Noname manuscript No.
(will be inserted by the editor)

Estimation from Indirect Observations under Stochastic Uncertainty in Observation Matrix

Yannis Bekri · Anatoli Juditsky · Arkadi Nemirovski

the date of receipt and acceptance should be inserted later

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 [31] and polyhedral estimate introduced in [27]. It was shown in [28] 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 stochastic uncertainty in the observation matrix. In this setting, we show how to bound the estimation risk by the optimal value of an efficiently solvable convex optimization problem; "presumably good" estimates are then obtained through optimization of the risk bounds with respect to estimate parameters.

2020 Mathematics Subject Classification: 62F35, 62G35, 90C90
Keywords: statistical linear inverse problems, robust estimation, observation matrix uncertainty

Communicated by Yurii Nesterov.

Yannis Bekri
LJK, Université Grenoble Alpes
Campus de Saint-Martin-d'Hères, 38401 France
yannis.bekri@univ-grenoble-alpes.fr

Anatoli Juditsky, Corresponding author
LJK, Université Grenoble Alpes
Campus de Saint-Martin-d'Hères, 38401 France
anatoli.juditsky@univ-grenoble-alpes.fr

Arkadi Nemirovski
Georgia Institute of Technology
Atlanta, Georgia 30332, USA
nemirovs@isye.gatech.edu

1 Introduction

In this paper we focus on the problem of recovering unknown signal x given noisy observation $\omega \in \mathbf{R}^m$,

$$\omega = Ax + \xi, \tag{1}$$

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 (1) (cf., e.g., [40,21,22,37,52,30,17,45]) 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. [7,5] 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 [6] and further developed in [24,26,27,28], 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 [26,27,28] in the operational framework.

- *Linear estimates.* Since their introduction in [31,32], *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., [44,8,11,9,20,50,53]). 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.* The idea of a *polyhedral estimate* goes back to [43] where it was shown (see also [41, 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 [14] and [45] and extended to the setting to follow in [27]. In this setting, 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

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

Our interest in these estimates stems from the results of [25,27,28] 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 subjected to random uncertainty. Specifically, we assume that

$$\omega = A[\eta]x + \xi, \tag{2}$$

where $\xi \in \mathbf{R}^m$ is zero mean random noise and

$$A[\eta] = A + \sum\nolimits_{\alpha=1}^{q} \eta_\alpha A_\alpha \in \mathbf{R}^{m\times n}, \tag{3}$$

where $A, A_1, ..., A_q$ are given matrices and $\eta \in \mathbf{R}^q$ is unknown random perturbation ("random uncertainty"). Observation model (2) with random uncertainty is related to the linear regression problem with random errors in regressors [1,2,12,13,33,48,51] 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., [3,4,10,16,17,38]).

In what follows, our goal is to extend the estimation constructions from [28] 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.

- The ϵ-risk (the maximum, over signals from $\mathcal{X}$, of the radii of $(1-\epsilon)$-confidence $\|\cdot\|$-balls) is analyzed in Section 2. This leads to novel computationally efficient techniques of design of presumably good, in terms of this risk, linear estimates under random perturbation of observation matrix.
- *Analysis problem* for polyhedral estimate

  Given contrast matrix H, find a provably tight efficiently computable upper bound on ϵ-risk of the associated estimate

  is the subject of Section 3.3, where it is solved "in the full range" of our assumptions (ellitopic $\mathcal{X}$, sub-Gaussian zero mean η and ξ). In contrast, 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 uncertainty in sensing matrix is present, the strategy to circumvent this difficulty developed in [28, 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 3.4–3.6.

[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 $\{x : \|Px\|_p \leq 1\}$, $p \geq 2$, 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.

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

$$\|u\| = \max_{\ell \leq L} \sqrt{u^T R_\ell u},$$

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 and uncertainty η are zero-mean sub-Gaussian: $\xi \sim \mathcal{SG}(0, \sigma^2 I)$ and $\eta \sim \mathcal{SG}(0, I)$, i.e.,

$$\mathbf{E}\left\{e^{t^T\xi}\right\} \leq \exp\left(\frac{\sigma^2}{2}\|t\|_2^2\right) \qquad \forall t \in \mathbf{R}^m, \tag{4}$$

$$\mathbf{E}\left\{e^{t^T\eta}\right\} \leq \exp\left(\frac{1}{2}\|t\|_2^2\right) \qquad \forall t \in \mathbf{R}^q. \tag{5}$$

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*

$$\mathrm{Risk}_\epsilon[\widehat{w}|\mathcal{X}] := \sup_{x\in\mathcal{X}} \inf\left\{\rho :\ \mathrm{Prob}_{\xi,\eta}\{\|Bx - \widehat{w}(A[\eta]x + \xi)\| > \rho\} \leq \epsilon\right\} \tag{6}$$

(the radius of the smallest $\|\cdot\|$-ball centered at $\widehat{w}(\omega)$ which covers Bx with probability $\geq 1 - \epsilon$, uniformly over $x \in \mathcal{X}$).

2 Design of Presumably Good Linear Estimate

2.1 Preliminaries: Ellitopes

Throughout this section, