) to imply that the value $\varkappa(L)\varkappa(K)\|\mathcal{A}\|_{\mathcal{X},\mathcal{B}}$ is lower bounded with the quantity

$$\max\left[\vartheta(2\kappa),\pi/2\right]^{-1}\left[\sum\nolimits_s \left\|\lambda\left(\mathcal{L}[\overline{S}^{1/2}Q^TS_sP\overline{T}^{1/2}]\right)\right\|_1+\sum\nolimits_t\|\overline{S}^{1/2}Q^TL_t^T\|_{\rm Fro}\|\overline{T}^{1/2}P^TR_t^T\|_{\rm Fro}\right].$$

Invoking the inequality in (5.81), we arrive at the second inequality in (5.74). The above reasoning assumed that $\kappa\geq 1$, with evident simplifications, it is applicable to the case of $\kappa=0$ as well. □

5.6.3 Proof of Proposition 5.3.1

We put $S=q_{\rm s}$ and $T=q-q_{\rm s}$. In the situation of Proposition 5.3.1 we want to tightly upper-bound quantity

$$\begin{array}{rcl}\mathfrak{s}(H) & = & \max\limits_{x\in\mathcal{X},\eta\in\mathcal{U}}\left\|H^TD[\eta]x\right\|\\ & = & \max\limits_{\ell\leq L}\max\limits_{x\in\mathcal{X},\eta\in\mathcal{U}}\left\{\sqrt{[H^TD[\eta]x]^TR_\ell[H^TD[\eta]x]}\right\}\\ & = & \max\limits_{\ell\leq L}\|\mathcal{A}_\ell[H]\|_{\mathcal{X},2},\end{array}$$

where $\|\cdot\|_{\mathcal{X},2}$ is the operator norm induced by $\|\cdot\|_{\mathcal{X}}$ on the argument and $\|\cdot\|_2$ on the image space and the uncertain matrix $\mathcal{A}_\ell[H]$ is given by

$$\mathcal{A}_\ell=\left\{\sum\nolimits_{s=1}^S\delta_s\underbrace{R_\ell^{1/2}H^TP_s^TQ_s}_{=:A_{s\ell}[H]}+\sum\nolimits_{t=1}^T\underbrace{R_\ell^{1/2}H^TP_{S+t}^T}_{L_{t\ell}^T[H]}\Delta_s\underbrace{Q_{S+t}}_{=:R_t}:\begin{array}{l}|\delta_s|\leq 1\quad,1\leq s\leq S\\ \|\Delta_s\|_{2,2}\leq 1\quad,1\leq t\leq T\end{array}\right\}$$

It follows that

$$\mathfrak{s}(H)=\max_{\ell\leq L}\|\mathcal{A}_\ell[H]\|_{\mathcal{X},2},$$

and Proposition 5.6.1 provides us with the efficiently computable convex in H upper bound $\overline{\mathfrak{s}}(H)$ on $\mathfrak{s}(H)$:

$$\begin{array}{rcl}\overline{\mathfrak{s}}(H) & = & \max\limits_{\ell\leq L}{\rm Opt}_\ell(H),\\ {\rm Opt}_\ell(H) & = & \min\limits_{\mu,\upsilon,\lambda,U_s,V_s,U^t,V^t}\left\{\frac{1}{2}[\mu+\phi_{\mathcal{T}}(\upsilon)]:\ \mu\geq 0,\ \upsilon\geq 0,\ \lambda\geq 0\right.\\ & & \left.\begin{array}{l}\left[\begin{array}{c|c}U_s & -A_{s\ell}[H]P\\ \hline -P^TA_{s\ell}^T[H] & V_s\end{array}\right]\succeq 0\\ \left[\begin{array}{c|c}U^t & -L_{t\ell}^T[H]\\ \hline -L_{t\ell}[H] & \lambda_tI_{p_{q_{\rm S}+t}}\end{array}\right]\succeq 0\\ V^t-\lambda_tP^TR_t^TR_tP\succeq 0\\ \mu I_\nu-\sum_sU_s-\sum_tU^t\succeq 0\\ \sum_k\upsilon_kT_k-\sum_sV_s-\sum_tV^t\succeq 0\end{array}\right\}\end{array}$$

and tightness factor of this bound does not exceed $\max[\vartheta(2\kappa),\pi/2]$ where $\kappa=\max\limits_{\alpha\leq q_{\rm s}}\min[p_\alpha,q_\alpha]$. □

5.6.4 Spectratopic version of Proposition 5.6.1

Proposition 5.6.1 admits a "spectratopic version," in which ellitopes $\mathcal{X}$ and $\mathcal{B}_*$ given by (5.72) are replaced by the pair of *spectratopes*

$$\begin{array}{rcl}\mathcal{X} & = & \{Py : y \in \mathcal{Y}\} \subset \mathbf{R}^n, \mathcal{Y} = \{y \in \mathbf{R}^N \ \& \ \exists t \in \mathcal{T} : T_k[y]^2 \preceq t_k I_{f_k}, k \le K\}, \\ & & T_k[y] = \sum_{j=1}^N y_j T^{jk},\ T^{jk} \in \mathbf{S}^{f_k}, \sum_k T_k^2[y] \succ 0 \ \forall y \ne 0 \end{array} \tag{5.83a}$$

$$\begin{array}{rcl}\mathcal{B}_* & = & \{Qz : z \in \mathcal{Z}\} \subset \mathbf{R}^m,\ \mathcal{Z} = \{z \in \mathbf{R}^M : \exists s \in \mathcal{S} :\ S_\ell^2[z] \preceq s_\ell I_{d_\ell},\ \ell \le L\}, \\ & & S_\ell[z] = \sum_{j=1}^M z_j S^{jk\ell},\ S^{jk\ell} \in \mathbf{S}^{d_\ell}, \sum_\ell S_\ell^2[z] \succ 0 \ \forall z \ne 0 \end{array} \tag{5.83b}$$

The spectratopic version of the statement reads as follows:

Proposition 5.6.2 *Given uncertain matrix (5.71) and spectratopes (5.83a) and (5.83b), consider convex optimization problem*

$$\begin{array}{rl}\mathrm{Opt} = \min\limits_{\mu,\upsilon,\lambda,U_s,V_s,U^t,V^t} & \left\{\frac{1}{2}[\phi_{\mathcal{S}}(\lambda[\mu]) + \phi_{\mathcal{T}}(\lambda[\upsilon])] : \right. \\ & \textit{subject to} \\ & \mu = \{M_\ell \in \mathbf{S}_+^{d_\ell}, \ell \le L\},\ \upsilon = \{\Upsilon_k \in \mathbf{S}_+^{f_k}, k \le K\},\ \lambda \ge 0 \\ & \left[\begin{array}{c|c} U_s & -Q^T A_s P \\ \hline -P^T A_s^T Q & V_s \end{array}\right] \succeq 0 \qquad (5.84a) \\ & \left[\begin{array}{c|c} U^t & -Q^T L_t^T \\ \hline -L_t Q & \lambda_t I_{p_t} \end{array}\right] \succeq 0,\ V^t - \lambda_t P^T R_t^T R_t P \succeq 0 \qquad (5.84b) \\ & \sum\limits_\ell S_\ell^{+,*}[M_\ell] - \sum\limits_s U_s - \sum\limits_t U^t \succeq 0 \qquad (5.84c) \\ & \sum\limits_k T_k^{+,*}[\Upsilon_k] - \sum\limits_s V_s - \sum\limits_t V^t \succeq 0 \qquad (5.84d) \end{array}$$

where

$$\lambda[\zeta] = [\mathrm{Tr}(Z_1); ...; \mathrm{Tr}(Z_I)] \ \textit{for}\ \zeta = \{Z_i \in \mathbf{S}^{k_i}, i \le I\}$$

and

$$S_\ell^{+,*}[V] = \left[\mathrm{Tr}(V S^{i\ell} S^{j\ell})\right]_{i,j \le M} \ \textit{for}\ V \in \mathbf{S}^{d_\ell},\ T_k^{+,*}[U] = \left[\mathrm{Tr}(U T^{ik} T^{jk})\right]_{i,j \le N} \ \textit{for}\ U \in \mathbf{S}^{f_k}.$$

Problem (5.84) is strictly feasible and solvable, and

$$\|\mathcal{A}\|_{\mathcal{X},\mathcal{B}} \le \mathrm{Opt} \le \varsigma\left(\sum\nolimits_k f_k\right) \varsigma\left(\sum\nolimits_\ell d_\ell\right) \max\left[\vartheta(2\kappa), \pi/2\right] \|\mathcal{A}\|_{\mathcal{X},\mathcal{B}}$$

where ϑ and κ are the same as in Proposition 5.6.1 and

$$\varsigma(J) = 2\sqrt{2\ln(2J)}.$$

Proof. For $Y \in \mathbf{S}^M$ and $X \in \mathbf{S}^N$ let us set

$$S_\ell^+[Y] = \sum_{i,j=1}^M Y_{ij} S^{i\ell} S^{j\ell},\ T_k^+[X] = \sum_{i,j=1}^N X_{ij} T^{ik} T^{jk},$$

so that

$$S_\ell^+[zz^T] = S_\ell^2[z],\ T_k^+[yy^T] = T_k^2[y] \tag{5.85}$$

and

$$\begin{array}{l}\mathrm{Tr}(VS_\ell^+[Y]) = \mathrm{Tr}(S_\ell^{+,*}[V]Y) \text{ for } V \in \mathbf{S}^{d_\ell}, Y \in \mathbf{R}^M,\\ \mathrm{Tr}(UT_k^+[X]) = \mathrm{Tr}(T_k^{+,*}[U]X) \text{ for } U \in \mathbf{S}^{f_k}, X \in \mathbf{R}^N.\end{array} \tag{5.86}$$

The proof of Proposition 5.6.2 is obtained from that (below referred to as "the proof") of Proposition 5.6.1 by the following modifications:

1. All references to (5.73) should be replaced with references to (5.84). Item [b] in 1^o of the proof now reads

   [b′] $x = Py$ with $T_k^2[y] \preceq \tau_k I_{f_k}$, $k \le K$, for some $\tau \in \mathcal{T}$ and $u = Qz$ for some z such that $S_\ell^2[z] \preceq \varsigma_\ell I_{d_\ell}$, $\ell \le L$, for some $\varsigma \in \mathcal{S}$.

   The last three lines in the chain (5.79) are replaced with

   $$\begin{aligned}\gamma &\le \frac{1}{2}\left[\sum_\ell \mathrm{Tr}([zz^T]S_\ell^{+,*}[M_\ell]) + \sum_k \mathrm{Tr}([yy^T]T_k^{+,*}[\Upsilon_k])\right] \text{ [by (5.84c) and(5.84d)]}\\ &= \frac{1}{2}\left[\sum_\ell \mathrm{Tr}(S_\ell^2[z]M_\ell) + \sum_k \mathrm{Tr}(T_k^2[y]\Upsilon_k)\right] \text{ [by (5.85) and (5.86)]}\\ &\le \frac{1}{2}\left[\sum_\ell \varsigma_\ell \mathrm{Tr}(M_\ell) + \sum_k \tau_k \mathrm{Tr}(\Upsilon_k)\right] \text{ [due to (b′) and } M_\ell \succeq 0, \Upsilon_k \succeq 0]\\ &\le \tfrac{1}{2}[\phi_{\mathcal{S}}(\lambda[\mu]) + \phi_{\mathcal{T}}(\lambda[\upsilon])] \text{ [since } \varsigma \in \mathcal{S}, \tau \in \mathcal{T}].\end{aligned}$$

2. Constraints (5.80) in (P) now read

   $$\left[\sum_\ell S_\ell^{+,*}[M_\ell] - \sum_s U_s - \sum_t U^t\right]^{\overline{S}} \succeq 0, \quad \left[\sum_k T_k^{+,*}[\Upsilon_k] - \sum_s V_s - \sum_t V^t\right]^{\overline{T}} \succeq 0.$$

   As a result, (5.81) becomes

   $$\begin{aligned}\mathrm{Opt} \le \max_{\overline{S},\overline{g},\overline{T},\overline{h}} \Bigg\{ &\sum_s \left\|\lambda(\mathcal{L}[\overline{S}^{1/2}Q^T A_s P \overline{T}^{1/2}])\right\|_1 + \sum_t \|\overline{S}^{1/2}Q^T L_t^T\|_{\mathrm{Fro}}\|\overline{T}^{1/2}P^T R_t^T\|_{\mathrm{Fro}} :\\ &\qquad \begin{array}{l}\overline{S} \succeq 0, \overline{g} \in \mathcal{S}, S_\ell^+[\overline{S}] \preceq \overline{g}_\ell I_{d_\ell},\ \ell \le L\\ \overline{T} \succeq 0, \overline{h} \in \mathcal{T}, T_k^+[\overline{T}] \preceq \overline{h}_k I_{f_k},\ k \le K\end{array} \Bigg\}\end{aligned} \tag{5.87}$$

3. The role of Lemma 6 in the proof is now played by the following fact.

   Lemma 7 *[118, Lemma 8] Let*

   $$\mathcal{V} = \{v \in \mathbf{R}^d : \exists r \in \mathcal{R} : R_j^2[v] \preceq r_j I_{\nu_j}, 1 \le j \le J\} \subset \mathbf{R}^d$$

   be a basic spectratope, $W \succeq 0$ be symmetric $d \times d$ matrix such that

   $$\exists r \in \mathcal{R} : R_j^+[W] \preceq r_j I_{\nu_j}, j \le J,$$

   and $\omega \sim \mathcal{N}(0, W)$. Denoting by $\rho(\cdot)$ the norm on $\mathbf{R}^d$ with the unit ball $\mathcal{V}$, we have

   $$\mathbf{E}\{\rho(\omega)\} \le \varsigma\left(\sum_j \nu_j\right),\ \varsigma(F) = 2\sqrt{2\ln(2F)}.$$

Chapter 6

Stochastic Mirror Descent for Large-Scale Sparse Recovery

Abstract

In this chapter we discuss an application of Stochastic Approximation to statistical estimation of high-dimensional sparse parameters. The proposed solution reduces to resolving a penalized stochastic optimization problem on each stage of a multistage algorithm; each problem being solved to a prescribed accuracy by the non-Euclidean Composite Stochastic Mirror Descent (CSMD) algorithm. Assuming that the problem objective is smooth and quadratically minorated and stochastic perturbations are sub-Gaussian, our analysis prescribes the method parameters which ensure fast convergence of the estimation error (the radius of a confidence ball of a given norm around the approximate solution). This convergence is linear during the first "preliminary" phase of the routine and is sublinear during the second "asymptotic" phase. We consider an application of the proposed approach to sparse Generalized Linear Regression problem. In this setting, we show that the proposed algorithm attains the optimal convergence of the estimation error under weak assumptions on the regressor distribution. We also present a numerical study illustrating the performance of the algorithm on high-dimensional simulation data.

6.1 Introduction

Our original motivation is the well known problem of (generalized) linear high-dimensional regression with random design. Formally, consider a dataset of N points $(\phi_i, \eta_i), i \in \{1, \ldots, N\}$, where $\phi_i \in \mathbf{R}^n$ are (random) features and $\eta_i \in \mathbf{R}$ are observations, linked by the following equation

$$\eta_i = \mathfrak{r}(\phi_i^T x_*) + \sigma\xi_i, \quad i \in [N] := \{1, \ldots, N\} \tag{6.1}$$

where $\xi_i \in \mathbf{R}$ are i.i.d. observation noises. The standard objective is to recover the unknown parameter $x_* \in \mathbf{R}^n$ of the Generalized Linear Regression (6.1) – which is assumed to belong to a given convex closed set X and to be *s-sparse,* i.e., to have at most $s \ll n$ non-vanishing entries from the data-set.

As mentioned before, we consider random design, where ϕ_i are i.i.d. random variables, so that the estimation problem of x_* can be recast as the following generic Stochastic Optimization problem:

$$g_* = \min_{x \in X} g(x), \quad \text{where} \quad g(x) = \mathbf{E}\big\{G\big(x, (\phi, \eta)\big)\big\}, \quad G(x, (\phi, \eta)) = \mathfrak{s}(\phi^T x) - \phi^T x\eta, \tag{6.2}$$

with $\mathfrak{s}(\cdot)$ any primitive of $\mathfrak{r}(\cdot)$, i.e., $\mathfrak{r}(t) = \mathfrak{s}'(t)$. The equivalence between the original and the stochastic optimization problems comes from the fact that x_* is a critical point of $g(\cdot)$, i.e., $\nabla g(x_*) = 0$ since, under mild assumptions, $\nabla g(x) = \mathbf{E}\{\phi[\mathfrak{r}(\phi^T x) - \mathfrak{r}(\phi^T x_*)]\}$. Hence, as soon as g has a unique minimizer (say, g is strongly convex over X), solutions of both problems are identical.

As a consequence, we shall focus on the generic problem (6.2), that has already been widely tackled. For instance, when given an observation sample (ϕ_i, η_i), $i \in [N]$, one may build a Sample Average Approximation (SAA) of the objective $g(x)$

$$\widehat{g}_N(x) = \frac{1}{N}\sum_{i=1}^{N} G(x, (\phi_i, \eta_i)) = \frac{1}{N}\sum_{i=1}^{N} [\mathfrak{s}(\phi_i^T x) - \phi_i^T x \eta_i] \tag{6.3}$$

and then solve the resulting problem of minimizing $\widehat{g}_N(x)$ over sparse x's. The celebrated ℓ_1-norm minimization approach allows to reduce this problem to convex optimization. We will provide a new algorithm adapted to this high-dimensional case, and instantiating it to the original problem 6.1.

Existing approaches and related works. Sparse recovery by Lasso and Dantzig Selector has been extensively studied [43, 47, 48, 135–137]. It computes a solution $\widehat{x}_N$ to the ℓ_1-penalized problem $\min_x \widehat{g}_N(x) + \lambda\|x\|_1$ where $\lambda \geq 0$ is the algorithm parameter [138]. This delivers "good solutions", with high probability for sparsity level s as large as $O\left(\frac{N\kappa_\Sigma}{\ln n}\right)$, as soon as the random regressors (the ϕ_i) are drawn independently from a normal distribution with a covariance matrix Σ such that $\kappa_\Sigma I \preceq \Sigma \preceq \rho\kappa_\Sigma I$[1], for some $\kappa_\Sigma > 0, \rho \geq 1$. However, computing this solution may be challenging in a very high-dimensional setting: even popular iterative algorithms, like coordinate descent, loops over a large number of variables. To mitigate this, randomized algorithms [139, 140], screening rules and working sets [141–143] may be used to diminish the size of the optimization problem at hand, while iterative thresholding [144–148] is a "direct" approach to enhance sparsity of the solution. Another approach relies on Stochastic Approximation (SA). As $\nabla G(x, (\phi_i, \eta_i)) = \phi_i(\mathfrak{r}(\phi_i^T x) - \eta_i)$ is an unbiased estimate of $\nabla g(x)$, iterative Stochastic Gradient Descent (SGD) algorithm may be used to build approximate solutions. Unfortunately, unless regressors ϕ are sparse or possess a special structure, standard SA leads to accuracy bounds for sparse recovery proportional to the dimension n which are essentially useless in the high-dimensional setting. This motivates non-Euclidean SA procedures, such as Stochastic Mirror Descent (SMD) [56], its application to sparse recovery enjoys almost dimension free convergence and it has been well studied in the literature. For instance, under bounded regressors and with sub-Gaussian noise, SMD reaches "slow rate" of sparse recovery of the type $g(\widehat{x}_N) - g_* = O\left(\sigma\sqrt{s\ln(n)/N}\right)$ where $\widehat{x}_N$ is the approximate solution after N iterations [149, 150]. Multistage routines may be used to improve the error estimates of SA under strong or uniform convexity assumptions [151–153]. However, they do not always hold, as in sparse Generalized Linear Regression, where they are replaced by Restricted Strong Convexity conditions. In that setting, the multistage procedure by [63] attains the rate $O\left(\frac{\sigma}{\kappa_\Sigma}\sqrt{\frac{s\ln n}{N}}\right)$ for the ℓ_2-error $\|\widehat{x}_N - x_*\|_2$ with high probability.[2] This is the best "asymptotic" rate attainable when solving (6.2). However, those algorithms have two major limitations. They both need a number of iterations to reach a given accuracy proportional to the initial error $R = \|x_* - x_0\|_1$ and the sparsity level s must be of order $O\left(\kappa_\Sigma\sqrt{\frac{N}{\ln n}}\right)$ for the sparse

[1] We use $A \preceq B$ for two symmetric matrices A and B if $B - A \succeq 0$, i.e. $B - A$ is positive semidefinite.

[2] Some flows in the proofs in [63] we fixed by [154].

linear regression. These limits may be seen as a consequence of dealing with *non-smooth* objective $g(x)$. Although it slightly restricts the scope of corresponding algorithms, we shall consider smooth objectives and algorithm for minimizing composite objectives (cf. [151, 155, 156]) to mitigate the aforementioned drawbacks of the multistage algorithms from [63, 154].

Principal contributions. We provide a refined analysis of *Composite Stochastic Mirror Descent (CSMD)* algorithms for computing sparse solutions to Stochastic Optimization problem leveraging smoothness of the objective. This leads to a new "aggressive" choice of parameters in a multistage algorithm with significantly improved performances compared to those in [63]. We summarize below some properties of the proposed procedure for problem (6.2).

Each stage of the algorithm is a specific CSMD recursion; They fall into two phases. During the first (preliminary) phase, the estimation error decreases linearly with the exponent proportional to $\frac{\kappa_\Sigma}{s\ln n}$. When it reaches the value $O\Big(\frac{\sigma s}{\sqrt{\kappa_\Sigma}}\Big)$, the second (asymptotic) phase begins, and its stages contain exponentially increasing number of iterations per stage, hence the estimation error decreases as $O\Big(\frac{\sigma s}{\kappa_\Sigma}\sqrt{\frac{\ln n}{N}}\Big)$ where N is the total iteration count.

Organization and notation The remaining of the paper is organized as follows. In Section 6.2, the general problem is set, and the multistage optimization routine and the study of its basic properties are presented. Then, in Section 6.3, we discuss the properties of the method and conditions under which it leads to "small error" solutions to sparse GLR estimation problems. Finally, a small simulation study illustrating numerical performance of the proposed routines in high-dimensional GLR estimation problem is presented in Section 6.3.3.

In the following, E is a Euclidean space and $\|\cdot\|$ is a norm on E; we denote $\|\cdot\|_*$ the conjugate norm (i.e., $\|x\|_* = \sup_{\|y\|\le 1}\langle y, x\rangle$). Given a positive semidefinite matrix $\Sigma \in \mathbf{S}_n$, for $x \in \mathbf{R}^n$ we denote $\|x\|_\Sigma = \sqrt{x^T\Sigma x}$ and for any matrix Q, we denote $\|Q\|_\infty = \max_{ij}|[Q]_{ij}|$. We use a generic notation c and C for absolute constants; a shortcut notation $a \lesssim b$ ($a \gtrsim b$) means that the ratio a/b (ratio b/a) is bounded by an absolute constant; the symbols $\bigvee,\bigwedge$ and the notation $(.)_+$ respectively refer to "maximum between", "minimum between" and "positive part".

6.2 Multistage Stochastic Mirror Descent for Sparse Stochastic Optimization

This section is dedicated to the formulation of the generic stochastic optimization problem, the description and the analysis of the generic algorithm.

6.2.1 Problem statement

Let X be a convex closed subset of an Euclidean space E and (Ω, P) a probability space. We consider a mapping $G : X \times \Omega \to \mathbf{R}$ such that, for all $\omega \in \Omega$, $G(\cdot,\omega)$ is convex on X and smooth, meaning that $\nabla G(\cdot,\omega)$ is Lipschitz continuous on X with a.s. bounded Lipschitz constant,

$$\forall x, x' \in X, \quad \|\nabla G(x,\omega) - \nabla G(x',\omega)\|_* \le \mathcal{L}(\omega)\|x - x'\|, \qquad \mathcal{L}(\omega) \le \nu \quad a.s.. \tag{6.4}$$

We define $g(x) := \mathbf{E}\{G(x,\omega)\}$, where $\mathbf{E}\{\cdot\}$ stands for the expectation with respect to ω, drawn from P. We shall assume that the mapping $g(\cdot)$ is finite, convex and differentiable on X and we

aim at solving the following stochastic optimization problem

$$\min_{x\in X}[g(x) = \mathbf{E}\{G(x,\omega)\}], \tag{6.5}$$

assuming it admits an s-sparse optimal solution x_* for some sparsity structure.

To solve this problem, stochastic oracle can be queried: when given at input a point $x \in X$, generates an $\omega \in \Omega$ from P and outputs $G(x,\omega)$ and $\nabla G(x,\omega) := \nabla_x G(x,\omega)$ (with a slight abuse of notations). We assume that the oracle is *unbiased*, i.e.,

$$\mathbf{E}\{\nabla G(x,\omega)\} = \nabla g(x), \qquad \forall x \in X.$$

To streamline presentation, we assume, as it is often the case in applications of stochastic optimization problem (6.5), that x_* is unconditional, i.e., $\nabla g(x_*) = 0$. or stated otherwise $\mathbf{E}\{\nabla G(x_*,\omega)\} = 0$; we also suppose the sub-Gaussianity of $\nabla G(x_*,\omega)$, namely that, for some $\sigma_* < \infty$

$$\mathbf{E}\Big\{\exp\Big(\|\nabla G(x_*,\omega)\|_*^2/\sigma_*^2\Big)\Big\} \le \exp(1). \tag{6.6}$$

6.2.2 Composite Stochastic Mirror Descent algorithm

As mentioned in the introduction, (stochastic) optimization over the set of sparse solutions can be done through "composite" techniques. We take a similar approach here, by transforming the generic problem (6.5) into the following *composite Stochastic Optimization problem*, adapted to some norm $\|\cdot\|$, and parameterized by $\kappa \ge 0$,

$$\min_{x\in X}\big[F_\kappa(x) := \tfrac{1}{2}g(x) + \kappa\|x\| = \tfrac{1}{2}\mathbf{E}\{G(x,\omega)\} + \kappa\|x\|\big]. \tag{6.7}$$

The purpose of this section is to derive a new (proximal) algorithm. We first provide necessary backgrounds and notations.

Proximal setup, Bregman divergences and Proximal mapping. Let B be the unit ball of the norm $\|\cdot\|$ and $\theta : B \to \mathbf{R}$ be a *distance-generating function (d.-g.f.)* of B, i.e., a continuously differentiable convex function which is strongly convex with respect to the norm $\|\cdot\|$,

$$\langle\nabla\theta(x) - \nabla\theta(x'), x - x'\rangle \ge \|x - x'\|^2, \quad \forall x, x' \in X.$$

We assume w.l.o.g. that $\theta(x) \ge \theta(0) = 0$ and denote $\Theta = \max_{\|z\|\le 1}\theta(z)$.

We now introduce a local and renormalized version of the d.-g.f. θ.

Definition 18 *For any $x_0 \in X$, let $X_R(x_0) := \{z \in X : \|z - x_0\| \le R\}$ be the ball of radius R around x_0. It is equipped with the d.-g.f. $\vartheta_{x_0}^R(z) := R^2\theta\left((z - x_0)/R\right)$.*

Note that $\vartheta_{x_0}^R(z)$ is strongly convex on $X_R(x_0)$ with modulus 1, $\vartheta_{x_0}^R(x_0) = 0$, and $\vartheta_{x_0}^R(z) \le \Theta R^2$.

Definition 19 *Given $x_0 \in X$ and $R > 0$, the Bregman divergence V associated to ϑ is defined by*

$$V_{x_0}(x,z) = \vartheta_{x_0}^R(z) - \vartheta_{x_0}^R(x) - \langle\nabla\vartheta_{x_0}^R(x), z - x\rangle, \quad x, z \in X.$$

We can now define *composite proximal mapping* on $X_R(x_0)$ [85, 156] with respect to some convex and continuous mapping $h : X \to \mathbf{R}$.

Definition 20 *The composite proximal mapping with respect to h and x is defined by*

$$\begin{aligned}\mathrm{Prox}_{h,x_0}(\zeta,x) &:= \operatorname*{argmin}_{z\in X_R(x_0)}\left\{\langle\zeta,z\rangle+h(z)+V_{x_0}(x,z)\right\}\\ &= \operatorname*{argmin}_{z\in X_R(x_0)}\left\{\langle\zeta-\nabla\vartheta^R_{x_0}(x),z\rangle+h(z)+\vartheta^R_{x_0}(z)\right\}\end{aligned}\tag{6.8}$$

If (6.8) can be efficiently solved to high accuracy and Θ is "not too large" (we refer to [85, 151, 157]); those setups will be called "prox-friendly". We now introduce the main building block of our algorithm, the Composite Stochastic Mirror Descent.

Composite Stochastic Mirror Descent algorithm. Given a sequence of positive *step sizes* $\gamma_i>0$, the *Composite Stochastic Mirror Descent* (CSMD) is defined by the following recursion

$$x_i = \mathrm{Prox}_{\gamma_i h,x_0}(\gamma_{i-1}\nabla G(x_{i-1},\omega_i),x_{i-1}),\quad x_0\in X.\tag{6.9}$$

After m steps of CSMD, the final output is $\widehat{x}_m$ (approximate solution) defined by

$$\widehat{x}_m=\frac{\sum_{i=0}^{m-1}\gamma_i x_i}{\sum_{i=0}^{m-1}\gamma_i}\tag{6.10}$$

For any integer $L\in\mathbf{N}$, we can also define the L-minibatch CSMD. Let $\omega_i^{(L)}=[\omega_i^1,...,\omega_i^L]$ be i.i.d. realizations of ω_i. The associated (average) stochastic gradient is then simply defined as

$$H\left(x_{i-1},\omega_i^{(L)}\right)=\frac{1}{L}\sum_{\ell=1}^{L}\nabla G(x_{i-1},\omega_i^\ell),$$

which yields the following recursion for the L-minibatch CSMD recursion:

$$x_i^{(L)} = \mathrm{Prox}_{\gamma_i h,x_0}\left(\gamma_{i-1}H\left(x_{i-1},\omega_i^{(L)}\right),x_{i-1}^{(L)}\right),\quad x_0\in X,\tag{6.11}$$

with its approximate solution $\widehat{x}_m^{(L)}=\sum_{i=0}^{m-1}\gamma_i x_i^{(L)}/\sum_{i=0}^{m-1}\gamma_i$ after m iterations.

From now on, we set $h(x)=\kappa\|x\|$.

Proposition 6.2.1 *If step-sizes are constant, i.e., $\gamma_i\equiv\gamma\le(4\nu)^{-1}$, $i=0,1,...$, and the initial point $x_0\in X$ such that $x_*\in X_R(x_0)$ then for any $t\gtrsim\sqrt{1+\ln m}$, with probability at least $1-4e^{-t}$*

$$F_\kappa(\widehat{x}_m)-F_\kappa(x_*)\lesssim m^{-1}\left[\gamma^{-1}R^2(\Theta+t)+\kappa R+\gamma\sigma_*^2(m+t)\right],\tag{6.12}$$

and the approximate solution $\widehat{x}_m^{(L)}$ of the L-minibatch CSMD satisfies

$$F_\kappa(\widehat{x}_m^{(L)})-F_\kappa(x_*)\lesssim m^{-1}\left[\gamma^{-1}R^2(\Theta+t)+\kappa R+\gamma\sigma_*^2\Theta L^{-1}(m+t)\right].\tag{6.13}$$

For the sake of clarity and conciseness, we denote $\mathrm{CSMD}(x_0,\gamma,\kappa,R,m,L)$ the approximate solution $\widehat{x}_m^{(L)}$ computed after m iterations of L-minibatch CSMD algorithm with initial point x_0, step-size γ, and radius R using recursion (6.11).

6.2.3 Main contribution: a multistage adaptive algorithm

Our approach to find sparse solution to the original stochastic optimization problem (6.5) consists in solving a sequence of auxiliary composite problems (6.7), with their sequence of parameters (κ, x_0, R) defined recursively. For the latter, we need to infer the quality of approximate solution to (6.5). To this end, we introduce the following *Reduced Strong Convexity* (RSC) assumption, satisfied in the motivating example (it is discussed in the appendix for the sake of fluency):

Assumption [RSC] There exist some $\delta > 0$ and $\rho < \infty$ such that for any feasible solution $\widehat{x} \in X$ to the composite problem (6.7),

$$\|\widehat{x} - x_*\| \leq \delta \left[\rho s \kappa + \frac{(F_\kappa(\widehat{x}) - F_\kappa(x_*))_+}{\kappa}\right]. \tag{6.14}$$

Given the different problem parameters $s, \nu, \delta, \rho, \kappa, R$ and some initial point $x_0 \in X$ such that $x_* \in X_R(x_0)$ Algorithm 1 works in stages. Each stage represents a run of CSMD algorithm with properly set penalty parameter κ. More precisely, at stage $k+1$, given the approximate solution $\widehat{x}_m^k$ of stage k, a new instance of CSMD is initialized on $X_{R_{k+1}}(x_0^{k+1})$ with $x_0^{k+1} = \widehat{x}_m^k$ and $R_{k+1} = R_k/2$.

Furthermore, those stages are divided into two phases which we refer to as *preliminary* and *asymptotic*:

Preliminary phase: During this phase, the step-sizes γ and the number of CSMD iterations per stage are fixed; the error of approximate solutions converges linearly with the total number of calls to stochastic oracle. This phase terminates when the error of approximate solution becomes independent of the initial error of the algorithm; then the asymptotic phase begins.

Asymptotic phase: In this phase, the step-size decreases and the length of the stage increases linearly; the solution converges sublinearly, with the "standard" rate $O\big(N^{-1/2}\big)$ where N is the total number of oracle calls. When expensive proximal computation (6.8) results in high numerical cost of the iterative algorithm, minibatches are used to keep the number of iterations per stage fixed.

In the algorithm description, $\overline{K}_1$ and $\overline{K}_2 \asymp 1+\log(\frac{N}{m_0})$ stand for the respective maximal number of stages of the two phases of the method, here, $m_0 \asymp s\rho\nu\delta^2(\Theta + t)$ is the length of stages of the first (preliminary) phase. The pseudo-code for the variant of the asymptotic phase with minibatches is given in Algorithm 2.

The following theorem states the main result of this paper, an upper bound on the precision of the estimator computed by our multistage method.

Theorem 6.2.1 *Assume that the total sample budget satisfies $N \geq m_0$, so that at least one stage of the preliminary phase of Algorithm 1 is completed, then for $t \gtrsim \sqrt{\ln N}$ the approximate solution $\widehat{x}_N$ of Algorithm 1 satisfies, with probability at least $1 - C(\overline{K}_1 + \overline{K}_2)e^{-t}$,*

$$\|\widehat{x}_N - x_*\| \lesssim R\exp\left\{-\frac{c}{\delta^2\rho\nu}\frac{N}{s(\Theta+t)}\right\} + \delta^2\rho\sigma_* s\sqrt{\frac{\Theta+t}{N}}.$$

The corresponding solution $\widehat{x}_N^{(b)}$ of the minibatch Algorithm 2 satisfies with probability $\geq 1 - C(\overline{K}_1 + \widetilde{K}_2)e^{-t}$

$$\|\widehat{x}_N^{(b)} - x_*\| \lesssim R\exp\left\{-\frac{c}{\delta^2\rho\nu}\frac{N}{s\,(\Theta+t)}\right\} + \delta^2\rho\sigma_* s\sqrt{\frac{\Theta\,(\Theta+t)}{N}}.$$

Algorithm 1 CSMD-SR

Initialization : Initial point $x_0 \in X$, step-size $\gamma = (4\nu)^{-1}$, initial radius R_0, confidence level t, total budget N.

Set $m_0 \asymp s\rho\nu\delta^2(\Theta + t)$, $\overline{K}_1 \asymp \ln\left(\frac{R_0^2\nu}{\delta^2\sigma_*^2\rho s}\right) \wedge \frac{N}{m_0}$, $L = 1$

if $R_0 \gtrsim \sigma_*\delta\sqrt{\frac{\rho s}{\nu}}$ **continue** with preliminary stage,
else proceed directly to asymptotic phase
end
for stage $k = 1, \ldots, \overline{K}_1$ **do** ▷ Preliminary Phase
  Set $\kappa_k \asymp R_k(\delta\rho s)^{-1}$
  Compute approximate solution $\widehat{x}^k_{m_0}$ =CSMD$(x_0, \gamma, \kappa_k, R_k, m_0, L)$ at stage k
  Reset the prox-center $x_0 = \widehat{x}^k_{m_0}$
  Set $R_k = R_{k-1}/2$
end for
Set $\widehat{x}_N = \widehat{x}^{\overline{K}_1}_{m_0}$, $B = N - m_0\overline{K}_1$, $m_1 \asymp m_0$
if $m_1 > B$ **output :** $\widehat{x}_N$ and **return**; **endif** ▷ Asymptotic Phase
Set $r_0 = R_{\overline{K}_1}$
Set $k = 1$
while $m_k \le B$ **do**
  Set $\kappa_k \asymp 2^{-k}\sigma_*(\rho\nu s)^{-1/2}$, $\gamma_k \asymp 4^{-k}\nu^{-1}$
  Compute approximate solution $\widehat{x}^k_{m_k}$ =CSMD$(x_0, \gamma_k, \kappa_k, r_k, m_k, L)$ at stage k
  Reset the prox-center $x_0 = \widehat{x}^k_{m_k}$
  Set $B = B - m_k$, $k = k + 1$, $r_k = r_{k-1}/2$, $m_k \asymp 4^k m_0$
end while

output : $\widehat{x}_N = \widehat{x}^k$

where $\widetilde{K}_2 \asymp 1 + \ln\left(\frac{N}{\Theta m_0}\right)$ *is the bound for the number of stages of the asymptotic phase of the minibatch algorithm.*

Remark 6.2.1 *Along with the oracle computation, proximal computation to be implemented at each iteration of the algorithm is an important part of the computational cost of the method. It becomes even more important during the asymptotic phase when number of iterations per stage increases exponentially fast with the stage count, and may result in poor real-time convergence. The interest of minibatch implementation of the second phase of the algorithm is in reducing drastically the number of iterations per asymptotic stage. The price to be paid is an extra factor* $\sqrt{\Theta}$ *that could also theoretically hinder convergence. However, in the problems of interest (sparse and group-sparse recovery, low rank matrix recovery)* Θ *is logarithmic in problem dimension. Furthermore, in our numerical experiments we did not observe any accuracy degradation when using the minibatch variant of the method.*

Algorithm 2 Asymptotic phase of CSMD-SR with minibatch

Input : The approximate solution $\widehat{x}_{m_0}^{\overline{K}_1}$ at the end of the preliminary stage, step-size parameter γ, radius at the end of the preliminary phase $R_{\overline{K}_1}$, initial batch size $\ell_1 \asymp \Theta$

Set $r_0 = R_{\overline{K}_1}$, $x_0 = \widehat{x}_{m_0}^{\overline{K}_1}$, $B = N - m_0\overline{K}_1$ ▷ Asymptotic Phase
$k = 1$
while $m_0\ell_k \le B$ **do**
  $\kappa_k \asymp 2^{-k}\sigma_*(\rho\nu s)^{-1/2}$
  Compute approximate solution $\widehat{x}_{m_0}^k$ =CSMD$(x_0, \gamma_k, \kappa_k, r_k, m_0, L = \ell_k)$ at stage k
  Reset the prox-center $x_0 = \widehat{x}_{m_0}^k$
  Set $B = B - m_0\ell_k$, $k = k + 1$, $r_k = r_{k-1}/2$, $\ell_k \asymp 4^k\ell_1$
end while

output: $\widehat{x}_N^{(b)} = \widehat{x}_{m_2}^k$

6.3 Sparse generalized linear regression by stochastic approximation

6.3.1 Problem setting

We now consider again the original problem of recovery of a s-sparse signal $x_* \in X \subset \mathbf{R}^n$ from random observations defined by

$$\eta_i = \mathfrak{r}(\phi_i^T x_*) + \sigma\xi_i, \quad i = 1, 2, ..., N, \tag{6.15}$$

where $\mathfrak{r} : \mathbf{R} \to \mathbf{R}$ is some non-decreasing and continuous "activation function", and $\phi_i \in \mathbf{R}^n$ and $\xi_i \in \mathbf{R}$ are mutually independent. We assume that ξ_i are sub-Gaussian, i.e., $\mathbf{E}\{e^{\xi_i^2}\} \le \exp(1)$, while regressors ϕ_i are bounded, i.e., $\|\phi_i\|_\infty \le \overline{\nu}$. We also denote $\Sigma = \mathbf{E}\{\phi_i\phi_i^T\}$, with $\Sigma \succeq \kappa_\Sigma I$ with some $\kappa_\Sigma > 0$, and $\|\Sigma_j\|_\infty \le \upsilon < \infty$.
We will apply the machinery developed in Section 6.2, with respect to

$$g(x) = \mathbf{E}\{\mathfrak{s}(\phi^T x) - x^T\phi\eta\}$$

where $\mathfrak{r}(t) = \nabla\mathfrak{s}(t)$ for some convex and continuously differentiable $\mathfrak{s}$, applied with the norm $\|\cdot\| = \|\cdot\|_1$ (hence $\|\cdot\|_* = \|\cdot\|_\infty$), from some initial point $x_0 \in X$ such that $\|x_* - x_0\|_1 \le R$. It remains to prove that the different assumptions of Section 6.2 are satisfied.

Proposition 6.3.1 *Assume that $\mathfrak{r}$ is $\overline{r}$-Lipschitz continuous and $\underline{r}$-strongly monotone (i.e., $|\mathfrak{r}(t) - \mathfrak{r}(t')| \ge \underline{r}|t - t'|$ which implies that $\mathfrak{s}$ is $\underline{r}$-strongly convex) then*

1. *[Smoothness] $G(\cdot, \omega)$ is $\mathcal{L}(\omega)$-smooth with $\mathcal{L}(\omega) \le \overline{r}\overline{\nu}^2$.*
2. *[Quadratic minoration] g satisfies*

$$g(x) - g(x_*) \ge \tfrac{1}{2}\underline{r}\|x - x_*\|_\Sigma^2. \tag{6.16}$$

3. *[Reduced Strong Convexity] Assumption [RSC] holds with $\delta = 1$ and $\rho = (\kappa_\Sigma\underline{r})^{-1}$.*
4. *[Sub-Gaussianity] $\nabla G(x_*, \omega_i)$ is $\sigma^2\overline{\nu}^2$-sub Gaussian.*

The proof is postponed to the appendix. The last point is a consequence of a generalization of the Restricted Eigenvalue property [48], that we detail below (as it gives insight on why Proposition 6.3.1 holds). This condition, that we state and call $\mathbf{Q}(\lambda,\psi)$ in the following Lemma 8, and is reminiscent of [83] with the corresponding assumptions of [158, 159].

Lemma 8 *Let $\lambda > 0$ and $0 < \psi \leq 1$, and suppose that for all subsets $I \subset \{1,...,n\}$ of cardinality smaller than s the following property is verified:*

$$\forall z \in \mathbf{R}^n \quad \|z_I\|_1 \leq \sqrt{\frac{s}{\lambda}}\|z\|_\Sigma + \tfrac{1}{2}(1-\psi)\|z\|_1 \qquad \mathbf{Q}(\lambda,\psi)$$

where z_I is obtained by zeroing all its components with indices $i \notin I$. If $g(\cdot)$ satisfies the quadratic minoration condition, *i.e., for some $\mu > 0$,*

$$g(x) - g(x_*) \geq \tfrac{1}{2}\mu\|x - x_*\|_\Sigma^2, \tag{6.17}$$

and that $\widehat{x}$ is an admissible solution to (6.7) satisfying, with probability at least $1-\varepsilon$,

$$F_\kappa(\widehat{x}) \leq F_\kappa(x_*) + \upsilon.$$

Then, with probability at least $1-\varepsilon$,

$$\|\widehat{x} - x_*\|_1 \leq \frac{s\kappa}{\lambda\mu\psi} + \frac{\upsilon}{\kappa\psi}. \tag{6.18}$$

Remark 6.3.1 *Condition $\mathbf{Q}(\lambda,\psi)$ generalizes the classical Restricted Eigenvalue (RE) property [48] and Compatibility Condition [135], and is the most relaxed condition under which classical bounds for the error of ℓ_1-recovery routines were established. Validity of $\mathbf{Q}(\lambda,\psi)$ with some $\lambda > 0$ is necessary for Σ to possess the celebrated* null-space property *[160]*

$$\exists \psi > 0 : \max_{I,\,|I|\leq s} \|z_I\|_1 \leq \tfrac{1}{2}(1-\psi)\|z\|_1 \ \ \forall z \in \mathrm{Ker}(\Sigma)$$

which is necessary and sufficient for the s-goodness of Σ (i.e., $\widehat{x} \in \mathrm{Argmin}_u \{\|u\| : \Sigma u = \Sigma x_*\}$ reproduces exactly every s-sparse signal x_* in the noiseless case).
When Σ possesses the nullspace property, $\mathbf{Q}(\lambda,\psi)$ may hold for Σ with nontrivial kernel; this is typically the case for random matrices [158, 161] such as rank deficient Wishart matrices, etc. When Σ is a regular matrix, condition $\mathbf{Q}(\lambda,\psi)$ may also holds with constant λ which is much higher that the minimal eigenvalue of Σ when the eigenspace corresponding to small eigenvalues of Σ does not contain vectors z with $\|z_I\|_1 > \frac{1}{2}(1-\psi)\|z\|_1$.

Special cases. The quadratic minoration bound (6.16) for $g(x) - g(x_*)$ is usually overly pessimistic. Indeed, consider for instance, Gaussian regressor $\phi \sim \mathcal{N}(0,\Sigma)$ (even if they are not a.s. bounded, this is for illustration purposes) and activation $\mathfrak{r}$, define for some $0 \leq \alpha \leq 1$ (with the convention, $0/0 = 0$)

$$\mathfrak{r}(t) = \begin{cases} t, & |t| \leq 1, \\ \mathrm{sign}(t)[\alpha^{-1}(|t|^\alpha - 1) + 1], & |t| > 1. \end{cases} \tag{6.19}$$

When passing from ϕ to $\varphi = \Sigma^{-1/2}\phi$ and from x to $z = \Sigma^{1/2}x$ and using the fact that

$$\varphi = \frac{zz^T}{\|z\|_2^2}\varphi + \underbrace{\left(I - \frac{zz^T}{\|z\|_2^2}\right)\varphi}_{=:\chi}$$

with independent $\frac{zz^T}{\|z\|_2^2}\varphi$ and χ, with $\mathbf{E}\{\chi\} = 0$, we obtain

$$\begin{aligned} H(z) =& \mathbf{E}\{\varphi[\mathfrak{r}(\varphi^T z)]\} = \mathbf{E}\left\{\frac{zz^T}{\|z\|_2^2}\varphi\,\mathfrak{r}(\varphi^T z)\right\} \\ =& \frac{z}{\|z\|_2}\mathbf{E}\left\{\varsigma\mathfrak{r}(\varsigma\|z\|_2)\right\} = \frac{\Sigma^{1/2}x}{\|x\|_\Sigma}\mathbf{E}\left\{\varsigma\mathfrak{r}(\varsigma\|x\|_\Sigma)\right\} \end{aligned}$$

where $\varsigma \sim \mathcal{N}(0,1)$. Thus, $H(\Sigma^{1/2}x)$ is proportional to $\frac{\Sigma^{1/2}x}{\|x\|_\Sigma}$ with coefficient

$$h\big(\|x\|_\Sigma\big) = \mathbf{E}\left\{\varsigma\mathfrak{r}(\varsigma\|x\|_\Sigma)\right\}.$$

Figure 6.1 represents the mapping h for different values of α (on the left), along with the dependence on r of moduli of strong monotonicity of corresponding mappings H on the centered at the origin $\|\cdot\|_2$-ball of radius r (on the right).

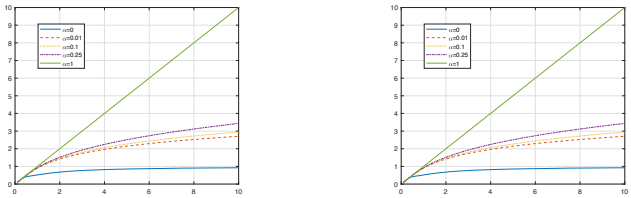


Figure 6.1: Given the activation function $\mathfrak{r}$ in (6.19) and $\alpha = (0, 0.01, 0.1, 0.25, 1)$; left plot: mappings h; right plot: moduli of strong monotonicity of mappings H on $\{z : \|z\|_2 \le r\}$ as function of r.

In the case of linear regression where $\mathfrak{r}(t) = t$, it holds

$$\begin{aligned} g(x) &= \mathbf{E}\big\{\tfrac{1}{2}(\phi^T x)^2 - x^T\phi\eta\big\} \\ &= \tfrac{1}{2}\mathbf{E}\big\{(\phi^T(x_* - x))^2 - (\phi^T x_*)^2\big\} \\ &= \tfrac{1}{2}(x - x_*)^T\Sigma(x - x_*) - \tfrac{1}{2}x_*^T\Sigma x_* \\ &= \tfrac{1}{2}\|x - x_*\|_\Sigma^2 - \tfrac{1}{2}\|x_*\|_\Sigma^2 \end{aligned}$$

and $\nabla G(x,\omega) = \phi\phi^T(x - x_*) - \sigma\xi\phi$. In this case $\mathcal{L}(\omega) \le \|\phi\phi^T\|_\infty \le \overline{\nu}^2$.

6.3.2 Stochastic Mirror Descent algorithm

In this section, we describe the statistical properties of approximate solutions of Algorithm 1 when applied to the sparse recovery problem. We shall use the following distance-generating function of the ℓ_1-ball of $\mathbf{R}^n$ (cf. [151, Section 5.7.1])

$$\theta(x) = \frac{c}{p}\|x\|_p^p, \quad p = \begin{cases} 2, & n = 2 \\ 1 + \frac{1}{\ln(n)}, & n \ge 3, \end{cases} \quad c = \begin{cases} 2, & n = 2, \\ e\ln n, & n \ge 3. \end{cases} \tag{6.20}$$

It immediately follows that θ is strongly convex with modulus 1 w.r.t. the norm $\|\cdot\|_1$ on its unit ball, and that $\Theta \le e\ln n$. In particular, Theorem 6.2.1 entails the following statement.

Proposition 6.3.2 *For $t \gtrsim \sqrt{\ln N}$, assuming the samples budget is large enough, i.e., $N \geq m_0$ (so that at least one stage of the preliminary phase of Algorithm 1 is completed), the approximate solution $\widehat{x}_N$ output satisfies with probability at least $1 - Ce^{-t}\ln N$,*

$$\|\widehat{x}_N - x_*\|_1 \lesssim R\exp\left\{-c\frac{\underline{r}\kappa_\Sigma}{\overline{r}\overline{\nu}^2}\frac{N}{s(\ln n + t)}\right\} + \frac{\sigma\overline{\nu}s}{\underline{r}\kappa_\Sigma}\sqrt{\frac{\ln n + t}{N}} \tag{6.21}$$

The corresponding solution $\widehat{x}_N^{(b)}$ of the minibatch variant of the algorithm satisfies with probability $\geq 1 - Ce^{-t}\ln N$,

$$\|\widehat{x}_N^{(b)} - x_*\|_1 \lesssim R\exp\left\{-c\frac{\underline{r}\kappa_\Sigma}{\overline{r}\overline{\nu}^2}\frac{N}{s\,(\ln n + t)}\right\} + \frac{\sigma\overline{\nu}s}{\underline{r}\kappa_\Sigma}\sqrt{\frac{\ln n\,(\ln n + t)}{N}}$$

Remark 6.3.2 *Bounds for the ℓ_1-norm of the error $\widehat{x}_N - x_*$ (or $\widehat{x}_N^{(b)} - x_*$) established in Proposition 6.3.2 allows us to quantify prediction error $g(\widehat{x}_N) - g(x_*)$ (and $g(\widehat{x}_N^{(b)}) - g(x_*)$), and also lead to bounds for $\|\widehat{x}_N - x_*\|_\Sigma$ and $\|\widehat{x}_N - x_*\|_2$ (respectively, for $\|\widehat{x}_N^{(b)} - x_*\|_\Sigma$ and $\|\widehat{x}_N^{(b)} - x_*\|_2$). For instance, Proposition 6.2.1 in the present setting implies the bound on the prediction error after N steps of the algorithm that reads*

$$g(\widehat{x}_N) - g(x_*) \lesssim \frac{R^2\kappa_\Sigma\underline{r}}{s}\exp\left\{-\frac{c\kappa_\Sigma\underline{r}}{\delta^2\overline{r}\overline{\nu}^2}\frac{N}{s(\Theta + t)}\right\} + \frac{\sigma^2\overline{\nu}^2 s(\Theta + t)}{\kappa_\Sigma\underline{r}N}$$

with probability $\geq 1 - C\ln N e^{-t}$. We conclude by (6.16) that

$$\begin{aligned}\|\widehat{x}_N - x_*\|_2^2 &\leq \kappa_\Sigma^{-1}\|\widehat{x}_N - x_*\|_\Sigma^2 \leq 2\kappa_\Sigma^{-1}\underline{r}^{-1}[g(\widehat{x}_N) - g(x_*)]\\ &\lesssim \frac{R^2}{s}\exp\left\{-\frac{c\kappa_\Sigma\underline{r}}{\delta^2\overline{r}\overline{\nu}^2}\frac{N}{s(\Theta + t)}\right\} + \frac{\sigma^2\overline{\nu}^2 s(\Theta + t)}{\kappa_\Sigma^2\underline{r}^2 N}.\end{aligned}$$

In other words, the error $\|\widehat{x}_N - x_*\|_2$ converges geometrically to the "asymptotic rate" $\frac{\sigma\overline{\nu}}{\kappa_\Sigma\underline{r}}\sqrt{\frac{s(\Theta+t)}{N}}$ which is the "standard" rate established in the setting (cf. [48, 138, 144], etc).

Remark 6.3.3 *The proposed approach allows also to address the situation in which regressors are not a.s. bounded. For instance, consider the case of random regressors with i.i.d sub-Gaussian entries such that*

$$\forall j \leq n, \quad \mathbf{E}\left[\exp\left(\frac{[\phi_i]_j^2}{\varkappa^2}\right)\right] \leq 1.$$

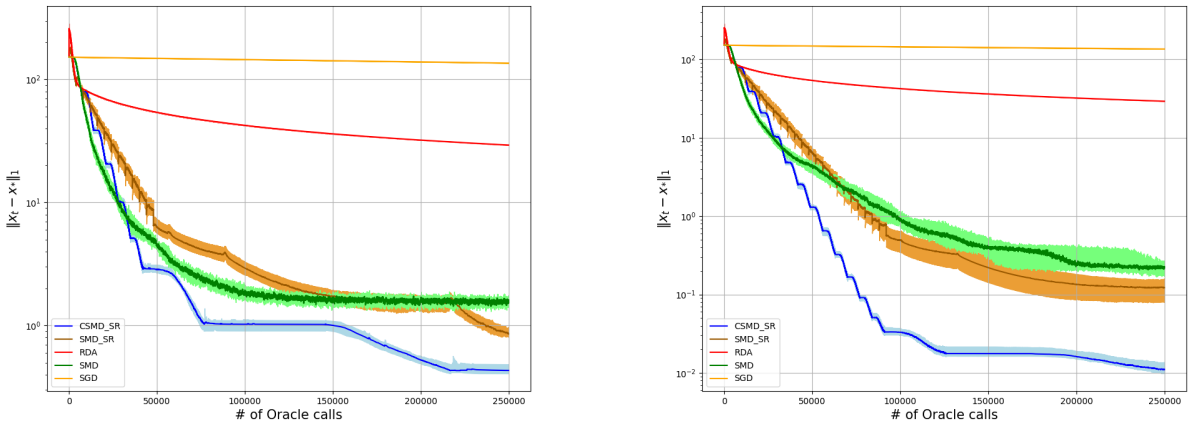
Using the fact that the maximum of uniform norms $\|\phi_i\|_\infty$, $1 \leq i \leq m$, concentrates around $\varkappa\sqrt{\ln mn}$ along with independence of noises ξ_i of ϕ_i, the "smoothness" and "sub-Gaussianity" assumptions of Proposition 6.3.2 can be stated "conditionally" to the event

$$\left\{\omega : \max_{i\leq m}\|\phi_i\|_\infty^2 \lesssim \varkappa^2(\ln[mn] + t)\right\}$$

of probability greater than $1 - e^{-t}$. For instance, when replacing the bound for the uniform norm of regressors with $\varkappa^2(\ln[mn] + t)$ in the definition of algorithm parameters and combining with appropriate deviation inequality for martingales (cf., e.g., [162]), one arrives at the bound for the error $\|\widehat{x}_N - x_*\|_1$ of Algorithm 1 which is similar to (6.21) of Proposition 6.3.2 in which $\overline{\nu}$ is replaced with $\varkappa\sqrt{\ln[mn] + t}$.

6.3.3 Numerical experiments

In this section, we present results of a small simulation study illustrating the theoretical part of the previous section. [3] We consider the GLR model (6.15) with activation function (6.19) where $\alpha = 1/2$. In our simulations, x_* is an s-sparse vector with s nonvanishing components sampled independently from the standard s-dimensional Gaussian distribution; regressors ϕ_i are sampled from a multivariate Gaussian distribution $\phi \sim \mathcal{N}(0, \Sigma)$, where Σ is a diagonal covariance matrix with diagonal entries $\sigma_1 \leq ... \leq \sigma_n$. In Figure 6.2 we report on the experiment in which we compare the performance of the CSMD-SR algorithm from Section 6.2.3 to that of four other methods. The contenders are (1) "vanilla" non-Euclidean SMD algorithm constrained to the ℓ_1-ball equipped with the distance generating function (6.20), (2) composite non-Euclidean dual averaging algorithm (p-Norm RDA) from [163], (3) multistage SMD-SR of [64], and (4) "vanilla" Euclidean SGD. The regularization parameter of the ℓ_1 penalty in (2) is set to the theoretically optimal value $\lambda = 2\sigma\sqrt{2\log(n)/T}$. The corresponding dimension of the parameter space is $n = 500000$, the sparsity level of the optimal point x_* is $s = 200$, and the "total budget" of oracle calls is $N = 250000$; we use the identity regressor covariance matrix ($\Sigma = I_n$) and $\sigma \in \{0.001, 0.1\}$. To reduce computation time we use the minibatch versions of the multi-stage algorithms—CSMD-SR and algorithm (3), the data to compute stochastic gradient realizations $\nabla G(x_i, \omega) = \phi(\mathfrak{r}(\phi^T x_i) - \eta)$ at the current search point x_i being generated "on the fly." We repeat simulations 20 times and plot the median value along with the first and the last deciles of the error $\|\widehat{x}_i - x_*\|_1$ at each iteration of the algorithm against the number of oracle calls.



$\sigma = 0.1$ $\sigma = 0.001$

Figure 6.2: Comparison between CSMD-SR and baseline algorithms in Generalized Linear Regression problem: ℓ_1 error as a function of the number of oracle calls

The proposed method outperforms other algorithms which struggle to reach the regime where the stochastic noise is dominant.

[3] The reader is invited to check Section 6.6 of the supplementary material for more experimental results.

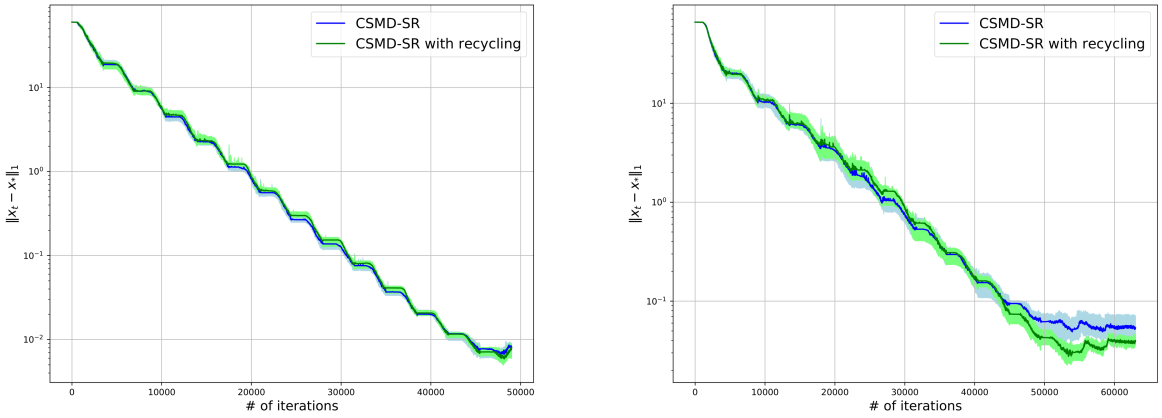


Figure 6.3: Preliminary stages of the CSMD-SR and its variant with data recycling: linear regression experiment (left pane), GLR with activation $\mathfrak{r}_{1/10}(t)$ (right pane).

In the second experiment we report on here, we study the behavior of the multistage algorithm derived from Algorithm 2 in which, instead of using independent data samples, we reuse the same data at each stage of the method. In Figure 6.3 we present results of comparison of the CSMD-SR algorithm with its variant with data recycle. This version is of interest as it attains fast the noise regime while using limited amount of samples. In our first experiment, we consider linear regression problem with parameter dimension $n = 100\,000$ and sparsity level $s = 75$ of the optimal solution; we consider the GLR model (6.15) with activation function $\mathfrak{r}_{1/10}(t)$ in the second experiment. We choose $\Sigma = I_n$ and $\sigma = 0.001$; we run 14 (preliminary) stages of the algorithm with $m_0 = 3500$ in the first simulation and $m_0 = 4500$ in the second. We believe that the results speak for themselves.

6.4 Proofs

We use notation $\mathbf{E}_i$ for conditional expectation given x_0 and $\omega_1, ..., \omega_i$.

6.4.1 Proof of Proposition 6.2.1

The result of Proposition 6.2.1 is an immediate consequence of the following statement.

Proposition 6.4.1 *Let*

$$f(x) = \tfrac{1}{2}g(x) + h(x), \quad x \in X.$$

In the situation of Section 6.2.2, let $\gamma_i \le (4\nu)^{-1}$ for all $i = 0, 1, ...$, and let $\widehat{x}_m$ be defined in (6.10), where x_i are iterations (6.9). Then for any $t \ge 2\sqrt{2 + \ln m}$ there is $\overline{\Omega}_m \subset \Omega$ such that

$\mathrm{Prob}(\overline{\Omega}_m) \geq 1 - 4e^{-t}$ *and for all* $\omega^m = [\omega_1, ..., \omega_m] \in \overline{\Omega}_m$,

$$\begin{aligned}\left(\sum_{i=0}^{m-1}\gamma_i\right)[f(\widehat{x}_m) - f(x_*)] &\leq \sum_{i=0}^{m-1}\left[\tfrac{1}{2}\gamma_i\langle\nabla g(x_i), x_i - x_*\rangle + \gamma_{i+1}(h(x_{i+1}) - h(x_*))\right] \\ &\leq V(x_0, x_*) + \gamma_0[h(x_0) - h(x_*)] - \gamma_m[h(x_m) - h(x_*)] \\ &\quad + V(x_0, x_*) + 15tR^2 + \sigma_*^2\left[7\sum_{i=0}^{m-1}\gamma_i^2 + 24t\overline{\gamma}^2\right]. \end{aligned} \tag{6.22}$$

In particular, when using the constant stepsize strategy with $\gamma_i \equiv \gamma$, $0 < \gamma \leq (4\nu)^{-1}$, *one has*

$$\begin{aligned}&\tfrac{1}{2}[g(\widehat{x}_m) - g(x_*)] + [h(\widehat{x}_m) - h(x_*)] \\ &\quad \leq \frac{V(x_0, x_*) + 15tR^2}{\gamma m} + \frac{h(x_0) - h(x_m)}{m} + \gamma\sigma_*^2\left(7 + \frac{24t}{m}\right). \end{aligned} \tag{6.23}$$

Proof. Denote $H_i = \nabla G(x_{i-1}, \omega_i)$. In the sequel, we use the shortcut notation $\vartheta(z)$ and $V(x, z)$ for $\vartheta_{x_0}^R(z)$ and $V_{x_0}(x, z)$ when exact values x_0 and R are clear from the context.

1^o. From the definition of x_i and of the composite prox-mapping (6.8) (cf. Lemma A.1 of [85]), we conclude that there is $\eta_i \in \partial h(x_i)$ such that

$$\langle\gamma_{i-1}H_i + \gamma_i\eta_i + \nabla\vartheta(x_i) - \nabla\vartheta(x_{i-1}), z - x_i\rangle \geq 0, \quad \forall\, z \in \mathcal{X},$$

implying, as usual [164], that $\forall z \in X$

$$\langle\gamma_{i-1}H_i + \gamma_i\eta_i, x_i - z\rangle \leq V(x_{i-1}, z) - V(x_i, z) - V(x_{i-1}, x_i).$$

In particular,

$$\begin{aligned}&\gamma_{i-1}\langle H_i, x_{i-1} - x_*\rangle + \gamma_i\langle\eta_i, x_i - x_*\rangle \\ &\leq V(x_{i-1}, x_*) - V(x_i, x_*) - V(x_{i-1}, x_i) + \gamma_{i-1}\langle H_i, x_{i-1} - x_i\rangle \\ &\leq V(x_{i-1}, x_*) - V(x_i, x_*) + \tfrac{1}{2}\gamma_{i-1}^2\|H_i\|_*^2. \end{aligned}$$

Observe that due to the Lipschitz continuity of $\nabla G(\cdot, \omega)$ one has

$$\nu\langle\nabla G(x, \omega) - \nabla G(x', \omega), x - x'\rangle \geq \|\nabla G(x, \omega) - \nabla G(x', \omega)\|_*^2, \quad \forall x, x' \in \mathcal{X}, \tag{6.24}$$

so that

$$\begin{aligned}\|\nabla G(x, \omega)\|_*^2 &\leq 2\|\nabla G(x, \omega) - \nabla G(x_*, \omega)\|_*^2 + 2\|\nabla G(x_*, \omega)\|_*^2 \\ &\leq 2\nu\langle\nabla G(x, \omega) - \nabla G(x_*, \omega), x - x_*\rangle + 2\|\nabla G(x_*, \omega)\|_*^2 \\ &= 2\nu\langle\nabla G(x, \omega), x - x_*\rangle - 2\nu\langle\nabla G(x_*, \omega), x - x_*\rangle + 2\|\nabla G(x_*, \omega)\|_*^2 \end{aligned}$$

so that

$$\begin{aligned}&\gamma_{i-1}\langle H_i, x_{i-1} - x_*\rangle + \gamma_i\langle\eta_i, , x_i - x_*\rangle \\ &\leq V(x_{i-1}, x_*) - V(x_i, x_*) + \gamma_{i-1}^2[\nu\langle H_i, x_{i-1} - x_*\rangle - \nu\zeta_i + \tau_i] \end{aligned}$$

where $\zeta_i = \langle \nabla G(x_*, \omega_i), x_{i-1} - x_* \rangle$ and $\tau_i = \|\nabla G(x_*, \omega)\|_*^2$. As a result, by convexity of h we have for $\gamma_i \le (4\nu)^{-1}$

$$\begin{aligned}
&\tfrac{3}{4}\gamma_{i-1}\langle \nabla g(x_{i-1}), x_{i-1} - x_* \rangle + \gamma_i [h(x_i) - h(x_*)] \\
&\le (\gamma_{i-1} - \gamma_{i-1}^2 \nu)\langle \nabla g(x_{i-1}), x_{i-1} - x_* \rangle + \gamma_i \langle \eta_i, x_i - x_* \rangle \\
&\le V(x_{i-1}, x_*) - V(x_i, x_*) + (\gamma_{i-1} - \gamma_{i-1}^2 \nu)\langle \xi_i, x_{i-1} - x_* \rangle + \gamma_{i-1}^2 [\tau_i - \nu \zeta_i]
\end{aligned}$$

where we put $\xi_i = H_i - \nabla g(x_{i-1})$. When summing from $i = 1$ to m we obtain

$$\begin{aligned}
&\sum_{i=1}^m \gamma_{i-1} \Big(\tfrac{3}{4}\langle \nabla g(x_{i-1}), x_{i-1} - x_* \rangle + [h(x_{i-1}) - h(x_*)] \Big) \\
&\qquad \le V(x_0, x_*) + \underbrace{\sum_{i=1}^m [\gamma_{i-1}^2 (\tau_i - \nu \zeta_i) + \gamma_{i-1}(1 - \gamma_{i-1}\nu)\langle \xi_i, x_{i-1} - x_* \rangle]}_{=: R_m} \\
&\qquad\qquad + \gamma_0 [h(x_0) - h(x_*)] - \gamma_m [h(x_m) - h(x_*)]. \qquad (6.25)
\end{aligned}$$

2^o. We have

$$\begin{aligned}
\gamma_{i-1}\langle \xi_i, x_{i-1} - x_* \rangle &= \gamma_{i-1} \overbrace{\langle [\nabla G(x_{i-1}, \omega_i) - \nabla G(x_*, \omega_i)] - \nabla g(x_{i-1}), x_{i-1} - x_* \rangle}^{\upsilon_i} \\
&\quad + \gamma_{i-1}\langle \nabla G(x_*, \omega_i), x_{i-1} - x_* \rangle \\
&= \gamma_{i-1}[\upsilon_i + \zeta_i],
\end{aligned}$$

so that

$$R_m = \sum_{i=1}^m \gamma_{i-1}^2 \tau_i + \sum_{i=1}^m (\gamma_{i-1} - \gamma_{i-1}^2 \nu)\upsilon_i + \sum_{i=1}^m (\gamma_{i-1} - 2\nu\gamma_{i-1}^2)\zeta_i =: r_m^{(1)} + r_m^{(2)} + r_m^{(3)}. \qquad (6.26)$$

Note that $r_m^{(3)}$ is a sub-Gaussian martingale. Indeed, one has $\mathbf{E}_{i-1}\{\zeta_i\} = 0$ a.s.,[4] and

$$|\zeta_i| \le \|x_{i-1} - x_*\| \, \|\nabla G(x_*, \omega)\|_*,$$

so that by the sub-Gaussian hypothesis (6.6), $\mathbf{E}_{i-1}\Big\{ \exp\Big(\underbrace{\frac{\zeta_i^2}{4R^2\sigma_*^2}}_{\nu_*^2} \Big) \Big\} \le \exp(1)$. As a result (cf. the proof of Proposition 4.2 in [165]),

$$\forall t \quad \mathbf{E}_{i-1}\Big\{ e^{t\zeta_i} \Big\} \le \exp\big(t\mathbf{E}_{i-1}\{\zeta_i\} + \tfrac{3}{4}t^2\nu_*^2 \big) = \exp\big(3t^2 R^2 \sigma_*^2 \big),$$

and applying (6.31a) to $S_m = r_m^{(3)}$ with

$$r_m = 6R^2\sigma_*^2 \sum_{i=0}^{m-1} (\gamma_i - 2\nu\gamma_i^2)^2 \le 6R^2\sigma_*^2 \sum_{i=0}^{m-1} \gamma_i^2$$

[4] We use notation $\mathbf{E}_{i-1}$ for the conditional expectation given $x_0, \omega_1, ..., \omega_{i-1}$.

we conclude that for some $\Omega_m^{(3)}$ such that $\mathrm{Prob}(\Omega_m^{(3)}) \geq 1 - e^{-t}$ and all $\omega^m \in \Omega_m^{(3)}$

$$r_m^{(3)} \leq 2\sqrt{3tR^2\sigma_*^2 \sum_{i=0}^{m-1} \gamma_i^2} \leq 3tR^2 + 3\sigma_*^2 \sum_{i=0}^{m-1} \gamma_i^2. \tag{6.27}$$

Next, again by (6.6), due to the Jensen inequality, $\mathbf{E}_{i-1}\{\tau_i\} \leq \sigma_*^2$, and

$$\mathbf{E}_{i-1}\left\{\exp\left(t\|\nabla G(x_*, \omega_i)\|_*\right)\right\} \leq \exp\left(t\mathbf{E}_{i-1}\{\|\nabla G(x_*, \omega_i)\|_*\} + \tfrac{3}{4}t^2\sigma_*^2\right) \leq \exp\left(t\sigma_* + \tfrac{3}{4}t^2\sigma_*^2\right).$$

Thus, when setting

$$\mu_i = \gamma_{i-1}\sigma_*, \quad s_i^2 = \tfrac{3}{2}\gamma_{i-1}\sigma_*^2, \quad \bar{s} = \max_i \gamma_i s_i,$$

$M_m = r_m^{(1)}$, $v_m + h_m = \frac{21}{4}\sigma_*^4 \sum_{i=0}^{m-1} \gamma_i^4$, and applying the bound (6.31b) of Lemma 9 we obtain

$$r_m^{(1)} \leq 3\sigma_*^2 \sum_{i=0}^{m-1} \gamma_i^2 + \underbrace{\sqrt{21t\sigma_*^4 \sum_{i=0}^{m-1} \gamma_i^4}}_{=:\Delta_m^{(1)}} + 3t\overline{\gamma}^2\sigma_*^2$$

for $\overline{\gamma} = \max_i \gamma_i$ and $\omega^m \in \Omega_m^{(1)}$ where $\Omega_m^{(1)}$ is of probability at least $1 - e^{-x}$. Because

$$\overline{\gamma}^2 \sum_{i=0}^{m-1} \gamma_i^2 \geq \sum_{i=0}^{m-1} \gamma_i^4,$$

whenever $\sqrt{21t\sigma_*^4 \sum_{i=0}^{m-1} \gamma_i^4} \geq \sum_{i=0}^{m-1} \gamma_i^2$, one has $21t\overline{\gamma}^2 \geq \sum_{i=0}^{m-1} \gamma_i^2$ and

$$21t \sum_{i=0}^{m-1} \gamma_i^4 \leq 21t\overline{\gamma}^2 \sum_{i=0}^{m-1} \gamma_i^2 \leq (21t\overline{\gamma}^2)^2$$

Thus,

$$\Delta_m^{(1)} \leq \min\left[21t\sigma_*^2\overline{\gamma}^2, \sigma_*^2 \sum_{i=0}^{m-1} \gamma_i^2\right] \leq 21t\sigma_*^2\overline{\gamma}^2 + \sigma_*^2 \sum_{i=0}^{m-1} \gamma_i^2,$$

and

$$r_m^{(1)} \leq \sigma_*^2\left[4\sum_{i=0}^{m-1} \gamma_i^2 + 24t\overline{\gamma}^2\right] \tag{6.28}$$

for $\omega^m \in \Omega_m^{(1)}$.
Finally, by the Lipschitz continuity of ∇G (cf. (6.24)), when taking expectation w.r.t. the distribution of ω_i, we get

$$\begin{aligned}\mathbf{E}_{i-1}\{v_i^2\} &\leq 4R^2\mathbf{E}_{i-1}\{\|\nabla G(x_{i-1}, \omega_i) - \nabla G(x_*, \omega_i)\|_*^2\} \\ &\leq 4R^2\nu\mathbf{E}_{i-1}\{\langle \nabla G(x_{i-1}, \omega_i) - \nabla G(x_*, \omega_i), x_{i-1} - x_*\rangle\} = 4R^2\nu\langle \nabla g(x_{i-1}), x_{i-1} - x_*\rangle.\end{aligned}$$

On the other hand, one also has $|\upsilon_i| \le 2\nu\|x_{i-1}-x_i\|^2 \le 8\nu R^2$. We can now apply Lemma 10 with $\sigma_i^2 = 4\gamma_{i-1}^2 R^2\nu\langle\nabla g(x_{i-1}), x_{i-1}-x_*\rangle$ to conclude that for $t \ge 2\sqrt{2+\ln m}$

$$r_m^{(2)} \le 4\underbrace{\sqrt{tR^2\nu\sum_{i=0}^{m-1}\gamma_i^2\langle\nabla g(x_i), x_i-x_*\rangle}}_{=:\Delta_m^{(2)}} + 16t\nu R^2\overline{\gamma}$$

for all $\omega^m \in \Omega_m^{(2)}$ such that $\mathrm{Prob}(\Omega_m^{(2)}) \ge 1-2e^{-t}$. Note that

$$\Delta_m^{(2)} \le 2tR^2 + \tfrac{1}{4}\nu\sum_{i=0}^{m-1}\gamma_i^2\langle\nabla g(x_i), x_i-x_*\rangle,$$

and $\gamma_i \le (4\nu)^{-1}$, so that

$$r_m^{(2)} \le \nu\sum_{i=0}^{m-1}\gamma_i^2\langle\nabla g(x_i), x_i-x_*\rangle + 12tR^2 \le \tfrac{1}{4}\sum_{i=0}^{m-1}\gamma_i\langle\nabla g(x_i), x_i-x_*\rangle + 12tR^2 \tag{6.29}$$

for $\omega^m \in \Omega_m^{(2)}$.

3^o. When substituting bounds (6.27)–(6.29) into (6.26) we obtain

$$\begin{aligned} R_m &\le \tfrac{1}{4}\textstyle\sum_{i=0}^{m-1}\gamma_i\langle\nabla g(x_i), x_i-x_*\rangle + 12tR^2 + \sigma_*^2\left[4\sum_{i=0}^{m-1}\gamma_i^2 + 24t\overline{\gamma}^2\right] + 2\sqrt{3tR^2\sigma_*^2\textstyle\sum_{i=0}^{m-1}\gamma_i^2} \\ &\le \tfrac{1}{4}\textstyle\sum_{i=0}^{m-1}\gamma_i\langle\nabla g(x_i), x_i-x_*\rangle + 15tR^2 + \sigma_*^2\left[7\sum_{i=0}^{m-1}\gamma_i^2 + 24t\overline{\gamma}^2\right] \end{aligned}$$

for all $\omega^m \in \overline{\Omega}_m = \bigcap_{i=1}^3\Omega_m^{(i)}$ with $\mathrm{Prob}(\overline{\Omega}_m) \ge 1-4e^{-t}$ and $t \ge 2\sqrt{2+\ln m}$. When substituting the latter bound into (6.25) and utilizing the convexity of g and h we arrive at

$$\begin{aligned} &\left(\sum_{i=0}^{m-1}\gamma_i\right)\Big(\tfrac{1}{2}[g(\widehat{x}_m)-g(x_*)] + [h(\widehat{x}_m)-h(x_*)]\Big) \le \sum_{i=0}^{m-1}\gamma_i\Big(\tfrac{1}{2}[g(x_i)-g(x_*)] + [h(x_i)-h(x_*)]\Big) \\ &\le \sum_{i=1}^{m}\gamma_{i-1}\Big(\tfrac{1}{2}\langle\nabla g(x_{i-1}), x_{i-1}-x_*\rangle + [h(x_{i-1})-h(x_*)]\Big) \\ &\le V(x_0,x_*) + 15tR^2 + \sigma_*^2\left[7\sum_{i=0}^{m-1}\gamma_i^2 + 24t\overline{\gamma}^2\right] + \gamma_0[h(x_0)-h(x_*)] - \gamma_m[h(x_m)-h(x_*)]. \end{aligned}$$

In particular, for constant stepsizes $\gamma_i \equiv \gamma$ we get

$$\begin{aligned} &\tfrac{1}{2}[g(\widehat{x}_m)-g(x_*)] + [h(\widehat{x}_m)-h(x_*)] \\ &\quad\le \frac{V(x_0,x_*)+15tR^2}{\gamma m} + \frac{h(x_0)-h(x_m)}{m} + \gamma\sigma_*^2\left(7+\frac{24t}{m}\right). \end{aligned}$$

This implies the first statement of the proposition.

5^o. To prove the bound for the minibatch solution $\widehat{x}_m^{(L)} = \left(\sum_{i=0}^{m-1}\gamma_i\right)^{-1}\sum_{i=0}^{m-1}\gamma_i x_i^{(L)}$, it suffices to note that minibatch gradient observation $H(x,\omega^{(L)})$ is Lipschitz-continuous with Lipschitz constant ν, and that $H(x_*,\omega^{(L)})$ is sub-Gaussian with parameter σ_*^2 replaced with $\overline{\sigma}_{*,L}^2 \lesssim \frac{\Theta\sigma_*^2}{L}$, see Lemma 11. □

6.4.2 Deviation inequalities

Let us assume that $(\xi_i,\mathcal{F}_i)_{i=1,2,...}$ is a sequence of sub-Gaussian random variables satisfying[5]

$$\mathbf{E}_{i-1}\left\{e^{t\xi_i}\right\} \le e^{t\mu_i+\frac{t^2s_i^2}{2}}, \quad a.s. \tag{6.30}$$

for some *nonrandom* μ_i, s_i, $s_i \le \overline{s}$. We denote by $S_n = \sum_{i=1}^n \xi_i - \mu_i$, $r_n = \sum_{i=1}^n s_i^2$, $v_n = \sum_{i=1}^n s_i^4$, $M_n = \sum_{i=1}^n \xi_i^2 - (s_i^2+\mu_i^2)$, and $h_n = \sum_{i=1}^n 2\mu_i^2 s_i^2$. The following well known result is provided for reader's convenience.

Lemma 9 *For all $x>0$ one has*

$$\text{Prob}\left\{S_n \ge \sqrt{2xr_n}\right\} \le e^{-x}, \tag{6.31a}$$

$$\text{Prob}\left\{M_n \ge 2\sqrt{x(v_n+h_n)} + 2x\overline{s}^2\right\} \le e^{-x}. \tag{6.31b}$$

Proof. The inequality (6.31a) is straightforward. To prove (6.31b), note that for $t < \frac{1}{2}\overline{s}^{-2}$ and $\eta\sim\mathcal{N}(0,1)$ independent of $\xi_0,...,\xi_n$, we have:

$$\begin{aligned}\mathbf{E}_{i-1}\left\{e^{t\xi_i^2}\right\} &= \mathbf{E}_{i-1}\left\{\mathbf{E}_\eta\left\{e^{\sqrt{2t}\xi_i\eta}\right\}\right\} = \mathbf{E}_\eta\left\{\mathbf{E}_{i-1}\left\{e^{\sqrt{2t}\xi_i\eta}\right\}\right\}\\ &\le \mathbf{E}_\eta\left\{\exp\left\{\sqrt{2t}\eta\mu_i + t\eta^2 s_i^2\right\}\right\} = (1-2ts_i^2)^{-1/2}\exp\left\{\frac{t\mu_i^2}{1-2ts_i^2}\right\} \text{ a.s.},\end{aligned}$$

and because, cf [65, Lemma 1],

$$-\tfrac{1}{2}\ln(1-2ts_i^2) + \frac{t\mu_i^2}{1-2ts_i^2} - t(s_i^2+\mu_i^2) \le \frac{t^2s_i^2(s_i^2+2\mu_i^2)}{1-2ts_i^2} \le \frac{t^2s_i^2(s_i^2+2\mu_i^2)}{1-2t\overline{s}^2},$$

one has for $t < \frac{1}{2}\overline{s}^{-2}$

$$\mathbf{E}\left\{e^{tM_n}\right\} \le \exp\left\{\frac{t^2(v_n+h_n)}{1-2t\overline{s}^2}\right\}.$$

By Lemma 8 of [166], this implies that

$$\text{Prob}\left\{M_n \ge 2\sqrt{x(v_n+h_n)} + 2x\overline{s}^2\right\} \le e^{-x}$$

for all $x>0$. □

Now, suppose that ζ_i, $i=1,2,...$ is a sequence of random variables satisfying

$$\mathbf{E}_{i-1}\{\zeta_i\} = \mu_i,\ \mathbf{E}_{i-1}\{\zeta_i^2\} \le \sigma_i^2,\ \ |\zeta_i| \le 1 \ \text{ a.s.} \tag{6.32}$$

Denote $M_n = \sum_{i=1}^n[\zeta_i - \mu_i]$ and $q_n = \sum_{i=1}^n \sigma_i^2$. Note that $q_n \le n$.

[5] Here, same as above, we denote $\mathbf{E}_{i-1}$ the expectation conditional to $\mathcal{F}_{i-1}$.

Lemma 10 *Let $x \geq 1$; one has*

$$\mathrm{Prob}\left\{M_n \geq \sqrt{2xq_n} + x\right\} \leq \left[e\left(2x \ln\left[\frac{9n}{2x}\right] + 1\right) + 1\right] e^{-x}.$$

In particular, for $x \geq 4\sqrt{2 + \ln n}$ one has

$$\mathrm{Prob}\left\{M_n \geq \sqrt{2xq_n} + x\right\} \leq 2e^{-x/2}.$$

Proof. In the premise of the lemma, applying Bernstein's inequality for martingales [162, 167] we obtain for all $x > 0$ and $u > 0$,

$$\mathrm{Prob}\left\{M_n \geq \sqrt{2xu} + \frac{x}{3},\, q_n \leq u\right\} \leq e^{-x}.$$

We conclude that

$$\mathrm{Prob}\left\{M_n \geq x,\, q_n \leq \frac{2x}{9}\right\} \leq e^{-x},$$

and for any $u > 0$

$$\mathrm{Prob}\left\{M_n \geq \sqrt{2(x+1)q_n} + \frac{x}{3},\, u \leq q_n \leq u\big(1 + 1/x\big)\right\} \leq e^{-x},$$

so that

$$\delta_n(x;u) := \mathrm{Prob}\left\{M_n \geq \sqrt{2xq_n} + \frac{x}{3},\, u \leq q_n \leq u\big(1 + 1/x\big)\right\} \leq e^{-x+1}.$$

Let now $u_0 = 2x/9$, $u_j = \min\{n, (1+1/x)^j u_0\}$, $j = 0, ..., J$, with

$$J = \left\lfloor \ln\left[n/u_0\right] \ln^{-1}[1 + 1/x] \right\rfloor.$$

Note that $\ln[1 + 1/x] \geq 1/(2x)$ for $x \geq 1$, so that

$$J \leq \ln\left[n/u_0\right] \ln^{-1}[1 + 1/x] + 1 \leq 2x \ln\left[n/u_0\right] + 1.$$

On the other hand,

$$\begin{array}{rcl} \mathrm{Prob}\left\{M_n \geq \sqrt{2xq_n} + x\right\} & \leq & e^{-x} + \sum_{j=1}^{J} \delta_n(x;u_j) \leq e^{-x} + Je^{-x+1} \\ & \leq & \left[e\Big(2x \ln\left[\frac{9n}{2x}\right] + 1\Big) + 1\right] e^{-x} \end{array}$$

Finally, we verify explicitly that for $x \geq 4\sqrt{2 + \ln n}$ one has

$$\left[e\Big(2x \ln\left[\frac{9n}{2x}\right] + 1\Big) + 1\right] e^{-x/2} \leq 2,$$

implying that for such x

$$\mathrm{Prob}\left\{M_n \geq \sqrt{2xq_n} + x\right\} \leq 2e^{-x/2}. \qquad \square$$

Let $(\xi_i)_{i=1,...}$ be a sequence of independent random vectors in $\mathbf{R}^n$ such that

$$\mathbf{E}_{i-1}\left\{\exp\left(\frac{\|\xi_i\|_*^2}{s^2}\right)\right\} \leq \exp(1),$$

and let $\eta = \sum_{i=1}^m \xi_i$, $m \in \mathbf{Z}_+$. We are interested in "sub-Gaussian characteristics" of r.v. $\zeta = \langle u, \eta\rangle$ for some $u \in \mathbf{R}^n$, $\|u\| \le R$, and of $\tau = \|\eta\|_*$.
Because $\mathbf{E}\{\langle u, \xi_i\rangle\} = 0$ and $|\langle u, \xi_i\rangle| \le \|u\|\,\|\xi_i\|_*$, for all t one has (cf.,e.g., Proposition 4.2 of [165])

$$\mathbf{E}\Big\{e^{t\langle u,\eta\rangle}\Big\} = \prod_{i=1}^m \mathbf{E}\Big\{e^{t\langle u,\xi_i\rangle}\Big\} \le \prod_{i=1}^m \exp\left(\tfrac{3}{4}t^2s^2\right) = \exp\left(\tfrac{3}{4}mt^2s^2\right).$$

Let ξ_ℓ, $\ell = 1, 2, ...$ be a sequence of independent random vectors $\xi_\ell \in E$, such that $\mathbf{E}\{\xi_\ell\} = 0$ and $\mathbf{E}\Big\{e^{\|\xi_\ell\|_*^2/s^2}\Big\} \le \exp(1)$. Denote $\eta_j = \sum_{\ell=1}^j \xi_\ell$. We have the following result.

Lemma 11

$$\forall L \in \mathbf{Z}_+, w \quad \mathbf{E}\left\{\exp\left(\frac{\|\eta_L\|_*^2}{18\Theta s^2 L}\right)\right\} \le \exp(1) \tag{6.33}$$

where $\Theta = \max_{\|z\|\le 1}\theta(z)$ *for the d.-g.f.* θ *of the unit ball of norm* $\|\cdot\|$ *in* E, *as defined in Section 6.2.2.*

Proof. Let for $\eta \in E$, $\pi(\eta) = \sup_{\|z\|\le 1}[\langle \eta, z\rangle - \theta(z)]$. Observe that for all $\beta > 0$,

$$\|\eta_L\|_* = \sup_{\|z\|\le 1}\langle \eta_L, z\rangle \le \max_{\|z\|\le 1} \beta\theta(z) + \beta\pi(\eta_L/\beta) \le \beta\Theta + \beta\pi\big(\frac{\eta_L}{\beta}\big). \tag{6.34}$$

On the other hand, we know (cf. [59, Lemma 1]) that π is smooth with $\|\nabla\pi\| \le 1$, and $\nabla\pi$ is Lipschitz-continuous w.r.t. to $\|\cdot\|_*$, i.e.,

$$\|\nabla\pi(z) - \nabla\pi(z')\| \le \|z - z'\|_* \quad \forall z, z' \in E.$$

As a consequence of Lipschitz continuity of π, when denoting $\pi_\beta(\eta) = \beta\pi\big(\frac{\eta}{\beta}\big)$, we have

$$\pi_\beta(\eta_{j-1} + \xi_j) - \pi_\beta(\eta_{j-1}) \le \|\xi_j\|_*,$$

so that $\mathbf{E}\left\{\exp\left([\pi_\beta(\eta_j) - \pi_\beta(\eta_{j-1})]^2/s^2\right)\right\} \le \exp(1)$. Furthermore,

$$\pi_\beta(\eta_{j-1} + \xi_j) \le \pi_\beta(\eta_{j-1}) + \langle \nabla\pi_\beta(\eta_{j-1}), \xi_j/\beta\rangle + \|\xi_j\|_*^2/\beta,$$

and, because η_{j-1} does not depend on ξ_j and $\mathbf{E}\{\|\xi_j\|_*^2\} \le s^2$, we get

$$\mathbf{E}_{j-1}\{\pi_\beta(\eta_j) - \pi_\beta(\eta_{j-1})\} \le s^2/\beta.$$

By [165, Proposition 4.2] we conclude that random variables $\delta_j = \pi_\beta(\eta_j) - \pi_\beta(\eta_{j-1})$ satisfy for all $t \ge 0$,

$$\mathbf{E}_{j-1}\left\{e^{t\delta_j}\right\} \le \exp\left(ts^2\beta^{-1} + \tfrac{3}{4}t^2s^2\right).$$

Consequently,

$$\mathbf{E}\left\{e^{t\pi_\beta(\eta_L)}\right\} \le \mathbf{E}\left\{e^{t\pi_\beta(\eta_{L-1})}\right\}\exp\left(ts^2\beta^{-1} + \tfrac{3}{4}t^2s^2\right) \le \exp\left(ts^2L\beta^{-1} + \tfrac{3}{4}t^2s^2L\right).$$

When substituting the latter bound into (6.34), we obtain for $\beta^2 = s^2L/\Theta$

$$\mathbf{E}\left\{e^{t\|\eta_L\|_*}\right\} \le \exp\left(2ts\sqrt{\Theta L} + \tfrac{3}{4}t^2s^2L\right) \quad \forall t \ge 0. \tag{6.35}$$

To complete the proof of the lemma, it remains to show that (6.35) implies (6.33). This is straightforward. Indeed, for $\chi\sim\mathcal{N}(0,1)$, $\alpha>0$ and $\zeta=\|\eta_L\|_*$ one has

$$\begin{aligned}\mathbf{E}\left\{e^{\alpha\zeta^2}\right\}&=\mathbf{E}\left\{\mathbf{E}_\eta\left(e^{\sqrt{2\alpha}\zeta\chi}\right)\right\}=\mathbf{E}_\chi\left\{\mathbf{E}\left\{e^{\sqrt{2\alpha}\zeta\chi}\right\}\right\}\\&\le\mathbf{E}_\chi\left\{\exp\left(2\sqrt{2\alpha\Theta L}\,s\chi+\tfrac{3}{2}\alpha Ns^2\chi^2\right)\right\}=(1-3\alpha Ls^2)^{-1/2}\exp\left\{\frac{4\alpha\Theta Ls^2}{1-3\alpha Ls^2}\right\}\end{aligned}$$

When setting $\alpha=(18\Theta s^2L)^{-1}$, we conclude that

$$\mathbf{E}\left\{e^{\alpha\zeta^2}\right\}\le\exp(1)$$

due to $\Theta\ge1/2$. □

6.4.3 Proof of Theorem 6.2.1

We start with analysing the behaviour of the approximate solution $\widehat{x}^k_{m_0}$ at the stages of the preliminary phase of the procedure.

Lemma 12 *Let $m_0=\lceil64\delta^2\rho\nu s(4\Theta+60t)\rceil$ (here $\lceil a\rceil$ stands for the smallest integer greater or equal to a), $\gamma=(4\nu)^{-1}$, and let t satisfy $t\ge4\sqrt{2+\log(m_0)}$.*

Suppose that $R\ge2\delta\sigma_*\sqrt{6\rho s/\nu}$, that initial condition x_0 of Algorithms 1 and 2 satisfies $\|x_0-x_*\|\le R$, and that at the stage k of the preliminary phase we choose

$$\kappa_k=R_{k-1}\sqrt{\frac{\nu(4\Theta+60t)}{\rho sm_0}}\tag{6.36}$$

where $(R_k)_{k\ge0}$ is defined recursively:

$$R_{k+1}=\tfrac{1}{2}R_k+\frac{16\sigma_*^2\delta^2\rho s}{\nu R_k},\quad R_0=R.$$

Then the approximate solution $\widehat{x}^k_{m_0}$ at the end of the kth stage of the CSMD-SR algorithm satisfies, with probability $\ge1-4ke^{-t}$

$$\|\widehat{x}^k_{m_0}-x_*\|\quad\le R_k\le2^{-k}R+4\sigma_*\delta\sqrt{2\rho s/\nu}.\tag{6.37}$$

In particular, the estimate $\widehat{x}^{\overline{K}_1}_{m_0}$ after $\overline{K}_1=\left\lceil\frac{1}{2}\log_2\left(\frac{R^2\nu}{32\sigma_*^2\delta^2\rho s}\right)\right\rceil$ stages satisfies with probaility at least $1-4\overline{K}_1e^{-t}$

$$\|\overline{x}^{\overline{K}_1}_{m_0}-x_*\|\le8\sigma_*\delta\sqrt{2\rho s/\nu}.\tag{6.38}$$

Proof of the lemma.
1^o. Note that initial point x_0 satisfies $x_0\in X_R(x_*)$. Suppose that the initial point $x_0^k=\widehat{x}^{k-1}_{m_0}$ of the kth stage of the method satisfy $x_0^k\in X_{R_{k-1}}(x_*)$ with probability $1-4(k-1)e^{-t}$. In other words, there is a set $\mathcal{B}_{k-1}\subset\Omega$, $\mathrm{Prob}(\mathcal{B}_{k-1})\ge1-4(k-1)e^{-t}$, such that for all $\overline{\omega}^{k-1}=[\omega_1;...;\omega_{m_0(k-1)}]\subset\mathcal{B}_{k-1}$ one has $x_0^k\in X_{R_{k-1}}(x_*)$. Let us show that upon termination of the $k$the stage $\widehat{x}^k_{m_0}$ satisfy $\|x^k_{m_0}-x_*\|\le R_k$ with probability $1-4ke^{-t}$. By Proposition 6.4.1 (with $h(x)=\kappa_k\|x\|$) we conclude

that for some $\overline{\Omega}_k \subset \Omega$, $\mathrm{Prob}(\overline{\Omega}_k) \geq 1 - 4e^{-t}$, solution $\widehat{x}^k_{m_0}$ after m_0 iterations of the stage satisfies, for all for all $\omega^k = [\omega_{(k-1)m_0+1}, ..., \omega_{km_0}] \in \overline{\Omega}_k$,

$$F(\widehat{x}^k_{m_0}) - F(x_*) \leq \tfrac{1}{m_0}\left(\nu R^2_{k-1}(4\Theta + 60t) + \kappa_k R_{k-1}\right) + \frac{\sigma^2_*}{\nu}\left(\tfrac{7}{4} + \tfrac{6t}{m_0}\right).$$

When using the relationship (6.14) of Assumption [RSC] we now get

$$\|\widehat{x}^k_{m_0} - x_*\| \leq \delta\left[\rho s\kappa_k + \frac{R_{k-1}}{m_0} + \frac{\nu R^2_{k-1}}{\kappa_k m_0}(4\Theta + 60t) + \frac{\sigma^2_*}{\nu\kappa_k}\left(\tfrac{7}{4} + \tfrac{6t}{m_0}\right)\right]. \tag{6.39}$$

Note that κ_k as defined in (6.36) satisfies $\kappa_k \leq R_{k-1}(8\delta\rho s)^{-1}$, while $\kappa_k m_0 \geq 8\delta(4\Theta + 60t)R_{k-1}\nu$. Because $m_0 \geq 3840t$ due to $\rho\nu \geq 1$ and $\delta \geq 1$, one also has $\left(\frac{7}{4} + \frac{6t}{m_0}\right)\kappa_k^{-1} < 16\delta\rho s/R_{k-1}$. When substituting the above bounds into (6.39) we obtain

$$\|\widehat{x}^k_{m_0} - x_*\| \leq \delta R_{k-1}\left(\tfrac{1}{4\delta} + \tfrac{1}{m_0}\right) + \frac{16\delta^2\rho s\sigma^2_*}{R_{k-1}\nu} \leq \tfrac{1}{2}R_{k-1} + \frac{16\delta^2\rho s\sigma^2_*}{R_{k-1}\nu} = R_k. \tag{6.40}$$

We conclude that $\widehat{x}^k_{m_0} \in X_{R_k}(x_*)$ for all $\overline{\omega}^k \in \mathcal{B}_k = \mathcal{B}_{k-1} \cap \overline{\Omega}_k$, and

$$\mathrm{Prob}(\mathcal{B}_k) \geq \mathrm{Prob}(\mathcal{B}_{k-1}) - \mathrm{Prob}(\overline{\Omega}^c_k) \geq 1 - 4ke^{-t}.$$

2^o. Let now $a = 16\delta^2\rho s\sigma^2_*/\nu$, and let us study the behaviour of the sequence

$$R_k = \frac{R_{k-1}}{2} + \frac{a}{R_{k-1}} =: f(R_{k-1}), \quad R_0 = R \geq \sqrt{2a}.$$

Function f admits a fixed point at $R = \sqrt{2a}$ which is also the minimum of f, so one has $R_k \geq \sqrt{2a}$ $\forall k$. Thus,

$$d_k := R_k - \sqrt{2a} = \frac{R_{k-1} - \sqrt{2a}}{2} + \frac{2a - \sqrt{2a}R_{k-1}}{2R_{k-1}} \leq \tfrac{1}{2}d_{k-1} \leq 2^{-k}d_0 \leq 2^{-k}(R - \sqrt{2a}).$$

We deduce that $R_k \leq 2^{-k}R_0 + \sqrt{2a}$ which is (6.37). Finally, after running $\overline{K}_1$ stages of the preliminary phase, the estimate $\widehat{x}^{\overline{K}_1}_{m_0}$ satisfies

$$\|\widehat{x}^{\overline{K}_1}_{m_0} - x_*\| \leq 8\delta\sigma_*\sqrt{2\rho s/\nu}. \qquad \square$$

We turn next to the analysis of the asymptotic phase of Algorithm 2. We assume that the preliminary phase of the algorithm has been completed.

Lemma 13 *Let t be such that $t \geq 4\sqrt{2 + \log(m_1)}$, with $m_1 = \lceil 81\delta^2\rho s\nu(4\Theta + 60t)\rceil$, $\gamma = (4\nu)^{-1}$, and let $\ell_k = \lceil 10 \times 4^{k-1}\Theta\rceil$. We set*

$$\kappa_k = r_{k-1}\sqrt{\frac{\nu(4\Theta + 60t)}{\rho s m_1}}, \quad r_k = 2^{-k}r_0, \quad r_0 = 8\delta\sigma_*\sqrt{2\rho s/\nu}.$$

Then the approximate solution by Algorithm 2 $\widehat{x}^k_{m_1}$ at the end of the kth stage of the asymptotic phase satisfies, with probability $\geq 1 - 4(\overline{K}_1 + k)e^{-t}$, $\|\widehat{x}^k_{m_1} - x_*\| \leq r_k$, implying that

$$\|\widehat{x}^k_{m_1} - x_*\| \lesssim \delta^2\sigma_*\rho s\sqrt{\frac{\Theta\,(\Theta + t)}{N_k}}, \tag{6.41}$$

where $N_k = m_1\sum_{i=1}^k \ell_i$ is the total count of oracle calls for k asymptotic stages.

Proof of the lemma. Upon terminating the preliminary phase, the initial condition $x_0 = \widehat{x}_{m_0}^{\overline{K}_1}$ of the asymptotic phase satisfies (6.38) with probability greater or equal to $1 - 4\overline{K}_1 e^{-t}$. We are about to show that $\forall k \geq 1$, with probability at least $1 - 4(\overline{K}_1 + k)e^{-t}$,

$$\|\widehat{x}_{m_1}^k - x_*\| \leq r_k = 2^{-k} r_0, \quad r_0 = 8\delta\sigma_*\sqrt{2\rho s/\nu}.$$

The claim is obviously true for $k = 0$. Let let us suppose that it holds at stage $k - 1 \geq 0$, and let us prove that it also holds at stage k. To this end, we reproduce the argument used in the proof of Lemma 12, while taking into account that now ℓ_k observations are averaged at each iteration of the CSMD algorithm. Recall (cf. Lemma 11) that this amounts to replacing sub-Gaussian parameter σ_*^2 with $\overline{\sigma}_*^2 = 18\Theta\sigma_*^2/\ell_k$. When applying the result of Proposition 6.4.1 and the bound of (6.14) we conclude (cf. (6.39)) that, with probability $1 - (\overline{K}_1 + k)e^{-t}$,

$$\|\widehat{x}_{m_1}^k - x_*\| \leq \delta\left[\rho s\kappa_k + \frac{r_{k-1}}{m_1} + \frac{\nu r_{k-1}^2}{\kappa_k m_1}(4\Theta + 60t) + \frac{18\Theta\sigma_*^2}{\nu\kappa_k\ell_k}\left(\tfrac{7}{4} + \tfrac{6t}{m_1}\right)\right]$$

By simple algebra, we obtain the following analogue of (6.40):

$$\|\widehat{x}_{m_1}^k - x_*\| < \delta r_{k-1}\left(\tfrac{2}{9\delta} + \tfrac{1}{m_1}\right) + 10\frac{4^{-k+1}\delta^2\rho s\sigma_*^2}{r_{k-1}\nu} < \tfrac{r_{k-1}}{4} + \tfrac{r_{k-1}}{4} = r_k.$$

Observe that upon the end of the kth stage we used $N_k = m_1\sum_{i=1}^k \ell_k < 3m_1\Theta\sum_{j=1}^k 4^{j-1} \leq 4^k\Theta m_1$ observations of the asymptotic stage. As a consequence, $4^{-k} < \Theta m_1/N_k$ and

$$r_k = 2^{-k}r_0 \lesssim \delta^2\sigma_*\sqrt{\frac{\Theta(\Theta + t)s\nu\rho}{N_k}}.$$ □

Assuming that the preliminary phase of Algorithm 1 was completed, we now consider the asymptotic phase of the algorithm.

Lemma 14 *Let $t \geq 4\sqrt{2 + \log m_k}$, $m_k = \lceil 4^{k+4}(4\Theta + 60t)\delta^2\rho s\nu\rceil$,*

$$\gamma^k = \frac{r_{k-1}}{2\sigma_*}\sqrt{\frac{(4\Theta + 60t)}{2m_k}}, \quad \kappa_k^2 = \frac{5\sigma_* r_{k-1}}{\rho s}\sqrt{\frac{(4\Theta + 60t)}{m_k}} \tag{6.42}$$

where

$$r_k := 2^{-k}r_0, \quad r_0 = 8\delta\sigma_*\sqrt{2\rho s/\nu}.$$

Then the approximate solution $\widehat{x}_{m_k}^k$ upon termination of the kth asymptotic stage satisfies with probability $\geq 1 - 4(\overline{K}_1 + k)e^{-t}$

$$\|\widehat{x}_{m_k}^k - x_*\| \leq 2^{-k}r_0 \lesssim \quad 2^{-k}\sigma_*\delta\sqrt{\rho s\nu^{-1}} \lesssim \quad \delta^2\sigma_*\rho s\sqrt{\tfrac{\Theta + t}{N_k}} \tag{6.43}$$

where $N_k = \sum_{j=1}^k m_j$ is the total iteration count of k stages of the asymptotic phase.

Proof of the lemma.

We are to show that $\forall k \geq 0$, $\|\widehat{x}_{m_k}^k - x_*\| \leq r_k$ with probability $\geq 1 - 4(\overline{K}_1 + k)e^{-t}$ is true. By Lemma 12, the claim is true for $k = 0$ (at the start of the asymptotic phase, the initial condition $x_0 = \widehat{x}_{m_0}^{\overline{K}_1}$ satisfies the bound (6.38)). We now assume it to hold for $k - 1 \geq 0$, our objective is to

implement the recursive step $k-1 \to k$ of the proof. First, observe that the choice of γ^k in (6.42) satisfies $\gamma^k \le (4\nu)^{-1}$, $k = 1, ...$, so that Proposition 6.4.1 can be applied. From the result of the proposition and bound (6.14) we conclude (cf. (6.39)) that it holds, with probability $1-(\overline{K}_1+k)e^{-t}$,

$$\|\widehat{x}^k_{m_k} - x_*\| \le \delta\left[\rho s\kappa_k + \frac{r_{k-1}}{m_k} + \frac{r^2_{k-1}\,(4\Theta+60t)}{\gamma^k\kappa_k m_k} + 8\frac{\gamma^k\sigma_*^2}{\kappa_k}\right]$$

When substituting the value of γ^k from (6.42) we obtain

$$\|\widehat{x}^k_{m_k} - x_*\| \le \delta\left[\rho s\kappa_k + \frac{r_{k-1}}{m_k} + \frac{4\sigma_* r_{k-1}}{\kappa_k}\sqrt{\frac{2(4\Theta+60t)}{m_k}}\right],$$

which, by the choice of κ_k in (6.42), results in

$$\|\widehat{x}^k_{m_k} - x_*\|^2 \le 2\delta^2\left[10\rho s\sigma_* r_{k-1}\sqrt{\frac{4\Theta+60t}{m_k}} + \frac{r^2_{k-1}}{m_k^2}\right] \le \frac{r^2_{k-1}}{4} = r_k^2.$$

It remains to note that the total number $N_k = \sum_{j=1}^k m_j$ of iterations during k stages of the asymptotic phase satisfies $N_k \lesssim 4^k(\Theta+t)\delta^2\rho s\nu$, and $2^{-k} \lesssim \delta\sqrt{\frac{(\Theta+t)\rho s\nu}{N_k}}$, which along with definition of r_0 implies (6.43). □

Proof of Theorem 6.2.1. We can now terminate the proof of the theorem. Let us prove the accuracy bound of the theorem for the minibatch variant of the procedure.

Assume that the "total observation budget" N is such that only the preliminary phase of the procedure is implemented. This is the case when either $m_0\overline{K}_1 \ge N$, or $m_0\overline{K}_1 < N$ and $m_0\overline{K}_1 + m_1\ell_1 > N$. The output $\widehat{x}_N$ of the algorithm is then the last update of the preliminary phase, and by Lemma 12 it satisfies $\|\widehat{x}_N - x_*\| \le R2^{-k}$ where k is the count of completed stages. In the case of $m_0\overline{K}_1 \ge N$ this clearly implies that (recall that $N \ge m_0$) that $k \ge cN/m_0$ and, with probability $\ge 1-4ke^{-t}$

$$\|\widehat{x}_N - x_*\| \lesssim R\exp\left\{-\frac{c'N}{\delta^2\rho s\nu(\Theta+t)}\right\}. \tag{6.44}$$

On the other hand, when $m_0\overline{K}_1 < N < m_0\overline{K}_1 + m_1\ell_1$, by definition of m_1 and ℓ_1, one has $N \le Cm_0\overline{K}_1$, so that bound (6.44) still holds in this case.

Now, consider the case where at least one asymptotic stage has been completed. When $m_0\overline{K}_1 > \frac{N}{2}$ we still have $N \le Cm_0\overline{K}_1$, so that the bound (6.44) holds for the approximate solution $\widehat{x}^{(b)}_N$ at the end of the asymptotic stage. Otherwise, the number of oracle calls N_k of asymptotic stages satisfies $N_k \ge N/2$, and by (6.41) this implies that with probability $\ge 1-4(\overline{K}_1+\overline{K}_2)e^{-t}$,

$$\|\widehat{x}^{(b)}_N - x_*\| \lesssim \delta^2\sigma_*\rho s\sqrt{\frac{\Theta(\Theta+t)}{N}}.$$

To summarize, in both cases, the bound of Theorem 6.2.1 holds with probability at least $1-4(\overline{K}_1+\overline{K}_2)e^{-t}$.

The proof of the accuracy bound for the "standard" solution $\widehat{x}_N$ is completely analogous, making use of the bound (6.43) of Lemma 14 instead of (6.41). □

Remark 6.4.1 *Theorem 6.2.1 as stated in Section 6.2.3 does not say anything about convergence of $g(\widehat{x}_N)$ to $g(x_*)$. Such information can be easily extracted from the proof of the theorem. Indeed, observe that at the end of a stage of the method, one has, with probability $1-Cke^{-t}$,*

$$F_{\kappa_k}(\widehat{x}^k) - F_{\kappa_k}(x_*) \le \upsilon_k,$$

or

$$g(\widehat{x}^k) - g(x_*) \le \upsilon_k + \kappa_k(\|\widehat{x}^k\| - \|x_*\|) \le \upsilon_k + \kappa_k\|\widehat{x}^k - x_*\|$$

where $\widehat{x}^k$ is the approximate solution at the end of the stage k. One the other hand, at the end of the kth stage of the preliminary *phase one has $\|\widehat{x}^k - x_*\| \le R_k \le 2^{-k}R$, with $\kappa_k \lesssim R_k(\delta\rho s)^{-1} \le 2^{-k}R(\delta\rho s)^{-1}$ and $\upsilon_k \lesssim \frac{4^{-k}R^2}{\delta^2\rho s}$ implying that*

$$g(\widehat{x}^k) - g(x_*) \lesssim \upsilon_k + \frac{R_k^2}{\delta^2\rho s} \lesssim (\delta^{-2}+\delta^{-1})\frac{R^2}{\rho s}\exp\left\{-\frac{c}{\delta\rho\nu}\frac{N}{s(\Theta+t)}\right\}$$

where N is the current iteration count. Furthermore, at the end of the kth asymptotic *stage, one has, with probability $1-(\overline{K}_1+k)e^{-t}$, $\|\widehat{x}^k - x_*\| \le R_k \lesssim \delta^2\sigma_*\rho s\sqrt{\frac{\Theta+t}{m_k}}$, while $\kappa_k \asymp 2^{-k}\delta\sigma_*(\rho\nu s)^{-1/2} \lesssim \delta\sigma_*\sqrt{\frac{\Theta+t}{m_k}}$, and $\upsilon_k \lesssim \delta^2\sigma_*^2\rho s(\Theta+t)/m_k$. As a result, the corresponding $\widehat{x}^k$ satisfies*

$$g(\widehat{x}^k) - g(x_*) \le \upsilon_k + \kappa_k\|\widehat{x}^k - x_*\| \lesssim (\delta^2+\delta^3)\rho\sigma_*^2 s\frac{\Theta+t}{m_k}.$$

When putting the above bounds together, assuming that at least 1 stage of the algorithm was completed, we arrive at the bound after N steps:

$$g(\widehat{x}_N) - g(x_*) \lesssim (\delta^{-2}+\delta^{-1})\frac{R^2}{\rho s}\exp\left\{-\frac{c}{\delta^2\rho\nu}\frac{N}{s(\Theta+t)}\right\} + (\delta^2+\delta^3)\rho s\sigma_*^2\frac{\Theta+t}{N} \tag{6.45}$$

with probability $1-(\overline{K}_1+\overline{K}_2)e^{-t}$.

6.4.4 Proof of Proposition 6.3.1

1^o. Recall that $\mathfrak{r}$ is $\overline{r}$-Lipschitz continuous, i.e., for all $t, t' \in \mathbf{R}^m$

$$|\mathfrak{r}(t) - \mathfrak{r}(t')| \le \overline{r}|t - t'|.$$

As a result, for all $x, x' \in X$,

$$\|\phi[\mathfrak{r}(\phi_i^T x) - \mathfrak{r}(\phi_i^T x')]\|_\infty \le \overline{r}\|\phi_i\|_\infty|\phi_i^T(x-x')| \le \overline{r}\|\phi_i\|_\infty^2\|x-x'\|_1 \le \overline{r\nu}^2\|x-x'\|_1,$$

so that $\nabla G(x,\omega) = \phi[\mathfrak{r}(\phi^T x) - \eta]$ is Lipschitz continuous w.r.t. ℓ_1-norm with Lipschitz constant $\mathcal{L}(\omega) \le \overline{r\nu}^2$.

2^o. Due to strong monotonicity of $\mathfrak{r}$,

$$\begin{array}{rcl} g(x) - g(x_*) & = & \int_0^1 \nabla g(x_* + t(x-x_*))^T(x-x_*)dt \\ & = & \int_0^1 \mathbf{E}\Big\{\phi[\mathfrak{r}(\phi^T(x_*+t(x-x_*)) - \mathfrak{r}(\phi^T x_*)]\Big\}^T(x-x_*)dt \\ & \ge & \int_0^1 \underline{r}\mathbf{E}\big\{(\phi^T(x-x_*))^2\big\}tdt = \frac{1}{2}\underline{r}\|x-x_*\|_\Sigma^2, \end{array}$$

what is (6.16).

3^o. The sub-Gaussianity in the "batchless" case is readily given by $\nabla G(x_*, \omega_i) = \sigma\phi_i\xi_i$ with $\|\phi_i\xi_i\|_\infty \le \|\phi_i\|_\infty|\xi_i| \le \overline{\nu}\|\xi_i\|_2$ and

$$\mathbf{E}\left\{\exp\left(\frac{\|\nabla G(x_*,\omega_i)\|_\infty^2}{\sigma^2\overline{\nu}^2}\right)\right\} \le e$$

due to $\mathbf{E}\{e^{\xi_i^2}\} \le \exp(1)$. Because Θ variation of the d.-g.f. θ, as defined in (6.20), is bounded with $C\ln n$, by Lemma 11 we conclude that batch observation

$$H\left(x_*, \omega_i^{(L)}\right) = \frac{1}{L}\sum_{\ell=1}^{L}\nabla G(x_*,\omega_i^\ell) = \frac{1}{L}\sum_{\ell=1}^{L}\sigma\phi_i^\ell,\xi_i^\ell$$

is sub-Gaussian with parameter $\lesssim \sigma^2\overline{\nu}^2\ln n$.

4^o. In the situation of Section 6.3.1, Σ is positive definite, $\Sigma \succeq \kappa_\Sigma I$, $\kappa_\Sigma > 0$, and condition $\mathbf{Q}(\lambda,\psi)$ is satisfied with $\lambda = \kappa_\Sigma$ and $\psi = 1$. Because quadratic minoration condition (6.17) for g is verified with $\mu \ge \underline{r}$ due to (6.16), when applying the result of Lemma 8, we conclude that Assumption [RSC] holds with $\delta = 1$ and $\rho = (\kappa_\Sigma\underline{r})^{-1}$.[6] □

6.5 Properties of sparsity structures

6.5.1 Sparsity structures

The scope of results of Section 6.2 is much broader than "vanilla" sparsity optimization. We discuss here general notion of *sparsity structure* which provides a proper application framework for these results.

In what follows we assume to be given a *sparsity structure* [168] on E—a family $\mathcal{P}$ of projector mappings $P = P^2$ on E such that

A1.1 every $P \in \mathcal{P}$ is assigned a linear map $\overline{P}$ on E such that $P\overline{P} = 0$ and a nonnegative weight $\pi(P)$;

A1.2 whenever $P \in \mathcal{P}$ and $f, g \in E$ such that $\|f\|_* \le 1$, $\|g\|_* \le 1$,

$$\|P^*f + \overline{P}^*g\|_* \le 1$$

where for a linear map $Q : E \to F$, $Q^* : F \to E$ is the conjugate mapping.

Following [168], we refer to a collection of the just introduced entities and *sparsity structure on* E. For a nonnegative real s we set

$$\mathcal{P}_s = \{P \in \mathcal{P} : \pi(P) \le s\}.$$

Given $s \ge 0$ we call $x \in E$ *s-sparse* if there exists $P \in \mathcal{P}_s$ such that $Px = x$. Typically, one is interested in the following "standard examples":

1. "Vanilla (usual)" sparsity: in this case $E = \mathbf{R}^n$ with the standard inner product, $\mathcal{P}$ is comprised of projectors on all coordinate subspaces of $\mathbf{R}^n$, $\pi(P) = \mathrm{Rank}(P)$, and $\|\cdot\| = \|\cdot\|_1$.

[6] We refer to Section 6.5.2 and Lemma 15 for the proof of Lemma 8.

2. Group sparsity: $E = \mathbf{R}^n$, and we partition the set $\{1, ..., n\}$ of indices into K nonoverlapping subsets $I_1, ..., I_K$, so that to every $x \in \mathbf{R}^n$ we associate blocks x^k with corresponding indices in I_k, $k = 1, ..., K$. Now $\mathcal{P}$ is comprised of projectors $P = P_I$ onto subspaces $E_I = \{[x^1, ..., x^K] \in \mathbf{R}^n : x^k = 0 \, \forall k \notin I\}$ associated with subsets I of the index set $\{1, ..., K\}$. We set $\pi(P_I) = \mathrm{card}(I)$, and define $\|x\| = \sum_{k=1}^K \|x_k\|_2$—*block* ℓ_1/ℓ_2*-norm.*

3. Low rank structure: in this example $E = \mathbf{R}^{p\times q}$ with, for the sake of definiteness, $p \geq q$, and the Frobenius inner product. Here $\mathcal{P}$ is the set of mappings $P(x) = P_\ell x P_r$ where P_ℓ and P_r are, respectively, $q \times q$ and $p \times p$ orthoprojectors, $\overline{P}(x) = (I - P_\ell)x(I - P_r)$, and $\|\cdot\|$ is the nuclear norm $\|x\| = \sum_{i=1}^q \sigma_i(x)$ where $\sigma_1(x) \geq \sigma_2(x) \geq ... \geq \sigma_q(x)$ are singular values of x, $\|\cdot\|_*$ is the spectral norm, so that $\|x\|_* = \sigma_1(x)$, and $\pi(P) = \max[\mathrm{Rank}(P_\ell), \mathrm{Rank}(P_r)]$.

   In this case, for $\|f\|_* \leq 1$ and $\|g\|_* \leq 1$ one has

$$\|P^*(f)\|_* = \|P_\ell f P_r\|_* \leq 1, \quad \|\overline{P}^*(g)\|_* = \|(I - P_\ell)g(I - P_r)\|_* \leq 1,$$

   and because the images and orthogonal complements to the kernels of P and $\overline{P}$ are orthogonal to each other, $\|P^*(f) + \overline{P}^*(g)\|_* \leq 1$.

6.5.2 Condition $\mathbf{Q}(\lambda, \psi)$

We say that a positive semidefinite mapping $\Sigma : E \to E$ satisfies condition $\mathbf{Q}(\lambda, \psi)$ for given $s \in \mathbf{Z}_+$ if for some $\psi, \lambda > 0$ and all $P \in \mathcal{P}_s$ and $z \in E$

$$\|Pz\| \leq \sqrt{s/\lambda}\|z\|_\Sigma + \|\overline{P}z\| - \psi\|z\|. \tag{6.46}$$

Lemma 15 *Suppose that x_* is an optimal solution to (6.5) such that for some $P \in \mathcal{P}_s$, $\|(I - P)x_*\| \leq \delta$, and that condition $\mathbf{Q}(\lambda, \psi)$ is satisfied. Furthermore, assume that objective g of (6.5) satisfies the following minoration condition*

$$g(x) - g(x_*) \geq \mu\big(\|x - x_*\|_\Sigma\big)$$

where $\mu(\cdot)$ is monotone increasing and convex. Then a feasible solution $\widehat{x} \in \mathcal{X}$ to (6.7) such that

$$\mathrm{Prob}\left\{F_\kappa(\widehat{x}) - F_k(x_*) \leq \upsilon\right\} \geq 1 - \epsilon.$$

satisfies, with probability at least $1 - \epsilon$,

$$\|\widehat{x} - x_*\| \leq \frac{\mu^*\left(\kappa\sqrt{s/\lambda}\right) + \upsilon}{\kappa\psi} + \frac{2\delta}{\psi} \tag{6.47}$$

where $\mu^* : \mathbf{R}_+ \to \mathbf{R}_+$ is conjugate to $\mu(\cdot)$, $\mu^*(t) = \sup_{u\geq 0}[tu - \mu(u)]$.

Proof. When setting $z = \widehat{x} - x_*$ one has

$$\begin{aligned}\|\widehat{x}\| = \|x_* + z\| &= \|Px_* + (I - P)x_* + z\| \geq \|Px_* + z\| - \|(I - P)x_*\| \\ &\geq \|Px_*\| + \|\overline{P}z\| - \|Pz\| - \delta\end{aligned}$$

where we used the relation

$$\|Px_* + z\| \geq \|Px_*\| - \|Pz\| + \|\overline{P}z\|$$

(cf. Lemma 3.1 of [168] applied to $w = Px_*$). When using condition $\mathbf{Q}(\lambda, \psi)$ we obtain

$$\|\widehat{x}\| \geq \|Px_*\| - \sqrt{s/\lambda}\|z\|_\Sigma + \psi\|z\| - \delta,$$

so that $F_k(\widehat{x}) \leq F_k(x_*) + \upsilon$ implies

$$\begin{aligned}\kappa\left(\|Px_*\| + \psi\|z\| - \delta\right) &\leq \tfrac{1}{2}[g(x_*) - g(\widehat{x})] + \kappa\sqrt{s/\lambda}\|z\|_\Sigma + \kappa\|x_*\| + \upsilon \\ &\leq -\tfrac{1}{2}\mu(\|z\|_\Sigma) + \kappa\sqrt{s/\lambda}\|z\|_\Sigma + \kappa\|x_*\| + \upsilon \\ &\leq \tfrac{1}{2}\mu^*(2\kappa\sqrt{s/\lambda}) + \kappa\|x_*\| + \upsilon,\end{aligned}$$

and we conclude that

$$\kappa\psi\|z\| \leq \tfrac{1}{2}\mu^*(2\kappa\sqrt{s/\lambda}) + 2\kappa\delta + \upsilon$$

due to $\|x_*\| - \|Px_*\| \leq \|(I - P)x_*\| \leq \delta$. □

Note that when $\mu(u) = \frac{\mu}{2}u^2$, one has $\mu^*(t) = \frac{1}{2\mu}t^2$, and in the case of $\|\cdot\| = \|\cdot\|_1$, with probability $1 - \epsilon$,

$$\|\widehat{x} - x_*\|_1 \leq \frac{s\kappa}{\mu\lambda\psi} + \frac{\upsilon}{\kappa\psi} + \frac{2\delta}{\psi}.$$

This, in particular, implies bound (6.18) of Lemma 8.

Remark 6.5.1 *We discuss implications of condition $\mathbf{Q}(\lambda, \psi)$ and result of Lemma 15 for "usual" sparsity in Section 6.3 of the paper. Now, let us consider the case of the low rank sparsity. Let $z \in \mathbf{R}^{p\times q}$ with $p \geq q$ for the sake of definiteness. In this case, $\|\cdot\|$ is the nuclear norm, and we put $P(z) = P_\ell z P_r$ where P_ℓ and P_r are orthoprojectors of rank $s \leq q$ such that $\|(I - P)(x)\| = \|x_* - P_\ell x_* P_r\| \leq \delta$.*[7]

Furthermore, for a $p \times q$ matrix z let us put

$$\sigma^{(k)}(z) = \sum_{i=1}^{k} \sigma_i(z),\ 1 \leq k \leq q.$$

With the sparsity parameter s being a nonnegative integer,

$$\forall(z \in \mathbf{R}^{p\times q}, P \in \mathcal{P}_s): \quad \|P(z)\| \leq \sigma^{(s)}(z), \quad \|\overline{P}(z)\| \geq \|z\| - \sigma^{(2s)}(z).$$[8]

and we conclude that in the present situation condition

$$\sigma^{(s)}(z) + \sigma^{(2s)}(z) \leq \sqrt{s/\lambda}\|z\|_\Sigma + (1 - \psi)\|z\| \tag{6.48}$$

is sufficient for the validity of $\mathbf{Q}(\lambda, \psi)$. As a result, condition (6.48) with $\psi > 0$ is sufficient for applicability of the bound of Lemma 15. It may also be compared to the necessary and sufficient condition of "s-goodness of Σ" in [169]:

$$\exists\psi > 0:\ 2\sigma^{(s)}(z) \leq (1 - \psi)\|z\| \ \ \forall z \in \mathrm{Ker}(\Sigma).$$

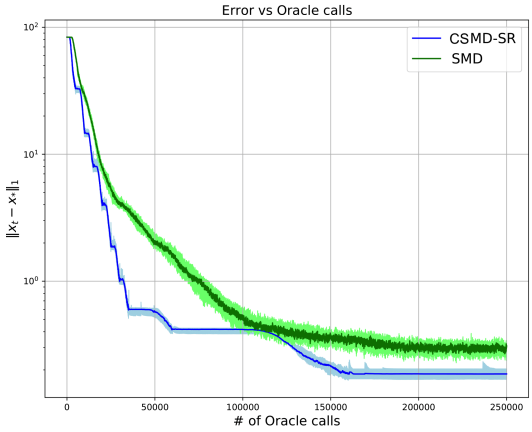
[7] E.g., choose P_ℓ and P_r as left and right projectors on the space generated by s principal left and right singular vectors of x_*, so that $\|x_* - P_\ell x_* P_r\| = \|(I - P_\ell)x_*(I - P_r)\| = \sum_{i=s+1}^{q}\sigma_i \leq \delta$.

[8] Indeed, let $P \in \mathcal{P}_s$, so that $\mathrm{Rank}(P_\ell) \leq s$ and $\mathrm{Rank}(P_r) \leq s$, and $\|P(z)\| = \|P_\ell z P_r\| \leq \sigma^{(s)}(z)$. Since the matrix $\overline{P}(z)$ differs from z by a matrix of rank at most $2s$, by the Singular Value Interlacing theorem we have $\sigma_i(\overline{P}(z)) \geq \sigma_{i+2s}(z)$, whence $\|\overline{P}(z)\| \geq \|z\| - \sigma^{(2s)}(z)$.

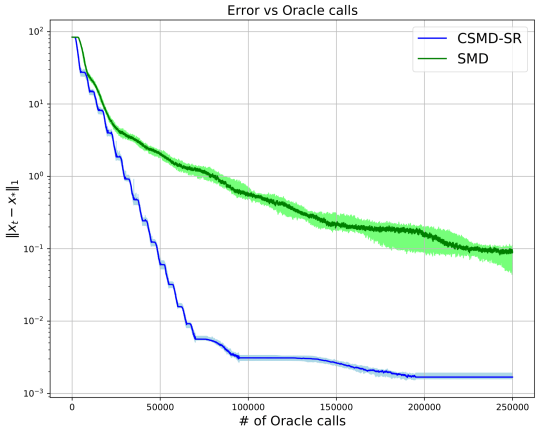
6.6 Supplementary numerical experiments

This section complements the numerical results appearing on the main body of the paper. We consider the setting in Section 6.3.3 of sparse recovery problem from GLR model observations (6.15). In the experiments below, we consider the choice (6.19) of activation function $\mathfrak{r}_\alpha(t)$ with values $\alpha = 1$ and $\alpha = 1/10$; value $\alpha = 1$ corresponds to linear regression with $\mathfrak{r}(t) = t$, whereas when $\alpha = 0.1$ activation have a flatter curve with rapidly decreasing with r modulus of strong convexity for $|t| \le r$. Same as before, in our experiments, the dimension of the parameter space is $n = 500\,000$, the sparsity level of the optimal point x_* is $s = 100$; we use the minibatch Algorithm 2 with the maximal number of oracle calls is $N = 250\,000$. In Figures 6.4 and 6.5 we report results for $\kappa_\Sigma \in \{0.1, 1\}$ and $\sigma \in \{0.001, 0.1\}$; the simulations are repeated 10 times, we trace the median of the estimation error $\|\widehat{x}_i - x_*\|_1$ along with its first and the last deciles against the number of oracle calls.

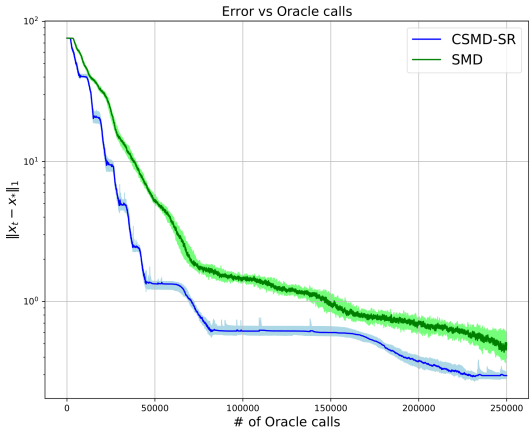
In our experiments, multistage algorithms exhibit linear convergence on initial iterations. Surprisingly, "standard" (non-Euclidean) SMD also converges fast in the "preliminary" regime. This may be explained by the fact that iteration x_i of the SMD obtained by the "usual" proximal mapping $\mathrm{Prox}(\gamma_{i-1}\nabla G(x_{i-1}, \omega_i), x_{i-1})$ is computed as a solution to the optimization problem with "penalty" $\theta(x) = c\|x\|_p^p$, $p = 1 + 1/\ln n$ which results in a "natural" sparsification of x_i. As iterations progress, such "sparsification" becomes insufficient, and the multistage routine eventually outperforms the SMD. Implementing the method for "flatter" nonlinear activation $\mathfrak{r}(t)$ or increased condition number of the regressor covariance matrix Σ requires increasing the length m_0 of the stage of the algorithm.



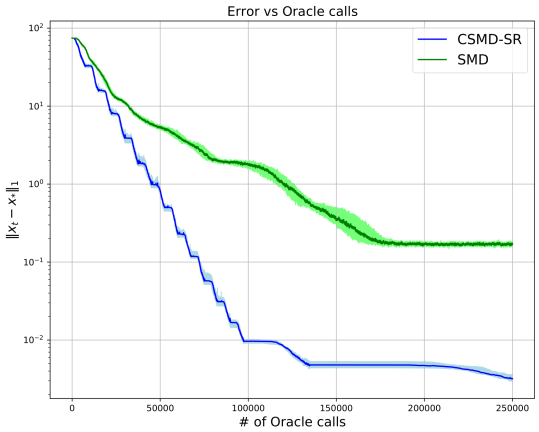
(a) $\kappa_\Sigma = 1, \sigma = 0.1, m_0 = 5000$



(b) $\kappa_\Sigma = 1, \sigma = 0.001, m_0 = 5000$

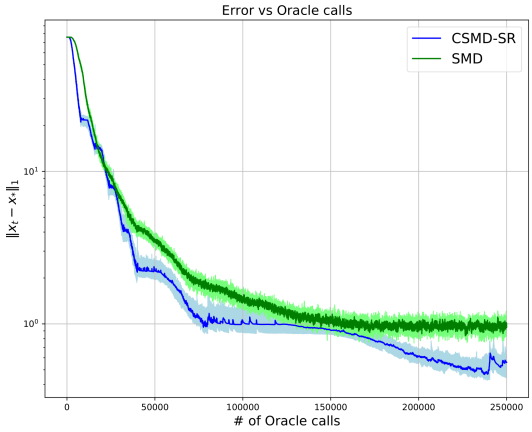


(c) $\kappa_\Sigma = 0.1, \sigma = 0.1, m_0 = 7500$

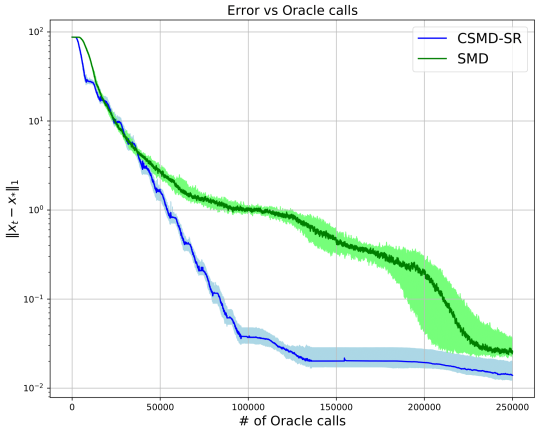


(d) $\kappa_\Sigma = 0.1, \sigma = 0.001, m_0 = 7500$

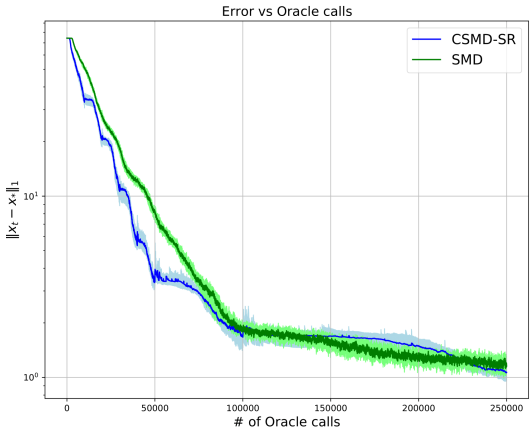
Figure 6.4: CSMD-SR and "vanilla" SMD in Linear Regression problem (activation function $\mathfrak{r}(t) = t$); ℓ_1 error as a function of the number of oracle calls



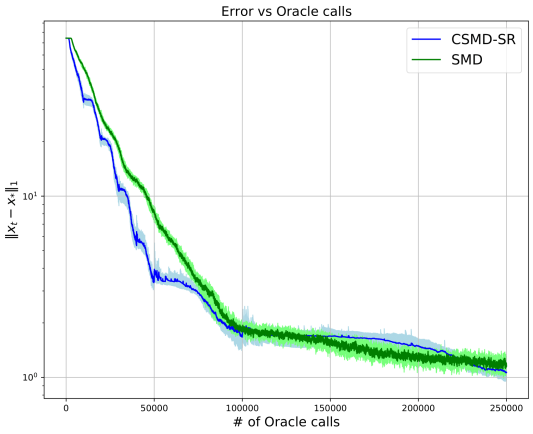
(a) $\kappa_\Sigma = 1, \sigma = 0.1, m_0 = 8000$



(b) $\kappa_\Sigma = 1, \sigma = 0.001, m_0 = 8000$



(c) $\kappa_\Sigma = 0.1, \sigma = 0.1, m_0 = 10\,000$



(d) $\kappa_\Sigma = 0.1, \sigma = 0.001, , m_0 = 10\,000$

Figure 6.5: CSMD-SR and "vanilla" SMD in Generalized Linear Regression problem: activation function $\mathfrak{r}_{1/10}(t)$; ℓ_1 error as a function of the number of oracle calls

Chapter 7

Extensions

7.1 Adaptive CSMD-SR via Lepski's Procedure

We present in this section an algorithm inspired by the CSMD-SR with hyper-parameters independent of problem parameters ρ and s, and thus adaptive to the latter. We call this new algorithm Ada-CSMD-SR.
More precisely, we are given N samples, a desired precision level $\epsilon \in (0,1)$ and a starting point x_0 with an initial prior R such that $\|x_0 - x_*\| \leq R$. We also assume that parameters $\sigma_*, \nu, \delta, \Theta$ are known, and that we have access to a stochastic approximation of objective function g's gradient, as in the previous chapter. Here, objective g satisfies the (RSC) assumption with *unknown* parameter ρ, and is minimized by x_* that is a sparse vector with *unknown* level of sparsity s. With this setting in place, we aim to produce an estimate $\widehat{x}^{(a)}$ with guarantees on the $1-\epsilon$ quantile of

$$\|\widehat{x}^{(a)} - x_*\|$$

that are almost the same as those we provide for the CSMD-SR estimate computable when knowing ρ and s.
In deterministic optimization, [170] proposes a first order algorithm adaptive to the smoothness of the objective, whereas in [152], the proposed multistage method is respectively adaptive to the uniform convexity-parameter. Authors in [152] also provide a stochastic variant of their algorithm, whereas [171] proposes a version of SGD adaptive to local strong convexity of the objective.
For the rest of this section, we will say that the pair (g, x_*) satisfies the (RSC) assumption with parameter ρs if (6.2.3) holds with parameters ρ and s.

7.1.1 Motivation

Observe that in our previous developments, one of the crucial hyper-parameter choice is the constant stage-length of the preliminary phase. In the situation where one knows $\nu, \Theta, \delta, \rho, s$, we advocate the choice

$$m = \lceil 64\delta^2 \nu \rho s(4\Theta + 60t) \rceil, \quad t \geq 4\sqrt{2 + \ln(m)}.$$

Observe that this hyper-parameter can replace the term ρs in other ones. Indeed, at the k-th stage of the preliminary phase, recall that

$$\kappa_k = R_{k-1}\sqrt{\frac{\nu(4\Theta + 60t)}{\rho s m}} \asymp R_{k-1}\frac{\delta\nu(\Theta + t)}{m},$$

$$R_k = \tfrac{1}{2}R_{k-1} + \frac{16\sigma_*^2\delta^2\rho s}{\nu R_{k-1}} \asymp \tfrac{1}{2}R_{k-1} + \left(\frac{\sigma_*}{\nu}\right)^2 \frac{m}{(\Theta+t)R_{k-1}}.$$

This illustrates that our algorithm's hyper-parameters can be formulated based on the stage-length m, rather than the parameters ρ and s. Hence, for any objective g and its minimizer x_*, an estimate that adapts to the former is also adaptive to both ρ and s. This leads us to focus on generating estimates adaptive to the stage-length m. It's worth mentioning that previous work in [172] introduces a deterministic multistage method that adjusts the length of each stage without knowing the sharpness parameters of the objective function. We begin by briefly analyzing a version of the CSMD-SR algorithm where hyper-parameters are modified to depend on m.

The stage-length dependent CSMD estimate.

Let integer $m \in [\lceil 4\delta, N]$, and define $\beta(m) := \frac{m}{(64\delta)^2\nu t}$, with

$$t := \max\left\{\Theta; 4\sqrt{2+\ln(m)}; t_\epsilon(m)\right\}, \tag{7.1}$$
$$t_\epsilon(m) := \ln\left(\frac{4}{\epsilon}\log_2\left(\left(\frac{2\nu R}{\sigma_*\sqrt{m}} \bigvee 1\right)\sqrt{1+\frac{8N}{3\Theta m}}\right)\right).$$

Observe that if (g, x_*) satisfies the RSC with $\rho s \le \beta(m)$, then

$$m \ge \lceil (64\delta)^2\nu t\rho s\rceil \ge \lceil 64\delta^2\nu\rho s(4\Theta + 60t)\rceil.$$

Thus, we say hat a specific stage-length m is adapted for (g, x_*) if the latter satisfies RSC with parameter $\beta(m)$. We now present the notations used to define a version of the CSMD-SR algorithm that depends on m. For $k \ge 0$, define the sequence of preliminary rates

$$R_k(m) := \frac{R_{k-1}(m)}{2} + \left(\frac{\sigma^*}{\nu}\right)^2 \frac{m}{128R_{k-1}(m)}, \quad R_0(m) = R, \tag{7.2}$$

and asymptotic rates

$$r_k(m) := \tfrac{\sigma^*}{\nu}\tfrac{\sqrt{m}}{2}2^{-k}. \tag{7.3}$$

Observe that for all $k \ge 1$, $R_k(m) \ge \frac{\sigma_*\sqrt{m}}{16\nu}$, implying in turn that $R_k(m) \le 2^{-k}R + \frac{\sigma_*\sqrt{m}}{4\nu}$. In particular,

$$R_{K_0(m)}(m) \le r_0(m), \text{ with } K_0(m) := \left(\left\lceil \log_2\left(\frac{4\nu R}{\sigma_*\sqrt{m}}\right)\right\rceil\right)_+.$$

With this definition, the total number of stages in the preliminary phase is

$$K_P(m) := \left\lfloor \frac{N}{m}\right\rfloor \bigwedge K_0(m).$$

For the asymptotic phase, we consider the mini-batch version of the CSMD-SR algorithm with batches of size $L_l = 4^{l-1}\lceil 9\Theta/8\rceil$ at its l-th stage. Therefore, in the asymptotic phase, one can compute a maximum of $K_A(m)$ stages, with the latter being defined as

$$K_A(m) := \left(\left\lfloor \frac{1}{2}\log_2\left(1+\frac{3}{\lceil 9\Theta/8\rceil}\left(\left\lfloor\frac{N}{m}\right\rfloor - K_P(m)\right)\right)\right\rfloor\right)_+ \tag{7.4}$$
$$:= \max\left\{k:\ k\ge 0,\ \sum_{l=0}^{k-1}\lceil 9\Theta/8\rceil 4^l \le \left(\left\lfloor\frac{N}{m}\right\rfloor - K_P(m)\right)\right\}.$$

Given these definitions, we define the total number of stages that can be completed with N samples as $K(m) := K_P(m) + K_A(m)$.
We now introduce the adaptive version of the CSMD-SR algorithm, beginning with the initial starting point $x_0 = \widehat{x}_0(m)$. For $k \in [1 : K(m)]$, if

- we are in the preliminary phase, i.e. $1 \leq k \leq K_0(m)$, we define the k-th stage's output be

$$\widehat{x}_k(m) = \text{CSMD}\left(\widehat{x}_{k-1}(m), \frac{1}{4\nu}, \kappa_k, R_{k-1}(m), m, 1\right), \tag{7.5}$$

  with $\kappa_k := 512\delta\nu t\frac{R_{k-1}(m)}{m}$.

- we are in the asymptotic phase, i.e. $k > K_0(m)$, we define the l-th stage's output

$$\widehat{x}_k(m) = \text{CSMD}\left(\widehat{x}_{k-1}(m), \frac{1}{4\nu}, \kappa_k, r_{l-1}(m), m, L_l\right), \tag{7.6}$$

  with $l = k - K_0(m)$ and $\kappa_k := 512\delta\nu t\frac{r_{l-1}(m)}{m}$.

Finally, we will denote the output of the final stage

$$\widehat{x}(m) := \widehat{x}_{K(m)}(m) \tag{7.7}$$

and refer to it as the CSMD-SR estimate associated with stage-length m. We also use the notation

$$R(m) := R_{K_P(m)}(m)\mathbf{1}_{\{K_A(m)=0\}} + r_{K_A(m)}(m)\mathbf{1}_{\{K_A(m)>0\}}, \tag{7.8}$$

for the high-probability upper bound on its error, as shown in the following

Proposition 7.1.1 *Let $m \in [\lceil 4\delta \rceil : N]$. Provided that (g, x_*) is such that $\rho s \leq \beta(m)$, the inequality*

$$\|\widehat{x}(m) - x_*\| \leq R(m) \tag{7.9}$$

holds with probability greater than $1 - \epsilon$.

Proof of 7.1.1:
We use the same arguments as for the proof of 12 and 13, adapted to the outlined choices of hyperparameters.
1°: Assume that we are at the $k+1$-th stage of the preliminary phase, starting with point $\widehat{x}_k(m)$ such that with probability greater than $1 - 4ke^{-t}$,

$$\|\widehat{x}_k(m) - x_*\| \leq R_k(m).$$

Inserting the values $\gamma = (4\nu)^{-1}$, $\kappa \equiv \kappa_{k+1}(m)$ into the equation (6.23) results in the upper bound

$$F_\kappa(\widehat{x}_{k+1}(m)) - F_\kappa(x_*) \leq \frac{64\nu t\left(R_k(m)\right)^2}{m}\left(1 + \frac{8\delta}{m}\right) + \frac{\sigma_*^2}{\nu}\left(\frac{7}{4} + \frac{6t}{m}\right).$$

which occurs on an event of probability greater than $1 - 4(k+1)e^{-t}$. Combining the latter with the RSC assumption being satisfied with $\rho s \leq \beta(m)$ yields

$$\|\widehat{x}_{k+1}(m) - x_*\| \leq R_k(m)\left(\frac{1}{4} + \frac{\delta}{m}\right) + \left(\frac{\sigma_*}{\nu}\right)^2\left(\frac{7}{4} + \frac{6t}{m}\right)\frac{m}{64t}\frac{1}{R_k(m)}.$$

Since $m \geq 4\delta \geq 4$ and (7.14) implies $t \geq 1$, one has that on the same event,

$$\|\widehat{x}_{k+1}(m) - x_*\| \leq \frac{R_{k+1}(m)}{2} + \left(\frac{\sigma_*}{\nu}\right)^2 \frac{m}{128 R_k(m)} = R_{k+1}(m).$$

From the results we've juste established, it follows through induction that at the conclusion of the preliminary phase, with a probability exceeding $1 - 4K_P(m)e^{-t}$

$$\|\widehat{x}_{K_P(m)}(m) - x_*\| \leq R_{K_P(m)}(m) \leq R 2^{-K_P(m)} + \frac{\sigma_* \sqrt{m}}{4\nu}, \tag{7.10}$$

holds true. Moreover, if $K_P(m) = K_0(m)$, the last inequality can be upper bounded by $\frac{\sigma_* \sqrt{m}}{2\nu} = r_0(m)$.

2°: Assume that $K_A(m) > 0$ and that we have already completed k stages, with $k \in [K_P(m); K(m) - 1]$. It follows that we are in the $l+1$-th stage of the asymptotic phase, with $l = k - K_0(m)$, with $\widehat{x}_k(m)$ as our starting point, such that

$$\text{Prob}\left[\|\widehat{x}_k(m) - x_*\| \leq r_l(m)\right] \geq 1 - 4ke^{-t}.$$

We use similar calculations and arguments as for the proof of 13. Recall that when using batches of size L, one can replace the sub-gaussianity parameter σ_*^2 by $18\Theta\sigma_*^2/L$. This yields that with probability greater than $1 - (k+1)e^{-t}$, one has that

$$\begin{aligned} \|\widehat{x}_{k+1}(m) - x_*\| &\leq \frac{r_l(m)}{2} + \left(\frac{\sigma_*}{\nu}\right)^2 \frac{m}{128 r_l(m)} \frac{18\Theta}{\lceil 9\Theta/8 \rceil 4^{l-1}} \\ &\leq r_l(m)\left(\frac{1}{2} + \frac{18\Theta}{36\Theta}\right) = r_{l+1}(m). \end{aligned} \tag{7.11}$$

3°: Setting $k = K(m)$ leads to

$$\text{Prob}\left[\|\widehat{x}_{K(m)}(m) - x_*\| \leq R(m)\right] \geq 1 - 4K(m)e^{-t}.$$

Noting that $K_0(m) \leq \log_2(\frac{2\nu R}{\sigma_* \sqrt{m}} \bigvee 1)$ and $K_A(m) \leq \log_2\left(\sqrt{1 + \frac{8N}{3\Theta m}}\right)$, we can derive

$$\begin{aligned} 4K(m)e^{-t} &\leq 4e^{-t}(K_0(m) + K_A(m)) \\ &\leq 4e^{-t} \log_2\left(\left(\frac{2\nu R}{\sigma_* \sqrt{m}} \bigvee 1\right)\sqrt{1 + \frac{8N}{3\Theta m}}\right) \leq \epsilon, \end{aligned} \tag{7.12}$$

where the final inequality is justified by definition (7.1). □

7.1.2 The adaptive CSMD-SR estimate

The adaptive estimate we propose is based on Lepski's [5] adaptive procedure. In our setting, the latter is applied to a collection of estimates $(\widehat{x}^{(i)})_{i=1}^I$ to select the best estimate in the context of (g, x_*). More formally, for an integer I, we assume that we are given the following grid of stage-lengths

$$\lceil 4\delta \rceil = m_1 < ... < m_I \leq N. \tag{7.13}$$

For all i, we define estimates $\widehat{x}^{(i)} := \widehat{x}(m_i)$, generated by the CSMD-SR algorithm presented above, with parameter

$$t_i := \max\left\{\Theta; 4\sqrt{2+\ln(m_i)}; t_\epsilon(m_i)\right\} + \ln(I) \tag{7.14}$$

instead of t. We also define the associated quantities $\beta_i := \frac{m_i}{(64\delta)^2 \nu t_i}$, $K^{(i)} := K(m_i)$, and $R^{(i)} := R(m_i)$.

Proposition 7.1.2 *Let collection of estimates $(\widehat{x}^{(i)})_{i=1}^I$ be as previously stated. The event "For all $i \in [1:I]$ such that (g, x_*) satisfies the RSC assumption with $\rho s \leq \beta_i$,*

$$\|\widehat{x}^{(i)} - x_*\| \leq R^{(i)}." \tag{7.15}$$

holds with probability greater than $1-\epsilon$.

Proof of 7.1.2:

1°: Using similar arguments as for the proof of (7.12), one has that

$$\begin{aligned} 4\sum_{i=1}^I K^{(i)} e^{-t_i} &\leq 4\sum_{i=1}^I \log_2\left(\left(\frac{2\nu R}{\sigma_*\sqrt{m_i}} \bigvee 1\right)\sqrt{1+\frac{8N}{3\Theta m_i}}\right) e^{-t_i} \\ &\leq \sum_{i=1}^I \frac{\epsilon}{I} = \epsilon. \end{aligned} \tag{7.16}$$

2°: We are now ready to prove (7.15). Let us call $\mathcal{E}_i$ the event
"If the RSC property is satisfied with parameters smaller than β_i, one has $\|\widehat{x}^{(i)} - x_*\| \leq R^{(i)}$."
Proposition 7.1.1 states that $\text{Prob}\left[\mathcal{E}_i\right] \geq 1 - 4K^{(i)}e^{-t_i}$. As the event we are interested in is $\mathcal{E} = \bigcap_{i=1}^I \mathcal{E}_i$, the result follows directly from the fact that

$$\text{Prob}\left[\mathcal{E}\right] \geq 1 - \text{Prob}\left[\bigcup_{i=1}^I \overline{\mathcal{E}_i}\right] \geq 1 - 4\sum_{i=1}^I K^{(i)} e^{-t_i} \geq 1-\epsilon,$$

where we have used (7.16) to prove the last inequality. □

Given the grid $(m_i)_{i=1}^I$, we now propose the following construction of our adaptive estimate

- For all $i \in [1:I]$, compute $\widehat{x}^{(i)}$, and set $\widehat{x}^{(I+1)} = x_0$, $R^{(I+1)} = R$.
- Define the set of admissible indexes

$$\mathcal{A} := \left\{i \in [1:I] : \ \forall j,\ i < j \leq 1+I,\ \|\widehat{x}^{(i)} - \widehat{x}^{(j)}\| \leq R^{(i)} + R^{(j)}\right\}, \tag{7.17}$$

  and let $\widehat{i} := \min \mathcal{A}$.
- Select $\widehat{x}^{(a)} := \widehat{x}^{(\widehat{i})}$ if $\mathcal{A}$ is non empty, and x_0 otherwise.

Prior to establishing guarantees on the error quantile of its estimation, we define the error of a CSMD-SR estimate that knows the value of ρs. With notation

$$m(t) := \lceil (64\delta)^2 \rho s \nu t \rceil, \tag{7.18}$$

we define parameters $t_* := \max\left\{\Theta; \bar{t}_*\right\}$ and $t_*^I := t_* + \ln(I)$, where

$$\bar{t}_* := \min\left\{t : t \geq \max\left\{4\sqrt{2+\ln(m(t))}; t_\epsilon(m(t))\right\}\right\}. \tag{7.19}$$

The next proposition requires that $R, \epsilon, \nu, \sigma_*, N, \delta$ are such that they verify the following assumptions.

Assumption 1:
For all m such that $\lceil 4\delta \rceil \leq m \leq N$, one has $R(m) < R(m+1) \leq 2R(m)$.

Assumption 2:
One can compute I and integers $(m_i)_{i=1}^{I+1}$ such that for all $i \in [1:I]$, $\lceil 4\delta \rceil \leq m_i < m_{i+1} \leq N$ and

$$2R^{(i)} \leq R^{(i+1)} \leq 4R^{(i)}, \tag{7.20}$$

and in particular

$$2R^{(I)} + R^{(I-1)} \leq R \leq 2R^{(I+1)} + R^{(I)}. \tag{7.21}$$

Assumption 3:
ρ is large enough so that $m(t_*) \geq \lceil 4\delta \rceil$.

Proposition 7.1.3 *With probability greater than $1-\epsilon$, one has*

$$\|\widehat{x}^{(a)} - x_*\| \leq 9R\left(m(t_*^I)\right) \bigwedge 3R \tag{7.22}$$

$$\leq 27R\left(m_*^I\right). \tag{7.23}$$

Remark 7.1.1 *Observe that assumptions 1 is not overly restrictive. Indeed, it can be computationally verified in a number of operation linear in N by simply calculating all the $R(m)$ for $m \in [\lceil 4\delta \rceil, N]$. Moreover, the second inequality in (7.20) being true essentially depends on ratio $\sigma_*/(\nu\sqrt{R})$. For instance, if one has*

$$\frac{\sigma_*^2 \lceil 4\delta \rceil}{\nu^2 R} \geq 1,$$

for all considered m, the resulting CSMD-SR algorithm will always be in the asymptotic phase, and one essentially needs to multiply m by 4 to perform $K_A(m)$ - 1 stages, which will result in multiplying the rate by two. On the other hand, when $\sigma_*/\nu \to 0$, our algorithm always stays in the preliminary phase, and in that case, we can not ensure the upper bound on $R(m+1)/R(m) \leq 2$ as $\lfloor N/(m+1) \rfloor - \lfloor N/m \rfloor$ can be greater than 1, and $R(m) \asymp R2^{-\lfloor N/m \rfloor}$.

Remark 7.1.2 *Note that if assumption 1 holds, assumption 2's fulfillement only depends on N being large enough. Indeed, starting with $m_1 = \lceil 4\delta \rceil$, one can sequentially choose m_{i+1} from the interval $[m_i + 1 : N]$ as the largest m satisfying (7.20), until condition (7.21) is achieved.*

Remark 7.1.3 *Note that if the last assumption does not hold, our procedure still yields the upper bound*

$$\|\widehat{x}^{(a)} - x_*\| \leq R(\lceil 4\delta \rceil),$$

corresponding to the smallest stage-length ensuring division of the error by 2 in the preliminary phase. Moreover, it is guaranteed to hold when $\rho \geq \frac{1}{4096\nu}$. In the situation of 6.3, the latter is true when

$$\kappa_\Sigma \leq \frac{4096\overline{r}}{\underline{r}}\overline{\nu}^2,$$

where κ_Σ is such that $\mathbf{E}\left[\phi\phi^T\right] \succeq \kappa_\Sigma I$.

Proof of 7.1.3:

1°: We first treat the case where it exists $\bar{i} \in [1:I]$ such that $\beta_{\bar{i}-1} < \rho s \leq \beta_{\bar{i}}$.
Using proposition 7.1.2, we have that on an event $\mathcal{E}$ of prability greater than $1-\epsilon$, the following is true:
For all $i \geq \bar{i}$ and $j > i$, one has that

$$\|\widehat{x}^{(i)} - \widehat{x}^{(j)}\| \leq \|\widehat{x}^{(i)} - x_*\| + \|\widehat{x}^{(j)} - x_*\| \leq R^{(i)} + R^{(j)}.$$

In particluar, this proves that $\bar{i} \in \mathcal{A}$, implying that on event $\mathcal{E}$, $\widehat{i} \leq \bar{i}$. We first treat the case where $\widehat{i} < \bar{i}$, where one has

$$\begin{aligned}
\|\widehat{x}^{(\widehat{i})} - x_*\| &\leq \|\widehat{x}^{(\widehat{i})} - \widehat{x}^{(\bar{i})}\| + \|\widehat{x}^{(\bar{i})} - x_*\| \\
&\leq 2R^{(\bar{i})} + R^{(\widehat{i})} \\
&\leq 2R^{(\bar{i})}(N) + \max_{l<\bar{i}} R^{(l)}(N) \qquad (7.24) \\
&\leq 2R^{(\bar{i})}(N) + R^{(\bar{i}-1)}(N), \qquad (7.25)
\end{aligned}$$

where last inequality is a direct consequence of sequence $\left(R^{(i)}\right)_{i=1}^{I}$ being increasing. Observe that in the case where $\widehat{i} = \bar{i}$, inequality (7.25) is still valid. By definition, one has that $m_{\bar{i}-1} < \rho s(64\delta)^2 \nu t_I \leq m_{\bar{i}}$. Using assumption 1, one also has

$$R^{(\bar{i}-1)} < R(m(t_*^I)) \leq R^{(\bar{i})},$$

which yields the desired result when combined with first part of assumption 2.
2°: If $\rho s \geq \beta_I$, then either x_0 is selected, or an index $j \in [1:I]$ is. If index j is selected, then

$$\|\widehat{x}^{(a)} - x_*\| \leq \|\widehat{x}^{(a)} - x_0\| + \|x_0 - x_*\| \leq R^{(j)} + 2R \leq 3R.$$

Observing that one has $2R^{(I+1)} + R^{(I)} \leq 9R^{(I)}$ implies that $R^{(I)} \geq R/9$, and that $\beta_I < \rho s$, one has

$$3R \leq 27R^{(I)} \leq 27R(m(t_*^I)),$$

which yields (7.23). □

Under the same assumptions, we state the main result of this section, an upper bound on the precision of our adaptive estimate.

Theorem 7.1.1 *One has with probability greater than* $1-\epsilon$ *that*

$$\|\widehat{x}^{(a)} - x_*\| \lesssim R\exp\left\{-\frac{cN}{\delta^2 \rho s \nu (t_* + \ln(I))}\right\} + \rho s \sigma_* \delta^2 \sqrt{\frac{\Theta(t_* + \ln(I))}{\nu N}}. \qquad (7.26)$$

Moreover, under assumption 2, the grid-size I *is such that*

$$I \leq \log_2\left(\frac{R}{R(\lceil 4\delta \rceil)}\right). \qquad (7.27)$$

Proof of Theorem 7.1.1: Using the same arguments as for the proof of 6.2.1, one has that (7.23) implies that with probability greater than $1-\epsilon$,

$$\|\widehat{x}^{(a)} - x_*\| \lesssim R \exp\left\{-\frac{cN}{\delta^2 \rho s \nu t_*^I}\right\} + \rho s \sigma_* \delta^2 \sqrt{\frac{\Theta t_*^I}{\nu N}}.$$

Moreover, observe that assumptions (7.21) and (7.20) implies that

$$R \geq 5R^{I-1} \geq \frac{5R_1}{4} 2^I \geq 2^I R_1,$$

which in turn implies (7.27) when noticing that $R_1 = R(\lceil 4\delta \rceil)$.

7.2 Analysis under Reduced Uniform Convexity hypothesis (RUC)

We present in this section an extension for the analysis of the *CSMD-SR* algorithm, introduced in the last chapter, when applying it to solve sparse recovery problem of the form (6.2). This section is justified by the introduction of a new condition on the regularity of the objective function g, that is linked to the notion of uniform convexity [152, 172–175]. We replace the quadratic lower bound assumption verified by the objective function g by a higher order polynomial lower bound. As discussed in the previous chapter, we will see that the $\mathbf{Q}(\lambda, \psi)$ condition and the new lower bound assumption can be integrated to establish a new assumption, similar to our approach with the RSC assumption. Before introducing this general assumption, we present for the sake of clarity a definition of uniform convexity for a differentiable function h.

Definition 21 *Let X be a convex closed subset of an Euclidean space E. A differentiable function $h : X \to \mathbf{R}$ is said to be (μ, p)-uniformly convex on X if there exists $p \geq 2$ and $\mu > 0$ such that $\forall x, y \in X$,*

$$h(x) - h(y) - \langle \nabla h(y), x - y \rangle \geq \frac{\mu}{p} \|x - y\|^p. \tag{7.28}$$

Note that this definition simply reduces to μ-strong convexity when h is $(\mu, 2)$-uniformly convex. Recall that in the previous chapter, the Reduced Strong Convexity (RSC) assumption was introduced to offer a comprehensive framework for analyzing sparse problems across various setups of sparsity structures. The RSC assumption takes it origins from Lemma 15 where the function $\mu(x) = \frac{\mu}{2}x^2$ is used to provide the quadratic lower bound on the suboptimality. This section is devoted to give the analysis of the CSMD-SR algorithm when the objective function g satisfies the minoration condition presented in Lemma 15 using $\mu(x) = \frac{\mu}{p}x^p$ with $p > 2$. This leads us to introduce the Reduced Uniform Convexity assumption (RUC).

Assumption [RUC] For a general norm $\|\cdot\|$ and two constants $p > 2$ and $q \in [1, 2)$ such that $1/p + 1/q = 1$, there exist $\delta \geq 1$, and problem dependent positive constants ν, ρ, α such that as long as feasible solution $\widehat{x} \in X$ to the composite problem ((6.7)) satisfies

$$\|\widehat{x} - x_*\| \leq R_{\mathrm{RUC}} := \Gamma_{p,n,\|\cdot\|} \cdot \left(\frac{\nu \alpha^p}{\rho^{p-1}}\right)^{\frac{1}{p-2}} \text{ and } (F_\kappa(\widehat{x}) - F_\kappa(x_*))_+ \leq \upsilon,$$

it holds that

$$\|\widehat{x} - x_*\| \leq \delta \kappa^{-1} \left[\rho s^{\frac{q}{2}} \kappa^q + \upsilon\right], \tag{7.29}$$

where $\Gamma_{p,n,\|\cdot\|}$ is a problem dependent constant.
In line with our approach in the previous chapter, where the CSMD-SR algorithm was formulated based on the RSC assumption, we will now leverage the RUC assumption in the following sections. This will allow us to explore new parameter choices aimed at adapting the multistage algorithm for this updated framework.

Remark 7.2.1 *Notice that when taking $p = 2$, we retrieve assumption RSC. The maximal radius R_{RUC} on which the assumption can hold goes to infinity, meaning that the condition holds on $\mathbf{R}^n$, while the final bound on $\|\widehat{x} - x_*\|$ remains unchanged. This is not surprising as assuming uniform convexity around the optimum tends to assuming strong convexity around the optimum when p goes to 2.*

7.2.1 An example motivating the RUC assumption

In this section we motivate the use of the RUC assumption by studying a theoretical example. We place ourselves in the "vanilla" sparsity setting that has been thoroughly studied in Section 6.5 of the previous chapter. Consider X a bounded convex set such that $X \subset E = \mathbf{R}^n$, and we have $\|\cdot\| = \|\cdot\|_1$. Our objective is to accurately recover an s-sparse unconditional ground truth x_*. Recall that these assumptions are made in our analysis :

- the stochastic gradients $\nabla G(\cdot, \omega)$ are assumed to be ν-Lipschitz almost surely, i.e.,

$$\forall x, x' \in X, \quad \|\nabla G(x,\omega) - \nabla G(x',\omega)\|_\infty \leq \mathcal{L}(\omega)\|x - x'\|_1, \qquad \mathcal{L}(\omega) \leq \nu \quad a.s.. \tag{7.30}$$

- for some $\Sigma \in \mathbf{S}^n_+$, g is lower bounded around x_* such that

$$\forall x \in X, \quad g(x) - g(x_*) \geq \frac{\mu}{p}\|x - x_*\|^p_\Sigma. \tag{7.31}$$

- There exists two positive constants λ, ψ, such that condition $\mathbf{Q}(\lambda, \psi)$ holds for the positive definite matrix Σ.

Let us note initially that the first assumption inherently suggests that the objective function g exhibits ν-smoothness. When combined with the fact that the optimal point x_* is unconditional, it follows that the subsequent inequality holds true

$$\forall x \in X, \quad g(x) - g(x_*) \leq \frac{\nu}{2}\|x - x_*\|^2_1. \tag{7.32}$$

Additionally, recall that both of the following results are valid : $\forall x \in \mathbf{R}^n$, $\|x\|_1 \leq n\|x\|_2$ and $\|x\|_\Sigma = \sqrt{x^T\Sigma x} = \|\Sigma^{1/2}x\|_2$. These relations enable us to recast the second assumption in the following manner:

$$\forall x \in X, \quad g(x) - g(x_*) \geq \frac{\mu\varsigma_{\min}(\Sigma)^{p/2}}{pn^p}\|x - x_*\|^p_1, \tag{7.33}$$

where $\varsigma_{\min}(\Sigma)$ represents the smallest eigenvalue of the matrix Σ.
Simple calculations reveal that (7.32) and (7.33) cannot be both true for all elements of X. The first and second assumptions are, however, valid whenever $x \in X$ statisfies the following condition

$$\forall x \in X, \quad \|x - x_*\|_1 \leq \left(\frac{p\nu}{2\mu}\left(\frac{n}{\sqrt{\varsigma_{\min}(\Sigma)}}\right)^p\right)^{\frac{1}{p-2}} =: \overline{R}. \tag{7.34}$$

This indicates that the compatibility of the first and second assumptions is confined to a particular region within X of radius $\overline{R}$.
Almost analogously, we can draw few consequences from the last assumption leading to assumption RUC being true. The first consequence being that s-sparsity of x_* and condition $\mathbf{Q}(\lambda, \psi)$ being true for some $\Sigma \succcurlyeq 0$ insures that for any point $\widehat{x} \in X$ the following inequality holds true:

$$\|\widehat{x}\|_1 - \|x_*\|_1 \geq \psi\|\widehat{x} - x_*\|_1 - \sqrt{\frac{s}{\lambda}}\|\widehat{x} - x_*\|_\Sigma.$$

The precedent inequality is obtained by a direct application of the inequality (6.46) presented in the last chapter and adapted to our "vanilla" sparsity setting. The second one is that lower bound (7.31) being true implies that as soon as $\widehat{x}$ is such that $F_\kappa(\widehat{x}) - F_\kappa(x_*) \leq \upsilon$, one has

$$\begin{aligned}
\upsilon &\geq \tfrac{1}{2}\tfrac{\mu}{p}\|\widehat{x} - x_*\|_\Sigma^p + \kappa\left(\psi\|\widehat{x} - x_*\|_1 - \sqrt{\tfrac{s}{\lambda}}\|\widehat{x} - x_*\|_\Sigma\right) \\
&\geq -\tfrac{1}{2}\max_{t\geq 0}\left\{2\kappa\sqrt{\tfrac{s}{\lambda}}t - \tfrac{\mu}{p}t^p\right\} + \kappa\psi\|\widehat{x} - x_*\|_1 \\
&= -\tfrac{(2\kappa\sqrt{s/\lambda})^q}{2q\mu^{q-1}} + \kappa\psi\|\widehat{x} - x_*\|_1,
\end{aligned}$$

since the Fenchel-Legendre transform of function $t \mapsto \frac{\mu}{p}t^p$ is $t \mapsto \frac{1}{q\mu^{q-1}}t^q$ where q is such that $p^{-1} + q^{-1} = 1$. The last inequality can then be rewritten as

$$\|\widehat{x} - x_*\|_1 \leq \frac{1}{\psi}\left[\frac{\upsilon}{\kappa} + \frac{1}{q}\Big(\frac{2\kappa}{\mu}\Big)^{q-1}\Big(\frac{s}{\lambda}\Big)^{\frac{q}{2}}\right],$$

as long as $\widehat{x}$ is at a $\|\cdot\|_1$-radius of at most

$$\overline{R} = \left(\frac{p\nu}{2\mu\lambda^{p/2}}\left(\frac{n\sqrt{\lambda}}{\sqrt{\varsigma_{\min}(\Sigma)}}\right)^p\right)^{\frac{1}{p-2}}.$$

Remark 7.2.2 *In the proposed motivating example, assumption RUC holds for* $\|\cdot\| = \|\cdot\|_1$ *and* $\delta = \frac{1}{\psi}$, $\rho = \frac{1}{q}\big(\frac{2}{\mu\lambda^{p/2}}\big)^{q-1}$, $\alpha = \sqrt{\lambda/\varsigma_{min}(\Sigma)}$, *and* $\Gamma_{p,n,\|\cdot\|} = pn^{p/(p-2)}/4$.

7.2.2 Prescribed choice of parameter and convergence results

In this section, we present the analysis of the CSMD-SR algorithm, focusing on its adaptation to the new setting where the RUC Assumption is valid. Similar to the analysis provided in the last chapter, the CSMD-SR algorithm is characterized by two distinct phases: a Preliminary phase and an Asymptotic phase. Accordingly, we will outline a convergence analysis and will prescribe a choice of parameter that is associated to each phase. For any stage $k \geq 1$, we introduce the notations $\kappa_k^{(P)}$ and $m_k^{(P)}$ to denote the penalization and length of the k-th stage in the CSMD-SR algorithm's Preliminary phase, respectively. Similarly, $\kappa_k^{(A)}$ and $m_k^{(A)}$ refer to the penalization and length of the k-th stage in its Asymptotic phase.
Next lemma provides the theoretical guarantees for the multistage method when applied in solving the sparse recovery problem.

Lemma 16 *Assume that problem's parameters and the algorithm's initialization point x_0 are such that $R := \|x_0 - x^*\|$ verifies*

$$r_0 \leq R \leq R_{RUC} \wedge R^{(P)}, \tag{7.35}$$

where

$$r_0 := C_1(p)\delta\sqrt{s}\rho^{\frac{p-1}{p}}\nu^{-1/p}\sigma_*^{2/p} \quad \text{and} \quad R^{(P)} := \frac{1}{4(p-1)^{\frac{p-1}{p-2}}}(\rho^{p-1}\nu)^{\frac{1}{p-2}}.$$

This ensures that both the RUC Assumption and some other condition that will be discussed in the proof of the Lemma are verified.
Consider the size of the Preliminary phase defined as $\overline{K}_1 := \left\lceil \log_2\left(\frac{2R}{r_0}\right)\right\rceil$. Now for $k \in \{1, \dots, \overline{K}_1\}$ and $t \geq 4\sqrt{2 + \ln\left(m_{\overline{K}_1}^{(P)}\right)}$ we define the value of the penalization parameter chosen at the k-th stage and the length of the k-th stage of the Preliminary phase such as

$$\kappa_k^{(P)} := \frac{1}{\sqrt{s}}\left(\frac{R_{k-1}^2\nu\,(4\Theta + 60t)}{m_k^{(P)}\rho(q-1)}\right)^{\frac{p-1}{p}} \quad \text{and} \quad m_k^{(P)} := \left\lceil \frac{4^p p^p}{(p-1)^{p-1}}\delta^p s^{\frac{p}{2}}\rho^{p-1}\nu\,(4\Theta + 60t)\,R_{k-1}^{2-p}\right\rceil,$$

where the sequence $(R_k)_{k\geq 0}$ verifies the following recursion

$$R_{k+1} = \frac{1}{2}R_k + \frac{C_0(p)\sigma_*^2}{\nu}\left(\frac{\delta\sqrt{s}\rho^{\frac{p-1}{p}}}{R_k^{\frac{p-1}{p}}}\right)^p, \text{ and } R_0 = R.$$

Given this setup, for any $k \in \{1, \dots, \overline{K}_1\}$, the approximate solution $\widehat{x}^k_{m_k^{(P)}}$ computed at the end of the k-th stage of the CSMD-SR algorithm satisfies with probability $\geq 1 - 4ke^{-t}$

$$\|\widehat{x}^k_{m_k^{(P)}} - x_*\| \leq R_k \leq 2^{-k}R + \tfrac{1}{2}C_1(p)\delta\nu^{-\frac{1}{p}}\sigma_*^{\frac{2}{p}}\sqrt{s}\rho^{\frac{p-1}{p}}. \tag{7.36}$$

In particular, the estimate $\widehat{x}^{\overline{K}_1}_{m_{\overline{K}_1}^{(P)}}$ computed after $\overline{K}_1$ stages of the preliminary phase, satisfies with probability at least $1 - 4\overline{K}_1e^{-t}$

$$\|\widehat{x}^{\overline{K}_1}_{m_{\overline{K}_1}^{(P)}} - x_*\| \leq C_1(p)\sigma_*^{\frac{2}{p}}\delta\sqrt{s}\rho^{\frac{p-1}{p}}\nu^{-\frac{1}{p}} = r_0. \tag{7.37}$$

The values of the 'constants' $C_0(p)$ and $C_1(p)$ can be found in the proof of the lemma.

The analysis of the Preliminary phase presented in Lemma 16 demonstrates that under the RUC Assumption, the stage lengths exhibit exponential growth with the stage count. This contrasts with the behavior appearing under the RSC Assumption, where stage lengths remain constant. As a result, the linear decay observed in the Preliminary phase under the RSC Assumption does not occur in analyses based on the RUC Assumption. This is presented later in the manuscript in Theorem 7.2.1.

Now we assume that the Preliminary phase of the algorithm is terminated, in other words that we have completed $\overline{K}_1$ stages of the Preliminary phase, we transition to analysis of the Asymptotic phase. For the sake of simplicity, the analysis is provided in the mini-batch setting.

Lemma 17 *Recall that the initial radius verifies $r_0 \leq R \leq R_{RUC} \wedge R^{(P)}$ as defined in Lemma 16. Consider the size of the Asymptotic phase defined as*

$$\overline{K}_2 := \max\left\{ k \,\middle|\, \sum_{i=1}^{k} m_i^{(A)} \ell_i \leq N - \sum_{i=1}^{\overline{K}_1} m_i^{(P)} \right\}.$$

Now for $k \in \{1, \dots, \overline{K}_2\}$ and $t \geq 4\sqrt{2 + \ln\left(m^{(P)}_{\overline{K}_2}\right)}$ we define the value of the penalization parameter, the length of the stages and the batch-size chosen at the k-th stage of the Asymptotic phase such as

$$\kappa_k^{(A)} = 2^{-k(p-1)} C_2(p) \left(\frac{\sigma_*^2}{s^{\frac{p}{2(p-1)}} \rho \nu} \right)^{\frac{p-1}{p}}, \qquad m_k^{(A)} = \left\lceil 2^{k(p-2)} C_3(p) \delta^2 (\nu\rho)^{\frac{2(p-1)}{p}} s(4\Theta + 60t) \sigma_*^{\frac{2(2-p)}{p}} \right\rceil, \tag{7.38}$$

$$\ell_k = \left\lceil 2^{kp} C_4(p) \Theta \right\rceil.$$

We set the sequence $(r_k)_{k \geq 0}$ such that it verifies the following recursion

$$r_k = 2^{-k} r_0, \text{ and } r_0 = C_1(p) \sigma_*^{\frac{2}{p}} \delta \sqrt{s} \rho^{\frac{p-1}{p}} \nu^{-\frac{1}{p}}.$$

Given this setting, the approximate solution produced by the CSMD-SR Algorithm denoted $\widehat{x}^k_{m_k^{(A)}}$, satisfies at the end of the k-th stage of the Asymptotic phase, for $k \in \{1, \dots, \overline{K}_2\}$, with probability $\geq 1 - 4(\overline{K}_1 + k)e^{-t}$, $\|\widehat{x}^k_{m_k^{(A)}} - x_*\| \leq r_k$, implying that

$$\|\widehat{x}^k_{m_k^{(A)}} - x_*\| \leq \left(\frac{C_5(p) \sigma_*^2 \delta^{2p} s^p \rho^{2(p-1)} \Theta \left(4\Theta + 60t\right)}{N_k} \right)^{\frac{1}{2(p-1)}}, \tag{7.39}$$

where $N_k = \sum_{i=1}^{k} m_i^{(A)} \ell_i$ is the total count of oracle calls for k asymptotic stages. The values of the constants $C_2(p), C_3(p), C_4(p), C_5(p)$ are provided in the proof of the lemma.

Similar to the result appearing in the first lemma, Lemma 17 shows that during the Asymptotic phase both the length of the stages and the minibatch size grow exponentially with the stage count.

Now we present the main result of the current analysis. In the following theorem, we present the rate of recovery achieved by the CSMD-SR algorithm under the RUC assumption when the sample size N is fixed in advance.

Theorem 7.2.1 *Assume that the total sample budget satisfies $N \geq m_1^{(P)}$, so that at least one stage of the Preliminary phase of the CSMD-SR Algorithm is completed, then for $t \geq 4\sqrt{2 + \ln(m^{(A)}_{\overline{K}_2})}$, the corresponding solution $\widehat{x}_N^{(b)}$ of the CSMDR-SR algorithm satisfies with probability at least $1 - 4(\overline{K}_1 + \overline{K}_2)e^{-t}$*

$$\|\widehat{x}_N^{(b)} - x_*\| \leq \left(\frac{C_6(p) \delta^p s^{\frac{p}{2}} \rho^{p-1} \nu (\Theta + t)}{N} \right)^{\frac{1}{p-2}} + \left(\frac{C_7(p) \sigma_*^2 \delta^{2p} s^p \rho^{2(p-1)} \Theta \left(\Theta + t\right)}{N} \right)^{\frac{1}{2(p-1)}}.$$

where $\overline{K}_2 := \max\left\{ k \,\middle|\, \sum_{i=1}^{k} m_i^{(A)} \ell_i \leq N - \sum_{i=1}^{\overline{K}_1} m_i^{(P)} \right\}$ is the count for the number of stages of the Asymptotic phase of the algorithm. Value of $C_6(p)$ and $C_7(p)$ can be found in the proof of the theorem.

Remark 7.2.3 *At first glance there seems to be a discrepancy between the settings where $p = 2$ and $p > 2$ since the term related to the Preliminary phase in the bound of Theorem 7.2.1 exhibits a sublinear decay whereas the same term of Theorem 6.2.1 exhibits a linear decay. This difference arises due to the majoration made in the proof of the theorem in equation (7.49) to obtain equation (7.50). Indeed, if we look back at (7.49) when p is close to 2 we have $2^{k(p-2)} - 1 \underset{p\to 2}{\sim} k(p-2)$ and similarly $2^{p-1} - 1 \underset{p\to 2}{\sim} p - 2$. Plugging everything together, we have $N \lesssim m_0 k$, which ultimately leads to the bound (6.44) represented in the precedent chapter.*

Remark 7.2.4 *In the same way as the CSMD-SR algorithm can be made adaptive to the quantity ρs under the RSC assumption, it can also made adaptive to the quantity $\rho s^{\frac{q}{2}}$ by using Lepski's adaptation protocole under the RUC assumption. Note that the algorithm can also be made adaptive to the uniform convexity parameter p since the convergence bounds are monotone in p.*

7.3 Appendix: proofs.

7.3.1 Proof of Lemma 16

Proof.

1^o. First let us start by proving that $\forall k \geq 0$, we have $R_k \leq 2^{-k}R + r_0/2$.

Set $\alpha = C_0(p)\sigma_*^2\delta^p\sqrt{s}^p\rho^{p-1}\nu^{-1}$, and let us study the behaviour of the sequence

$$R_k = \frac{R_{k-1}}{2} + \frac{\alpha}{R_{k-1}^{p-1}} =: f(R_{k-1}) \quad \forall k \geq 1, \quad R_0 = R.$$

One can easily check that function f is convex and admits a minimum at $\overline{R} := (2(p-1)\alpha)^{\frac{1}{p}}$. For any initial radius $R_0 > 0$, we have $\forall k \geq 0,\ R_{k+1} = f(R_k) \geq f(\overline{R})$, where $f(\overline{R}) = \frac{q}{2}(2(p-1)\alpha)^{\frac{1}{p}}$, we thus have $\forall k \geq 1,\ R_k \geq \frac{q}{2}(2(p-1)\alpha)^{\frac{1}{p}}$. Then, by using the precedent result, we can upper bound $R_{k-1}^{-(p-1)}$ as follows, $R_{k-1}^{-(p-1)} \leq \left(\frac{2}{q}\right)^{p-1}(2(p-1)\alpha)^{-\frac{1}{q}}$. We have then shown that

$$\forall k \geq 1,\ R_k = f(R_{k-1}) \leq \frac{1}{2}R_{k-1} + \left(\frac{2}{q}\right)^{p-1}(2(p-1))^{-\frac{1}{q}}\,\alpha^{\frac{1}{p}}$$

By plugging the value of α into the last inequality and by setting

$$C_1(p)/4 := C_0(p)^{1/p}\left(\frac{2}{q}\right)^{p-1}(2(p-1))^{-\frac{1}{q}}$$

we obtain that

$$\forall k \geq 1,\ R_k \leq \frac{1}{2}R_{k-1} + \frac{r_0}{4}. \tag{7.40}$$

Now by invoking the recursive relationship of the sequence $(R_k)_{k\geq 0}$ and last inequality, we immediately have that

$$\forall k \geq 1,\ R_k \leq 2^{-k}R + \frac{r_0}{4}\sum_{i=0}^{k-1}2^{-i} \leq 2^{-k}R + \frac{r_0}{2}. \tag{7.41}$$

2^o. We provide a brief explanation of the idea of the proof. Observe that under the hypothesis validating Proposition 6.2.1, for $t \geq 4\sqrt{2+\ln(m)}$, we have with probability at least $1-4e^{-t}$, for $\widehat{x}_m$ an approximate solution obtained after having applied m-step of the CSMD algorithm, that the following inequality holds true

$$F_\kappa(\widehat{x}_m) - F_\kappa(x_*) \leq \frac{R^2}{m\gamma}(\Theta + 15t) + \frac{\kappa R}{m} + \sigma_*^2\gamma\left(7 + \frac{24t}{m}\right) := \upsilon.$$

Recall that with the choice $\gamma = (4\nu)^{-1}$, the quantity υ becomes

$$\upsilon = \frac{R^2\nu}{m}(4\Theta + 60t) + \frac{\kappa R}{m} + \frac{\sigma_*^2}{\nu}\left(\frac{7}{4} + \frac{6t}{m}\right).$$

We can now use the *Reduced Uniform Convexity* assumption since the radius R is such that $R \leq R_{\mathrm{RUC}}$. Therefore the above value of υ together with result (7.29) results in

$$\|\widehat{x}_m - x_*\| \leq \delta\left[\kappa^{q-1}s^{\frac{q}{2}}\rho + \frac{R^2\nu}{\kappa m}(4\Theta + 60t) + \frac{R}{m} + \frac{\sigma_*^2}{\kappa\nu}\left(\frac{7}{4} + \frac{6t}{m}\right)\right]. \tag{7.42}$$

The rest of the proof is carried out by induction. It consists in applying result (7.42) for each stage of the algorithm and choosing its parameters accordingly to the statement of Lemma 16.
First note that initial point x_0 satisfies $x_0 \in X_R(x_*)$ with probability 1 by definition. Now suppose that the initial point $x_0^k = \widehat{x}^{k-1}_{m_{k-1}}$ of the kth stage of the method satisfy $x_0^k \in X_{R_{k-1}}(x_*)$ with probability $1-4(k-1)e^{-t}$. In other words, there is a set $\mathcal{B}_{k-1} \subset \Omega$, with $\mathrm{Prob}(\mathcal{B}_{k-1}) \geq 1-4(k-1)e^{-t}$, such that for all $\overline{\omega}^{k-1} = [\omega_1; ...; \omega_{m_{k-1}^{(P)}}] \subset \mathcal{B}_{k-1}$ one has $x_0^k \in X_{R_{k-1}}(x_*)$. Let us show that upon termination of the kth stage $\widehat{x}^k_{m_k^{(P)}}$ satisfy $\|\widehat{x}^k_{m_k^{(P)}} - x_*\| \leq R_k$ with probability $1-4ke^{-t}$. By result (7.42) we conclude that for some $\overline{\Omega}_k \subset \Omega$, $\mathrm{Prob}(\overline{\Omega}_k) \geq 1-4e^{-t}$, solution $\widehat{x}^k_{m_k^{(P)}}$ after $m_k^{(P)}$ iterations of the stage satisfies, for all for all $\omega^k = [\omega_{m_{k-1}^{(P)}+1}, ..., \omega_{m_k^{(P)}}] \in \overline{\Omega}_k$,

$$\|\widehat{x}^k_{m_k^{(P)}} - x_*\| \leq \delta\left[{\kappa_k^{(P)}}^{q-1}s^{\frac{q}{2}}\rho + \frac{R_{k-1}^2\nu}{\kappa_k^{(P)}m_k^{(P)}}(4\Theta + 60t) + \frac{R_{k-1}}{m_k^{(P)}} + \frac{\sigma_*^2}{\kappa_k^{(P)}\nu}\left(\frac{7}{4} + \frac{6t}{m_k^{(P)}}\right)\right]. \tag{7.43}$$

We now choose $\kappa_k^{(P)}$ in order to minimize the first two terms of equation (7.43), i.e.,

$$\kappa_k^{(P)} = \frac{1}{\sqrt{s}}\left(\frac{R_{k-1}^2\nu\,(4\Theta + 60t)}{m_k^{(P)}(q-1)\rho}\right)^{\frac{1}{q}}$$

Plugging the parameter value into (7.43) gives

$$\|\widehat{x}^k_{m_k^{(P)}} - x_*\| \leq \delta\left[qR_{k-1}^{\frac{2}{p}}\sqrt{s}\rho^{\frac{1}{q}}\left(\frac{\nu\,(4\Theta + 60t)}{(q-1)m_k^{(P)}}\right)^{\frac{1}{p}} + \frac{R_{k-1}}{m_k^{(P)}} + \frac{\sigma_*^2\sqrt{s}}{\nu R_{k-1}^{\frac{2}{q}}}\left(\frac{7}{4} + \frac{6t}{m_k^{(P)}}\right)\left(\frac{(q-1)\rho m_k^{(P)}}{\nu(4\Theta + 60t)}\right)^{\frac{1}{q}}\right]. \tag{7.44}$$

Now observe that by choosing the length of each stage such that $\forall k \in \{1, \ldots, \overline{K}_1\}$

$$m_k^{(P)} = \left\lceil\left(4p\delta\sqrt{s}\rho^{\frac{p-1}{p}}\right)^p \nu(p-1)^{1-p}\,(4\Theta + 60t)\,R_{k-1}^{2-p}\right\rceil,$$

we have

$$m_k^{(P)} \geq \left(4p\sqrt{s}\rho^{\frac{p-1}{p}}\delta\right)^p \nu(p-1)^{1-p}(4\Theta+60t)R_{k-1}^{2-p},$$

the value is specifically chosen to bound the first term of (7.44) by $\frac{R_{k-1}}{4}$. We will now establish that the second term is likewise bounded by $\frac{R_{k-1}}{4}$. Let us introduce $K_1 := \log_2\left(\frac{2R}{r_0}\right)$ such that $\overline{K}_1 = \lceil K_1 \rceil$. Recall from the initial part of the proof that we proved

$$\forall k \in \{1,\dots,\overline{K}_1\},\ R_{k-1} \leq 2^{-k+1}R + \frac{r_0}{2}.$$

Coupling the latter with the choice of the initial point x_0 such that $R \geq r_0$ and

$$2^{-\overline{K}_1+1}R \geq 2^{-K_1}R = \frac{r_0}{2} \geq 2^{-\overline{K}_1}R$$

boils down to state that

$$\forall k \in \{1,\dots,\overline{K}_1\},\ R_{k-1} \leq 2^{-k+2}R. \tag{7.45}$$

By using (7.45), the assumption $R \leq R_{\mathrm{RUC}} \wedge R^{(P)}$ and $p \geq 2$, we immediately obtain that the following inequalities holds true

$$\begin{aligned} R_{k-1}^{p-2} &\leq \left(2^{(-k+2)}R^{(P)}\right)^{p-2} \\ &\leq 2^{-k(p-2)}(p-1)^{1-p}(\rho^{p-1}\nu)^{-1}. \end{aligned}$$

The last inequality gives us that

$$(4p\sqrt{s}\delta)^p\rho^{p-1}\nu(p-1)^{1-p}R_{k-1}^{2-p} \geq 2^{k(p-2)}(4p\sqrt{s}\delta)^p \geq 1,$$

since $s,\delta \geq 1$, $p \geq 2$ and $k \in \{1,\dots,\overline{K}_1\}$. We have just proved that from the latter choice of the length of each stages of the Preliminary phase $m_k^{(P)}$ we have

$$m_k^{(P)} \geq (4\Theta+60t) > 60t > 339,$$

this implies that at the same time the following inequalities holds true

$$\frac{R_{k-1}}{m_k^{(P)}} \leq \frac{R_{k-1}}{4}, \quad \text{and} \quad \frac{6t}{m_k^{(P)}} \leq \frac{1}{4}.$$

By bringing all these results together we can show that

$$\|\widehat{x}^k_{m_k^{(P)}} - x_*\| \leq \frac{R_{k-1}}{2} + \frac{2\sigma_*^2\delta\sqrt{s}}{\nu R_{k-1}^{\frac{2}{q}}}\left(\frac{(q-1)\rho m_k^{(P)}}{\nu(4\Theta+60t)}\right)^{\frac{1}{q}} \tag{7.46}$$

By using the fact that for all $x \geq 1$, $\lceil x \rceil \leq 2x$, we can bound $m_k^{(P)}$ and obtain

$$\|\widehat{x}^k_{m_k^{(P)}} - x_*\| \leq \frac{R_{k-1}}{2} + \frac{\sigma_*^2}{\nu}\frac{\delta^p\sqrt{s}^p\rho^{p-1}}{R_{k-1}^{p-1}}\underbrace{q^{p-1}2^{\frac{1}{q}+2p-1}}_{=:C_0(p)} = R_k. \tag{7.47}$$

We conclude that $\widehat{x}^k_{m_k^{(P)}} \in X_{R_k}(x_*)$ for all $\overline{\omega}^k \in \mathcal{B}_k = \mathcal{B}_{k-1} \cap \overline{\Omega}_k$, and by application of the union bound we obtain

$$\mathrm{Prob}(\mathcal{B}_k) \geq \mathrm{Prob}(\mathcal{B}_{k-1}) - \mathrm{Prob}(\overline{\Omega}_k^c) \geq 1 - 4ke^{-t}.$$

Proof of results (7.36) and (7.37) follows immediately by bounding the sequence $(R_k)_{k\geq 0}$ with (7.41) and plugging the value of $\overline{K}_1$. This concludes the proof of the lemma. □

7.3.2 Proof of Lemma 17

Proof. Note that proof of Lemma 17 is very similar to the proof of Lemma 16. It starts with the exact same arguments as invoked in the precedent proof except that now we consider the sequence $(r_k)_{k\geq 0}$ and we replace at each stage the sub-Gaussian parameter σ_*^2 by $\overline{\sigma}_*^2 = \frac{18\sigma_*^2\Theta}{\ell_k}$ since we are using mini-batches of increasing size ℓ_k (cf. Lemma 11 presented in the previous chapter). In other words, for $k \in \{1, \dots, \overline{K}_2\}$ the claim (7.43) becomes

$$\|\widehat{x}^k_{m_k^{(A)}} - x_*\| \leq \delta \left[{\kappa_k^{(A)}}^{q-1} s^{\frac{q}{2}} \rho + \frac{\nu r_{k-1}^2}{m_k^{(A)} \kappa_k^{(A)}} (4\Theta + 60t) + \frac{r_{k-1}}{m_k^{(A)}} + \frac{18\sigma_*^2\Theta}{\kappa_k^{(A)} \ell_k \nu} \left(\frac{7}{4} + \frac{6t}{m_k^{(A)}} \right) \right],$$

We choose the parameters $\kappa_k^{(A)}$ and $m_k^{(A)}$ using the same arguments as presented in the proof of Lemma 16, except that now, after having chosen $\kappa_k^{(A)}$ and plugged its value in the previous relationship we obtain

$$\|\widehat{x}^k_{m_k^{(A)}} - x_*\| \delta \left[q r_{k-1}^{\frac{2}{p}} \sqrt{s} \rho^{\frac{1}{q}} \left(\frac{\nu (4\Theta + 60t)}{(q-1) m_k^{(P)}} \right)^{\frac{1}{p}} + \frac{r_{k-1}}{m_k^{(P)}} + \frac{18\sigma_*^2\Theta}{\kappa_k^{(A)} \ell_k \nu} \left(\frac{7}{4} + \frac{6t}{m_k^{(P)}} \right) \right]$$

The parameter $m_k^{(A)}$ is then chosen to bound each of the appearing first two terms of the last inequality by $\frac{r_{k-1}}{8}$. The latter choice of parameter still implies that

$$\frac{r_{k-1}}{m_k^{(A)}} \leq \frac{r_{k-1}}{8}, \text{ and } \frac{6t}{m_k^{(A)}} < \frac{1}{4}$$

since $r_0 \leq R \leq R_{\mathrm{RUC}} \wedge R^{(P)}$, and it results in the following bound

$$\|\widehat{x}^k_{m_k^{(A)}} - x_*\| \leq \frac{r_{k-1}}{4} + \frac{36\delta\sigma_*^2\Theta}{\kappa_k^{(A)} \ell_k \nu}.$$

Then we choose $\ell_k = \lceil \frac{144\delta\sigma_*^2\Theta}{r_{k-1}\kappa_k^{(A)}\nu} \rceil$ in order to bound the last term by $\frac{r_{k-1}}{4}$, this finally results in

$$\forall k \in \{1, \dots, \overline{K}_2\}, \quad \|\widehat{x}^k_{m_k^{(A)}} - x_*\| \leq \frac{r_{k-1}}{4} + \frac{r_{k-1}}{4} = \frac{r_{k-1}}{2} = r_k = 2^{-k} r_0. \tag{7.48}$$

This immediately results in (7.39). Values of the constants are provided below for the interested reader:

$$C_2(p) = \left(\frac{C_1(p)}{4q} \right)^{p-1}, \quad C_3(p) = 2(p-1)(4q)^p C_1(p)^{2-p}, \quad C_4(p) = 72 \times \frac{(4q)^{p-1}}{C_1(p)^p}.$$

Now let us express the total count of oracle call after k stages of asymptotic stage.

$$\begin{aligned} N_k &= \sum_{i=1}^k m_i^{(A)} \ell_i = C_3(p) C_4(p) \Theta (4\Theta + 60t) s \delta^2 (\rho\nu)^{\frac{2}{q}} \sigma_*^{\frac{2(q-2)}{q}} \sum_{i=1}^k 2^{2i(p-1)} \\ &\leq \frac{C_3(p) C_4(p)}{4^{p-1} - 1} \Theta (4\Theta + 60t) s \delta^2 (\rho\nu)^{\frac{2}{q}} \sigma_*^{\frac{2(q-2)}{q}} 2^{2(p-1)k}. \end{aligned}$$

By inverting the last inequality, we express 2^{-k} as a function of N_k up to a constant depending only on p, and result (7.39) follows after plugging this value and the value of r_0 within result (7.48). The multiplicative term appearing in (7.48) and denoted $C_5(p)$ is as follows :

$$C_5(p) := \frac{C_1(p)^{2(p-1)} C_3(p) C_4(p)}{4^{p-1} - 1}.$$

□

7.3.3 Proof of Theorem 7.2.1

Proof. Let assume that the "total observation budget" N is such that only the preliminary phase of the procedure is implemented. This is the case when either $\sum_{i=1}^{\overline{K}_1} m_i^{(P)} \geq N$, or $\sum_{i=1}^{\overline{K}_1} m_i^{(P)} < N$ and $\sum_{i=1}^{\overline{K}_1} m_i^{(P)} + m_1^{(A)} \ell_1 > N$. The output $\widehat{x}_N$ of the algorithm is then the last update of the Preliminary phase, and by Lemma 16 it satisfies $\|\widehat{x}_N - x_*\| \leq R_k \leq 2^{-k+1} R$ where k is the count of completed stages. In the first case we have that

$$N \leq \sum_{i=1}^{k} m_i^{(P)} < \tfrac{2}{q-1} \left(4q\sqrt{s}\rho^{\frac{1}{q}}\delta\right)^p \nu\,(4\Theta + 60t) \sum_{i=0}^{k-1} R_i^{2-p}$$

Note that a proof by induction readily demonstrates that $\forall k \in \{1, \ldots, \overline{K}_1\}$, we have $2^{-k} R \leq R_k$. Consequently, it follows that $2^{-k(p-2)} R^{p-2} \leq R_k^{p-2}$, which yields the following result:

$$\begin{aligned} N &< \tfrac{2}{q-1} \left(4q\sqrt{s}\rho^{\frac{1}{q}}\delta\right)^p \nu\,(4\Theta + 60t)\, R^{2-p} \sum_{i=0}^{k-1} 2^{i(p-2)} \\ &= \tfrac{2}{q-1} \left(4q\sqrt{s}\rho^{\frac{1}{q}}\delta\right)^p \nu\,(4\Theta + 60t)\, R^{2-p} \frac{2^{k(p-2)} - 1}{2^{p-2} - 1} \qquad (7.49) \\ &< \tfrac{2}{q-1} \left(4q\sqrt{s}\rho^{\frac{1}{q}}\delta\right)^p \nu\,(4\Theta + 60t)\, R^{2-p} \frac{2^{k(p-2)}}{2^{p-2} - 1}. \qquad (7.50) \end{aligned}$$

The following bound can be obtained by rearranging the last equation

$$2^{-k} \leq \frac{1}{R} \left(\tfrac{2}{q-1} \left(4q\sqrt{s}\rho^{\frac{1}{q}}\delta\right)^p \frac{\nu\,(4\Theta + 60t)}{N(2^{p-2} - 1)} \right)^{\frac{1}{p-2}}.$$

Finally, we have shown that with probability at least $1 - 4ke^{-t}$

$$\|\widehat{x}_N - x_*\| \leq \left(\frac{C_6(p) \delta^p s^{\frac{p}{2}} \rho^{p-1} \nu(\Theta + t)}{N} \right)^{\frac{1}{p-2}}, \qquad (7.51)$$

where $C_6(p) := 30(p-1)\frac{(8q)^p}{2^{p-2}-1}$.
On the other hand, when $\sum_{i=1}^{\overline{K}_1} m_i^{(P)} < N < \sum_{i=1}^{\overline{K}_1} m_i^{(P)} + m_1^{(A)} \ell_1$, by using the definition of $m_i^{(P)}, \overline{K}_1, m_1^{(A)}$ and ℓ_1, and by giving a similar reasoning as stated above, one has

$$N < 4^{p-1} C_4(p) C_3(p) \delta^2 (\nu\rho)^{\frac{2}{q}} s\Theta(4\Theta + 60t)\sigma_*^{\frac{2(q-2)}{q}} + \tfrac{2}{q-1} \left(4q\sqrt{s}\rho^{\frac{1}{q}}\delta\right)^p \nu\,(4\Theta + 60t)\, R^{2-p} \frac{2^{\overline{K}_1(p-2)}}{2^{p-2} - 1}.$$

Recall that we have

$$\overline{K}_1 = \left\lceil \log_2 \left(\frac{2R\nu^{\frac{1}{p}}}{C_1(p)\delta\sigma_*^{\frac{2}{p}}\sqrt{s}\rho^{\frac{1}{q}}} \right) \right\rceil \leq \log_2 \left(\frac{2R\nu^{\frac{1}{p}}}{C_1(p)\delta\sigma_*^{\frac{2}{p}}\sqrt{s}\rho^{\frac{1}{q}}} \right) + 1.$$

Plugging this into the last equation and using the fact that $2\Theta \geq 1$, we have

$$N < \underbrace{\left[4^{p-1}C_4(p)C_3(p) + \frac{2^{4p+5}q^p}{(q-1)(2^{p-2}-1)C_1(p)^{p-2}} \right]}_{=:\widetilde{C}_7(p)} \delta^2(\nu\rho)^{\frac{2}{q}} s\Theta(4\Theta + 60t)\sigma_*^{\frac{2(q-2)}{q}}.$$

Or similarly,

$$\nu^{-\frac{1}{p}} < \left(\frac{\widetilde{C}_7(p)\delta^2\rho^{\frac{2}{q}} s\Theta(4\Theta + 60t)\sigma_*^{\frac{2(q-2)}{q}}}{N} \right)^{\frac{1}{2(p-1)}}.$$

Note that as the Preliminary phase is terminated, bound (7.37) is valid with probability greater than $1 - 4\overline{K}_1 e^{-t}$, this together with the previous inequality results in the following bound

$$\|\widehat{x}_N - x_*\| \leq C_1(p)\sigma_*^{\frac{2}{p}}\delta\sqrt{s}\rho^{\frac{1}{q}}\nu^{-\frac{1}{p}} \leq \left(\frac{\overline{C}_7(p)\sigma_*^2\delta^{2p}s^p\rho^{2(p-1)}\Theta\left(\Theta + t\right)}{N} \right)^{\frac{1}{2(p-1)}}$$

where $\overline{C}_7(p) := 60C_1(p)^{2(p-1)}\widetilde{C}_7(p)$.

Now, consider the case where at least one asymptotic stage has been completed. When $\sum_{i=1}^{\overline{K}_1} m_i^{(P)} > \frac{N}{2}$ we still have $N \leq 2\sum_{i=1}^{\overline{K}_1} m_i^{(P)}$, so that the bound ((7.51)) holds for the approximate solution $\widehat{x}_N^{(b)}$ at the end of the Asymptotic stage with the same multiplicative constant. Otherwise, the number of oracle calls N_k of asymptotic stages satisfies $N_k \geq N/2$, and by ((7.39)) this implies that with probability $\geq 1 - 4(\overline{K}_1 + \overline{K}_2)e^{-t}$,

$$\|\widehat{x}_N^{(b)} - x_*\| \leq \left(\frac{120C_5(p)\sigma_*^2\delta^{2p}s^p\rho^{2(p-1)}\Theta\left(\Theta + t\right)}{N} \right)^{\frac{1}{2(p-1)}}.$$

We then set $C_7(p) := \overline{C}_7(p) \bigvee 120C_5(p)$ to obtain the value of the last multiplicative term.

To summarize, in both cases, after termination of the algorithm, the bound of Theorem 7.2.1 holds with probability at least $1 - 4(\overline{K}_1 + \overline{K}_2)e^{-t}$. This concludes the proof. □

Bibliography

1. Mourtada, J. Exact minimax risk for linear least squares, and the lower tail of sample covariance matrices. *The Annals of Statistics* **50,** 2157–2178 (2022) (Cited on pages 13, 16, 17).
2. Boucheron, S., Lugosi, G. & Massart, P. *Concentration Inequalities: A Nonasymptotic Theory of Independence,(2013)* (Cited on page 13).
3. Groetsch, C. The theory of Tikhonov regularization for Fredholm equations. *104p, Boston Pitman Publication* (1984) (Cited on page 13).
4. Tsybakov, A. B. & Tsybakov, A. B. Nonparametric estimators. *Introduction to Nonparametric Estimation,* 1–76 (2009) (Cited on pages 14, 43, 110).
5. Lepskii, O. On a problem of adaptive estimation in Gaussian white noise. *Theory of Probability & Its Applications* **35,** 454–466 (1991) (Cited on pages 15, 23, 38, 43, 45, 182).
6. Donoho, D. L. & Johnstone, I. M. Ideal spatial adaptation by wavelet shrinkage. *biometrika* **81,** 425–455 (1994) (Cited on pages 15, 21).
7. Donoho, D. L. & Johnstone, I. M. Adapting to unknown smoothness via wavelet shrinkage. *Journal of the american statistical association* **90,** 1200–1224 (1995) (Cited on pages 15, 43).
8. Donoho, D. L. & Johnstone, I. M. Minimax estimation via wavelet shrinkage. *The annals of Statistics* **26,** 879–921 (1998) (Cited on pages 15, 21).
9. Hastie, T., Tibshirani, R., Friedman, J. H. & Friedman, J. H. *The elements of statistical learning: data mining, inference, and prediction* (Springer, 2009) (Cited on page 15).
10. Lepskii, O. Asymptotically minimax adaptive estimation. II. Schemes without optimal adaptation: Adaptive estimators. *Theory of Probability & Its Applications* **37,** 433–448 (1993) (Cited on pages 15, 23, 43).
11. Lepskii, O. Asymptotically minimax adaptive estimation. i: Upper bounds. optimally adaptive estimates. *Theory of Probability & Its Applications* **36,** 682–697 (1992) (Cited on pages 15, 23, 43).
12. Massart, P. *Concentration inequalities and model selection: Ecole d'Eté de Probabilités de Saint-Flour XXXIII-2003* (Springer, 2007) (Cited on page 15).
13. Ben-Tal, A. & Nemirovski, A. *Lectures on modern convex optimization: analysis, algorithms, and engineering applications* (SIAM, 2001) (Cited on pages 16, 47, 99, 141, 209).
14. Nemirovskij, A. S. & Yudin, D. B. Problem complexity and method efficiency in optimization (1983) (Cited on page 16).
15. Nesterov, Y. & Nemirovskii, A. *Interior-point polynomial algorithms in convex programming* (SIAM, 1994) (Cited on page 16).

16. Grant, M., Boyd, S. & Ye, Y. *CVX: Matlab software for disciplined convex programming* 2009 (Cited on pages 16, 44, 92, 100).

17. Kuks, J. & Olman, W. Minimax linear estimation of regression coefficients (I). *Iswestija Akademija Nauk Estonskoj SSR* **20,** 480–482 (1971) (Cited on pages 17, 109, 110).

18. Arnold, B. F. & Stahlecker, P. Another view of the Kuks–Olman estimator. *Journal of statistical planning and inference* **89,** 169–174 (2000) (Cited on page 17).

19. Rao, C. R., Rao, C. R., Statistiker, M., Rao, C. R. & Rao, C. R. *Linear statistical inference and its applications* (Wiley New York, 1973) (Cited on page 17).

20. Rao, C. R. Estimation of parameters in a linear model. *The Annals of Statistics* **4,** 1023–1037 (1976) (Cited on page 17).

21. Pilz, J. Minimax linear regression estimation with symmetric parameter restrictions. *Journal of Statistical Planning and Inference* **13,** 297–318 (1986) (Cited on page 17).

22. Drygas, H. Spectral methods in linear minimax estimation. *Acta Applicandae Mathematica* **43,** 17–42 (1996) (Cited on page 17).

23. Pinsker, M. S. Optimal filtering of square-integrable signals in Gaussian noise. *Problemy Peredachi Informatsii* **16,** 52–68 (1980) (Cited on pages 17–19, 43, 110).

24. Efromovich, S. & Pinsker, M. Sharp-optimal and adaptive estimation for heteroscedastic nonparametric regression. *Statistica Sinica,* 925–942 (1996) (Cited on pages 17, 43, 110).

25. Johnstone, I. M. & Silverman, B. W. Speed of estimation in positron emission tomography and related inverse problems. *The Annals of Statistics* **18,** 251–280 (1990) (Cited on pages 17, 109).

26. Johnstone, I. M. & Silverman, B. W. Discretization effects in statistical inverse problems. *Journal of complexity* **7,** 1–34 (1991) (Cited on pages 17, 109).

27. Mair, B. A. & Ruymgaart, F. H. Statistical inverse estimation in Hilbert scales. *SIAM Journal on Applied Mathematics* **56,** 1424–1444 (1996) (Cited on pages 17, 109).

28. Donoho, D. L., Liu, R. C. & MacGibbon, B. Minimax risk over hyperrectangles, and implications. *The Annals of Statistics,* 1416–1437 (1990) (Cited on pages 18, 19, 43, 56, 110).

29. Hansen, P. C. The truncated SVD as a method for regularization. *BIT Numerical Mathematics* **27,** 534–553 (1987) (Cited on page 18).

30. Juditsky, A. & Nemirovski, A. near-optimality of linear recovery in gaussian observation scheme under $\|\cdot\|$ 2 2-loss. *The Annals of Statistics* **46,** 1603–1629 (2018) (Cited on pages 19, 21, 22, 42, 43, 46, 62, 92–94, 109, 110, 112).

31. Juditsky, A. & Nemirovski, A. Near-optimality of linear recovery from indirect observations. *Mathematical Statistics and Learning* **1,** 171–225 (2018) (Cited on pages 19, 21, 22, 42, 43, 110).

32. Juditsky, A. & Nemirovski, A. *Statistical inference via convex optimization* (Princeton University Press, 2020) (Cited on pages 19, 20, 22, 27, 41, 43, 45–47, 57, 73, 74, 92–96, 109–113, 116–118, 120).

33. Juditsky, A. & Nemirovski, A. On polyhedral estimation of signals via indirect observations (2020) (Cited on pages 21, 22, 41–43, 45, 92, 94, 109, 110, 117).

34. Donoho, D. L. Nonlinear solution of linear inverse problems by wavelet–vaguelette decomposition. *Applied and computational harmonic analysis* **2,** 101–126 (1995) (Cited on pages 21, 109).

35. Goldenshluger, A. & Nemirovski, A. On spatially adaptive estimation of nonparametric regression. *Mathematical methods of Statistics* **6,** 135–170 (1997) (Cited on page 21).

36. Nemirovski, A. Topics in non-parametric. *Ecole d'Eté de Probabilités de Saint-Flour* **28,** 85 (2000) (Cited on pages 21, 43, 110).

37. Bekri, Y., Juditsky, A. & Nemirovski, A. First order algorithms for computing linear and polyhedral estimates. *arXiv preprint arXiv:2312.14691* (2023) (Cited on page 24).

38. Juditsky, A. B., Bekri, Y. & Nemirovski, A. Robust signal recovery under uncertain-but-bounded perturbations in observation matrix (2023) (Cited on pages 24, 95).

39. Bekri, Y., Juditsky, A. B. & Nemirovsky, A. S. Estimation from indirect observations under stochastic uncertainty in observation matrix (2023) (Cited on page 24).

40. Rigollet, P. & Hütter, J.-C. High-dimensional statistics. *arXiv preprint arXiv:2310.19244* (2023) (Cited on page 25).

41. Candès, E. J. & Davenport, M. A. *How well can we estimate a sparse vector?* 2013. arXiv: 1104.5246 [cs.IT] (Cited on page 26).

42. Tibshirani, R. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology* **58,** 267–288 (1996) (Cited on page 26).

43. Candes, E. & Tao, T. The Dantzig selector: Statistical estimation when p is much larger than n (2007) (Cited on pages 26, 148).

44. Donoho, D. L. For most large underdetermined systems of linear equations the minimal ℓ_1-norm solution is also the sparsest solution. *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences* **59,** 797–829 (2006) (Cited on page 26).

45. Donoho, D. L. & Elad, M. Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ_1 minimization. *Proceedings of the National Academy of Sciences* **100,** 2197–2202 (2003) (Cited on page 26).

46. Candes, E. J. & Tao, T. Decoding by linear programming. *IEEE transactions on information theory* **51,** 4203–4215 (2005) (Cited on page 27).

47. Candes, E. J., Romberg, J. K. & Tao, T. Stable signal recovery from incomplete and inaccurate measurements. *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences* **59,** 1207–1223 (2006) (Cited on pages 28, 148).

48. Bickel, P. J., Ritov, Y. & Tsybakov, A. B. Simultaneous analysis of Lasso and Dantzig selector (2009) (Cited on pages 28, 29, 148, 155, 157).

49. Zhou, S. Restricted eigenvalue conditions on subgaussian matrices. *arXiv preprint ArXiv:0904.4723* (2009) (Cited on page 29).

50. Bellec, P. C., Lecué, G. & Tsybakov, A. B. Slope meets lasso: improved oracle bounds and optimality. *The Annals of Statistics* **46,** 3603–3642 (2018) (Cited on page 29).

51. Daubechies, I., Defrise, M. & De Mol, C. Sparsity-enforcing regularisation and ISTA revisited. *Inverse problems* **32,** 104001 (2016) (Cited on page 29).

52. Beck, A. & Teboulle, M. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM journal on imaging sciences* **2,** 183–202 (2009) (Cited on page 29).

53. Friedman, J., Hastie, T. & Tibshirani, R. Regularization paths for generalized linear models via coordinate descent. *Journal of statistical software* **33,** 1 (2010) (Cited on page 30).

54. Beck, A. & Tetruashvili, L. On the convergence of block coordinate descent type methods. *SIAM journal on Optimization* **23,** 2037–2060 (2013) (Cited on page 30).

55. Robbins, H. & Monro, S. A stochastic approximation method. *The annals of mathematical statistics,* 400–407 (1951) (Cited on page 30).

56. Nemirovskii, A. & IUDIN, D. Complexity of problems and effectiveness of methods of optimization(Russian book). *Moscow, Izdatel'stvo Nauka, 1979. 384* (1979) (Cited on pages 31, 123, 134, 148).

57. Bregman, L. M. The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming. *USSR computational mathematics and mathematical physics* **7,** 200–217 (1967) (Cited on page 31).

58. Lan, G. *First-order and stochastic optimization methods for machine learning* (Springer, 2020) (Cited on pages 32, 33).

59. Nesterov, Y. Primal-dual subgradient methods for convex problems. *Mathematical programming* **120,** 221–259 (2009) (Cited on pages 33, 166).

60. Iouditski, A. & Nesterov, Y. Primal-dual subgradient methods for minimizing uniformly convex functions. *arXiv preprint arXiv:1401.1792* (2014) (Cited on pages 33–35).

61. Nazin, A. V., Nemirovsky, A. S., Tsybakov, A. B. & Juditsky, A. B. Algorithms of robust stochastic optimization based on mirror descent method. *Automation and Remote Control* **80,** 1607–1627 (2019) (Cited on page 34).

62. Nguyen, N., Needell, D. & Woolf, T. Linear convergence of stochastic iterative greedy algorithms with sparse constraints. *IEEE Transactions on Information Theory* **63,** 6869–6895 (2017) (Cited on page 34).

63. Agarwal, A., Negahban, S. & Wainwright, M. J. Stochastic optimization and sparse statistical recovery: Optimal algorithms for high dimensions. *Advances in Neural Information Processing Systems* **25** (2012) (Cited on pages 35, 37, 148, 149).

64. Juditsky, A., Kulunchakov, A. & Tsyntseus, H. Sparse recovery by reduced variance stochastic approximation. *Information and Inference: A Journal of the IMA* **12,** 851–896 (2023) (Cited on pages 36–38, 158).

65. Laurent, B. & Massart, P. Adaptive estimation of a quadratic functional by model selection. *Annals of statistics,* 1302–1338 (2000) (Cited on pages 43, 60, 130, 164).

66. Goldenshluger, A. & Lepski, O. Bandwidth selection in kernel density estimation: oracle inequalities and adaptive minimax optimality (2011) (Cited on page 43).

67. Neumann, M. H. & von Sachs, R. in *Wavelets and statistics* 301–329 (Springer, 1995) (Cited on page 43).

68. Birgé, L. Estimating a density under order restrictions: Nonasymptotic minimax risk. *The Annals of Statistics,* 995–1012 (1987) (Cited on page 43).

69. Efromovich, S. *Nonparametric curve estimation: methods, theory, and applications* (Springer Science & Business Media, 2008) (Cited on pages 43, 110).

70. Ibragimov, I. A. & Has' Minskii, R. Z. *Statistical estimation: asymptotic theory* (Springer Science & Business Media, 2013) (Cited on pages 43, 110).

71. Wasserman, L. *All of nonparametric statistics* (Springer Science & Business Media, 2006) (Cited on pages 43, 110).

72. Nemirovskii, A. Nonparametric estimation of smooth regression functions. *Soviet Journal of Computer and Systems Sciences* **23,** 1–11 (1985) (Cited on pages 43, 110).

73. Grasmair, M., Li, H. & Munk, A. Variational multiscale nonparametric regression: Smooth functions (2018) (Cited on pages 43, 110).

74. Proksch, K., Werner, F. & Munk, A. Multiscale scanning in inverse problems (2018) (Cited on pages 43, 109, 110).

75. Nemirovski, A., Roos, C. & Terlaky, T. On maximization of quadratic form over intersection of ellipsoids with common center. *Mathematical programming* **86,** 463–473 (1999) (Cited on page 47).

76. Goldenshluger, A. A universal procedure for aggregating estimators. *The Annals of Statistics,* 542–568 (2009) (Cited on page 54).

77. Juditsky, A. & Nemirovski, A. Aggregating estimates by convex optimization. *Mathematical Statistics and Learning* **5,** 55–116 (2022) (Cited on page 54).

78. Rudelson, M. & Vershynin, R. Hanson-Wright inequality and sub-Gaussian concentration (2013) (Cited on pages 55, 105, 129).

79. Zajkowski, K. Bounds on tail probabilities for quadratic forms in dependent sub-gaussian random variables. *Statistics & Probability Letters* **167,** 108898 (2020) (Cited on page 55).

80. Klemelä, J. Sharp adaptive estimation of quadratic functionals. *Probability theory and related fields* **134,** 539–564 (2006) (Cited on page 56).

81. Cai, T. T. & Low, M. G. Nonquadratic estimators of a quadratic functional (2005) (Cited on page 56).

82. Cai, T. T. & Low, M. G. Optimal adaptive estimation of a quadratic functional (2006) (Cited on page 56).

83. Juditsky, A. & Nemirovski, A. Accuracy Guarantees for ℓ_1-Recovery. *IEEE Transactions on Information Theory* **57,** 7818–7839 (2011) (Cited on pages 72, 73, 155).

84. Ben-Tal, A., Boyd, S. & Nemirovski, A. Extending scope of robust optimization: Comprehensive robust counterparts of uncertain problems. *Mathematical Programming* **107,** 63–89 (2006) (Cited on page 76).

85. Nesterov, Y. & Nemirovski, A. On first-order algorithms for l1/nuclear norm minimization. *Acta Numerica* **22,** 509–575 (2013) (Cited on pages 100, 150, 151, 160).

86. Ben-Tal, A. & Nemirovski, A. Non-Euclidean restricted memory level method for large-scale convex optimization. *Mathematical Programming* **102,** 407–456 (2005) (Cited on page 102).

87. Kiwiel, K. C. Proximal level bundle methods for convex nondifferentiable optimization, saddle-point problems and variational inequalities. *Mathematical Programming* **69,** 89–109 (1995) (Cited on page 102).

88. Hsu, D., Kakade, S. & Zhang, T. A tail inequality for quadratic forms of subgaussian random vectors (2012) (Cited on pages 105, 129).

89. Spokoiny, V. & Zhilova, M. Sharp deviation bounds for quadratic forms. *Mathematical Methods of Statistics* **22,** 100–113 (2013) (Cited on pages 105, 129).

90. Natterer, F. *The mathematics of computerized tomography* (SIAM, 2001) (Cited on page 109).

91. Vogel, C. R. *Computational methods for inverse problems* (SIAM, 2002) (Cited on page 109).

92. Kaipio, J. & Somersalo, E. *Statistical and computational inverse problems* (Springer Science & Business Media, 2006) (Cited on page 109).

93. Hoffmann, M. & Reiss, M. Nonlinear estimation for linear inverse problems with error in the operator (2008) (Cited on pages 109, 110).

94. Cohen, A., Hoffmann, M. & Reiss, M. Adaptive wavelet Galerkin methods for linear inverse problems. *SIAM Journal on Numerical Analysis* **42,** 1479–1501 (2004) (Cited on page 109).

95. Juditsky, A. B. & Nemirovski, A. S. Nonparametric estimation by convex programming (2009) (Cited on page 110).

96. Bennani, M., Brunet, M.-C. & Chatelin, F. De l'utilisation en calcul matriciel de modèles probabilistes pour la simulation des erreurs de calcul. *Comptes rendus de l'Académie des sciences. Série 1, Mathématique* **307,** 847–850 (1988) (Cited on page 110).

97. Carroll, R. & Ruppert, D. The use and misuse of orthogonal regression in linear errors-in-variables models. *The American Statistician* **50,** 1–6 (1996) (Cited on page 110).

98. Fan, J. & Truong, Y. K. Nonparametric regression with errors in variables. *The Annals of Statistics,* 1900–1925 (1993) (Cited on page 110).

99. Gleser, L. J. Estimation in a multivariate" errors in variables" regression model: large sample results. *The Annals of Statistics,* 24–44 (1981) (Cited on page 110).

100. Kukush, A., Markovsky, I. & Van Huffel, S. Consistency of the structured total least squares estimator in a multivariate errors-in-variables model. *Journal of Statistical Planning and Inference* **133,** 315–358 (2005) (Cited on page 110).

101. Stewart, G. W. Stochastic perturbation theory. *SIAM review* **32,** 579–610 (1990) (Cited on page 110).

102. Van Huffel, S. & Lemmerling, P. *Total least squares and errors-in-variables modeling: analysis, algorithms and applications* (Springer Science & Business Media, 2013) (Cited on page 110).

103. Cavalier, L. & Hengartner, N. W. Adaptive estimation for inverse problems with noisy operators. *Inverse problems* **21,** 1345 (2005) (Cited on page 110).

104. Cavalier, L. & Raimondo, M. Wavelet deconvolution with noisy eigenvalues. *IEEE Transactions on signal processing* **55,** 2414–2424 (2007) (Cited on page 110).

105. Efromovich, S. & Koltchinskii, V. On inverse problems with unknown operators. *IEEE Transactions on Information Theory* **47,** 2876–2894 (2001) (Cited on page 110).

106. Hall, P. & Horowitz, J. L. Nonparametric methods for inference in the presence of instrumental variables (2005) (Cited on page 110).

107. Marteau, C. Regularization of inverse problems with unknown operator. *Mathematical Methods of Statistics* **15,** 415–443 (2006) (Cited on page 110).

108. Cope, J. & Rust, B. W. Bounds on solutions of linear systems with inaccurate data. *SIAM Journal on Numerical Analysis* **16,** 950–963 (1979) (Cited on page 111).

109. Higham, N. J. *Accuracy and stability of numerical algorithms* (SIAM, 2002) (Cited on page 111).

110. Kreinovich, V., Lakeyev, A. V. & Noskov, S. I. Optimal solution of interval linear systems is intractable (NP-hard). *Interval Computations* **1,** 6–14 (1993) (Cited on page 111).

111. Nazin, S. A. & Polyak*, B. T. Interval parameter estimation under model uncertainty. *Mathematical and Computer Modelling of Dynamical Systems* **11,** 225–237 (2005) (Cited on page 111).

112. Neumaier, A. *Interval methods for systems of equations* **37** (Cambridge university press, 1990) (Cited on page 111).

113. Oettli, W. & Prager, W. Compatibility of approximate solution of linear equations with given error bounds for coefficients and right-hand sides. *Numerische Mathematik* **6,** 405–409 (1964) (Cited on page 111).

114. Polyak, B. T. in *Directions in Mathematical Systems Theory and Optimization* 249–260 (Springer, 2002) (Cited on page 111).

115. Bertsekas, D. & Rhodes, I. Recursive state estimation for a set-membership description of uncertainty. *IEEE Transactions on Automatic Control* **16,** 117–128 (1971) (Cited on page 111).

116. Casini, M., Garulli, A. & Vicino, A. Feasible parameter set approximation for linear models with bounded uncertain regressors. *IEEE Transactions on Automatic Control* **59,** 2910–2920 (2014) (Cited on page 111).

117. Cerone, V. Feasible parameter set for linear models with bounded errors in all variables. *Automatica* **29,** 1551–1555 (1993) (Cited on page 111).

118. Juditsky, A., Kotsalis, G. & Nemirovski, A. Tight computationally efficient approximation of matrix norms with applications. *Open Journal of Mathematical Optimization* **3,** 1–38 (2022) (Cited on pages 111, 118, 125–127, 142, 145).

119. Kurzhanski, A. & Vályi, I. *Ellipsoidal calculus for estimation and control* (Springer, 1997) (Cited on page 111).

120. Matasov, A. I. *Estimators for uncertain dynamic systems* (Springer Science & Business Media, 2012) (Cited on page 111).

121. Milanese, M., Norton, J., Piet-Lahanier, H. & Walter, É. *Bounding approaches to system identification* (Springer Science & Business Media, 2013) (Cited on page 111).

122. Nazin, S. A. & Polyak, B. T. Ellipsoid-based parametric estimation in the linear multidimensional systems with uncertain model description. *Automation and Remote Control* **68,** 993–1005 (2007) (Cited on page 111).

123. Tempo, R. & Vicino, A. Optimal algorithms for system identification: a review of some recent results. *Mathematics and computers in simulation* **32,** 585–595 (1990) (Cited on page 111).

124. Ben-Tal, A., El Ghaoui, L. & Nemirovski, A. *Robust optimization* (Princeton university press, 2009) (Cited on pages 111, 141).

125. Juditsky, A. & Nemirovski, A. On Design of Polyhedral Estimates in Linear Inverse Problems. *SIAM Journal on Mathematics of Data Science* **6,** 76–96 (2024) (Cited on page 120).

126. Lerasle, M. & Oliveira, R. I. Robust empirical mean estimators. *arXiv preprint arXiv:1112.3914* (2011) (Cited on page 123).

127. Hsu, D. & Sabato, S. *Heavy-tailed regression with a generalized median-of-means* in *International Conference on Machine Learning* (2014), 37–45 (Cited on page 123).

128. Minsker, S. Geometric median and robust estimation in Banach spaces (2015) (Cited on pages 123, 135).

129. Lecué, G. & Lerasle, M. Robust machine learning by median-of-means: theory and practice (2020) (Cited on page 123).

130. Tropp, J. A. *et al.* An introduction to matrix concentration inequalities. *Foundations and Trends® in Machine Learning* **8,** 1–230 (2015) (Cited on page 132).

131. Guigues, V., Juditsky, A. & Nemirovski, A. Hypothesis testing via Euclidean separation (2020) (Cited on page 134).

132. Boyd, S., El Ghaoui, L., Feron, E. & Balakrishnan, V. *Linear matrix inequalities in system and control theory* (SIAM, 1994) (Cited on page 138).

133. Ben-Tal, A. & Nemirovski, A. On tractable approximations of uncertain linear matrix inequalities affected by interval uncertainty. *SIAM Journal on Optimization* **12,** 811–833 (2002) (Cited on page 141).

134. Ben-Tal, A., Nemirovski, A. & Roos, C. Extended matrix cube theorems with applications to μ-theory in control. *Mathematics of Operations Research* **28,** 497–523 (2003) (Cited on page 141).

135. Van De Geer, S. A. & Bühlmann, P. On the conditions used to prove oracle results for the lasso (2009) (Cited on pages 148, 155).

136. Candès, E. J. & Plan, Y. Near-ideal model selection by ℓ-1 minimization (2009) (Cited on page 148).

137. Candes, E. J. & Plan, Y. A probabilistic and RIPless theory of compressed sensing. *IEEE transactions on information theory* **57,** 7235–7254 (2011) (Cited on page 148).

138. Meier, L., Van De Geer, S. & Bühlmann, P. The group lasso for logistic regression. *Journal of the Royal Statistical Society Series B: Statistical Methodology* **70,** 53–71 (2008) (Cited on pages 148, 157).

139. Baes, M., Burgisser, M. & Nemirovski, A. A randomized mirror-prox method for solving structured large-scale matrix saddle-point problems. *SIAM Journal on Optimization* **23,** 934–962 (2013) (Cited on page 148).

140. Juditsky, A., Kılınç Karzan, F. & Nemirovski, A. Randomized first order algorithms with applications to ℓ-1-minimization. *Mathematical Programming* **142,** 269–310 (2013) (Cited on page 148).

141. Ghaoui, L. E., Viallon, V. & Rabbani, T. Safe feature elimination for the lasso and sparse supervised learning problems. *arXiv preprint arXiv:1009.4219* (2010) (Cited on page 148).

142. Kowalski, M., Weiss, P., Gramfort, A. & Anthoine, S. *Accelerating ISTA with an active set strategy* in *OPT 2011: 4th International Workshop on Optimization for Machine Learning* (2011), 7 (Cited on page 148).

143. Mairal, J. *Sparse coding for machine learning, image processing and computer vision* PhD thesis (École normale supérieure de Cachan-ENS Cachan, 2010) (Cited on page 148).

144. Agarwal, A., Negahban, S. & Wainwright, M. J. Fast global convergence rates of gradient methods for high-dimensional statistical recovery. *Advances in Neural Information Processing Systems* **23** (2010) (Cited on pages 148, 157).

145. Blumensath, T. & Davies, M. E. Iterative hard thresholding for compressed sensing. *Applied and computational harmonic analysis* **27,** 265–274 (2009) (Cited on page 148).

146. Jain, P., Tewari, A. & Kar, P. On iterative hard thresholding methods for high-dimensional m-estimation. *Advances in neural information processing systems* **27** (2014) (Cited on page 148).

147. Barber, R. F. & Ha, W. Gradient descent with non-convex constraints: local concavity determines convergence. *Information and Inference: A Journal of the IMA* **7,** 755–806 (2018) (Cited on page 148).

148. Liu, H. & Foygel Barber, R. Between hard and soft thresholding: optimal iterative thresholding algorithms. *Information and Inference: A Journal of the IMA* **9,** 899–933 (2020) (Cited on page 148).

149. Shalev-Shwartz, S. & Tewari, A. *Stochastic methods for l 1 regularized loss minimization* in *Proceedings of the 26th Annual International Conference on Machine Learning* (2009), 929–936 (Cited on page 148).

150. Srebro, N., Sridharan, K. & Tewari, A. Smoothness, low noise and fast rates. *Advances in neural information processing systems* **23** (2010) (Cited on page 148).

151. Juditsky, A., Nemirovski, A., *et al.* First order methods for nonsmooth convex large-scale optimization, I: general purpose methods. *Optimization for Machine Learning* **30,** 121–148 (2011) (Cited on pages 148, 149, 151, 156).

152. Juditsky, A. & Nesterov, Y. Deterministic and stochastic primal-dual subgradient algorithms for uniformly convex minimization. *Stochastic Systems* **4,** 44–80 (2014) (Cited on pages 148, 179, 186).

153. Ghadimi, S. & Lan, G. Optimal stochastic approximation algorithms for strongly convex stochastic composite optimization, II: shrinking procedures and optimal algorithms. *SIAM Journal on Optimization* **23,** 2061–2089 (2013) (Cited on page 148).

154. Gaillard, P. & Wintenberger, O. *Sparse accelerated exponential weights* in *Artificial Intelligence and Statistics* (2017), 75–82 (Cited on pages 148, 149).

155. Lei, Y. & Tang, K. Stochastic composite mirror descent: Optimal bounds with high probabilities. *Advances in Neural Information Processing Systems* **31** (2018) (Cited on page 149).

156. Nesterov, Y. Gradient methods for minimizing composite functions. *Mathematical programming* **140,** 125–161 (2013) (Cited on pages 149, 150).

157. Nemirovski, A., Juditsky, A., Lan, G. & Shapiro, A. Robust stochastic approximation approach to stochastic programming. *SIAM Journal on optimization* **19,** 1574–1609 (2009) (Cited on page 151).

158. Raskutti, G., Wainwright, M. J. & Yu, B. Restricted eigenvalue properties for correlated Gaussian designs. *The Journal of Machine Learning Research* **11,** 2241–2259 (2010) (Cited on page 155).

159. Dalalyan, A. & Thompson, P. Outlier-robust estimation of a sparse linear model using ℓ-1-penalized Huber's M-estimator. *Advances in neural information processing systems* **32** (2019) (Cited on page 155).

160. Cohen, A., Dahmen, W. & DeVore, R. Compressed sensing and best k-term approximation. *Journal of the American mathematical society* **22,** 211–231 (2009) (Cited on page 155).

161. Rauhut, H. Compressive sensing and structured random matrices. *Theoretical foundations and numerical methods for sparse recovery* **9,** 92 (2010) (Cited on page 155).

162. Bercu, B. *et al. Concentration inequalities for martingales* (Springer, 2015) (Cited on pages 157, 165).

163. Xiao, L. Dual averaging method for regularized stochastic learning and online optimization. *Advances in Neural Information Processing Systems* **22** (2009) (Cited on page 158).

164. Chen, G. & Teboulle, M. Convergence analysis of a proximal-like minimization algorithm using Bregman functions. *SIAM Journal on Optimization* **3,** 538–543 (1993) (Cited on page 160).

165. Juditsky, A. & Nemirovski, A. S. Large deviations of vector-valued martingales in 2-smooth normed spaces. *arXiv preprint arXiv:0809.0813* (2008) (Cited on pages 161, 166).

166. Birgé, L. & Massart, P. Minimum contrast estimators on sieves: exponential bounds and rates of convergence. *Bernoulli,* 329–375 (1998) (Cited on page 164).

167. Fan, X., Grama, I. & Liu, Q. Hoeffding's inequality for supermartingales. *Stochastic Processes and their Applications* **122,** 3545–3559 (2012) (Cited on page 165).

168. Juditsky, A., Karzan, F. K. & Nemirovski, A. On a unified view of nullspace-type conditions for recoveries associated with general sparsity structures. *Linear Algebra and its Applications* **441,** 124–151 (2014) (Cited on pages 172, 174).

169. Recht, B., Xu, W. & Hassibi, B. Null space conditions and thresholds for rank minimization. *Mathematical programming* **127,** 175–202 (2011) (Cited on page 174).

170. Nesterov, Y. Universal gradient methods for convex optimization problems. *Mathematical Programming* **152,** 381–404 (2015) (Cited on page 179).

171. Bach, F. Adaptivity of averaged stochastic gradient descent to local strong convexity for logistic regression. *The Journal of Machine Learning Research* **15,** 595–627 (2014) (Cited on page 179).

172. Roulet, V. & d'Aspremont, A. Sharpness, restart and acceleration. *Advances in Neural Information Processing Systems* **30** (2017) (Cited on pages 180, 186).

173. Polyak, B. T. *Existence theorems and convergence of minimizing sequences for extremal problems with constraints* in *Doklady Akademii Nauk* **166** (1966), 287–290 (Cited on page 186).

174. Azé, D. & Penot, J.-P. *Uniformly convex and uniformly smooth convex functions* in *Annales de la Faculté des sciences de Toulouse: Mathématiques* **4** (1995), 705–730 (Cited on page 186).

175. Vladimirov, A., Nesterov, Y. E. & Chekanov, Y. N. On uniformly convex functionals. *Vestnik Moskov. Univ. Ser. XV Vychisl. Mat. Kibernet* **3,** 12–23 (1978) (Cited on page 186).

Chapter 8

Executive summary on Conic Programming

8.1 Cones

A *cone* in Euclidean space E is a nonempty set K which is stable w.r.t. taking linear combinations of its element with nonnegative coefficients[1]. More precisely, subset $K \subset E$ is a cone if it is nonempty, and

- $x, y \in K \Rightarrow x + y \in K$;
- $x \in K, \lambda \geq 0 \Rightarrow \lambda x \in K$.

Note that definition of a cone directly implies that is also a convex set. We call a cone K

- *regular* if it is closed,
- *pointed* if it is such that $K \bigcap [-K] = \{0\}$ and it possesses a nonempty interior

For $K \subset E$ a cone, we introduce its *dual cone* K^* defined as

$$K^* = \{y \in E : \langle y, x\rangle \geq 0, \forall x \in K\} \qquad [\langle \cdot, \cdot \rangle \text{ is inner product on } E.]$$

Observe that by definition, K^* is a closed cone. Moreover, one always has that $K \subset (K^*)^*$. It is also well known that

- if K is a closed cone, it holds $K = (K^*)^*$,
- K is a regular cone if and only if K^* is so.

Examples of regular cones "useful in applications" are as follows:

1. *Nonnegative orthants* $\mathbf{R}^d_+ = \left\{x \in \mathbf{R}^d : x \geq 0\right\}$;
2. *Lorentz cones* $\mathbf{L}^d_+ = \left\{x \in \mathbf{R}^d : x_d \geq \sqrt{\sum_{i=1}^{d-1} x_i^2}\right\}$;

[1] We refer to such linear combinations as *conic*.

3. *Semidefinite cones* $\mathbf{S}^d_+$ of positive semidefinite symmetric $d \times d$ matrices. Semidefinite cone $\mathbf{S}^d_+$ lives in the space $\mathbf{S}^d$ of symmetric matrices equipped with the Frobenius inner product

$$\langle A, B\rangle = \mathrm{Tr}(AB^T) = \mathrm{Tr}(AB) = \sum_{i,j=1}^{d} A_{ij}B_{ij}, \quad A, B \in \mathbf{S}^d.$$

   All cones listed so far are self-dual, i.e., such that they equal their dual.

4. Let $\|\cdot\|$ be a norm on $\mathbf{R}^n$. The set $\{[x;t] \in \mathbf{R}^n \times \mathbf{R} : t \geq \|x\|\}$ is a regular cone, and its dual cone is $\{[y;\tau] : \|y\|_* \leq \tau\}$, where

$$\|y\|_* = \max_x \{x^T y : \|x\| \leq 1\}$$

   is the norm on $\mathbf{R}^n$ conjugate to $\|\cdot\|$.

An additional example of a regular cone useful in our developements is the *conic hull* of a convex compact set. Let $\mathcal{T}$ be a convex compact set with a nonempty interior in Euclidean space E. We define its *closed conic hull* as

$$\mathbf{T} = \mathrm{cl}\, \underbrace{\left\{[t;\tau] \in E^+ = E \times \mathbf{R} : \tau > 0, t/\tau \in \mathcal{T}\right\}}_{K^o(\mathcal{T})}.$$

One directly has that $\mathbf{T}$ is a regular cone. Additionally, the latter can be obtained as the union of convex set $K^o(\mathcal{T})$ and the origin of E_+. Moreover, one can "see $\mathcal{T}$ in $\mathbf{T}$:"—the former being equal to the cross-section of $\mathbf{T}$ by the hyperplane $\tau = 1$ in $E_+ = \{[t;\tau]\}$:

$$\mathcal{T} = \{t \in E : [t;1] \in \mathbf{T}\}.$$

Finally, the cone $\mathbf{T}_*$ dual to $\mathbf{T}$ is given by

$$\mathbf{T}_* = \{[g;s] \in \mathbf{E}^+ : s \geq \phi_{\mathcal{T}}(-g)\},$$

where

$$\phi_{\mathcal{T}}(g) = \max_{t \in \mathcal{T}} \langle g, t\rangle$$

is the support function of $\mathcal{T}$.

8.1.1 Conic problems and their duals

The primal

Given regular cones $(K_i \subset E_i)_{i=1}^m$, consider an optimization problem of the form

$$\mathrm{Opt}(P) = \min\left\{\langle c, x\rangle : \begin{array}{l} A_i x - b_i \in K_i,\ i = 1, ..., m \\ Rx = r \end{array}\right\}, \qquad (P)$$

where $x \mapsto A_i x - b_i$ are affine mappings acting from some Euclidean space E to the spaces E_i. Problem (P) is called a *conic problem on the cones* $K_1, ..., K_m$; the constraints $A_i x - b_i \in K_i$ on x are called *conic constraints*. A conic problem (P) is *strictly feasible* if it admits a *strictly feasible* solution $\bar{x}$, meaning that the latter

- satisfies the equality constraints $R\bar{x} = r$
- satisfies *strictly* the conic constraints, i.e., $A_i\bar{x} - b_i \in \mathrm{int}K_i$.

The dual

Observe that one can obtain lower bounds on the optimal value $\mathrm{Opt}(P)$ of *primal* problem (P) in the following way. Let x be a feasible solution to (P). Given aggregation weitghs $y_i \in K_i^*$, called *Lagrange multiplier*, one can proceed to the *linear aggregation of constraints* by associating each corresponding conic constraint $A_i x - b_i \in K_i$. More precisely, we have

$$\forall i, \;\; 1 \le i \le m, \;\; \langle y_i, A_i x \rangle \ge \langle y_i, b_i \rangle.$$

Similarly, we equip the system $Rx = r$ of equality constraints in (P) with Lagrange multiplier z, a vector of the same dimension as r. Naturally, we also have

$$z^T R x \ge r^T z.$$

Summing up all resulting inequalities, we arrive at the scalar linear inequality

$$\left\langle R^* z + \sum\nolimits_i A_i^* y_i, x \right\rangle \ge r^T z + \sum\nolimits_i \langle b_i, y_i \rangle \qquad (!)$$

where A_i^* are the conjugates to A_i: $\langle y, A_i x \rangle_{E_i} \equiv \langle A_i^* y, x \rangle_E$, and R^* is the conjugate of R. By construction, $(!)$ is satisfied for every x feasible for primal problem. If we impose on the y_i's and z the additional equality constraint

$$R^* z + \sum_i A_i^* y_i = c,$$

$(!)$ yields the lower bound, which is valid for all x feasible for the primal problem,

$$\langle c, x \rangle \ge r^T z + \sum_i \langle b_i, y_i \rangle.$$

From this lower bound, we introduce the *conic dual* of (P) as the problem

$$\mathrm{Opt}(D) = \max_{y_i, z} \left\{ r^T z + \sum_i \langle b_i, y_i \rangle : \begin{array}{l} y_i \in K_i^*, \, 1 \le i \le m \\ R^* z + \sum_{i=1}^m A_i^* y_i = c \end{array} \right\} \qquad (D)$$

of maximizing the latter.

The relations between the primal and the dual conic problems are the subject of the standard *Conic Duality Theorem* as follows:

Theorem 8.1.1 (Conic Duality Theorem) *Consider conic problem (P) (where all K_i are regular cones) along with its dual problem (D). Then*

1. *Duality is symmetric: the dual problem (D) is conic, and the conic dual of (D) is (equivalent to) (P);*
2. *Weak duality: It always holds $\mathrm{Opt}(D) \le \mathrm{Opt}(P)$*
3. *Strong duality: If one of the problems (P), (D) is strictly feasible and bounded, then the other problem in the pair is solvable, and the optimal values of the problems are equal to each other. In particular, if both (P) and (D) are strictly feasible, then both problems are solvable with equal optimal values.*

The proof of the Conic Duality Theorem can be found in numerous sources, e.g., in [13, Section 2.4].

[1] *For a minimization problem, boundedness means that the objective is bounded from below on the feasible set, for a maximization problem, that it is bounded from above on the feasible set.*

8.1.2 Schur Complement Lemma

When dealing with conic optimization problems, We make extensive use of the following simple lemma:

Lemma 18 *[Schur Complement Lemma] A symmetric block matrix*

$$A = \left[\begin{array}{c|c} P & Q^T \\ \hline Q & R \end{array}\right]$$

with $R \succ 0$ is positive (semi)definite if and only if the matrix $P - Q^T R^{-1} Q$ also is.

Proof. Let u, v be respectively of the same sizes as P and R. Using direct computations and the fact that $R \succ 0$, we have that

$$\begin{aligned} \min_v \left\{ [u;v]^T A [u;v] \right\} &= u^T P u + \min_v \left\{ 2v^T Q u + v^T R v \right\} \\ &= u^T [P - Q^T R^{-1} Q] u. \end{aligned}$$

It follows that the quadratic form associated with A is nonnegative everywhere if and only if the quadratic form with the matrix $[P - Q^T R^{-1} Q]$ is nonnegative everywhere. □

List of Figures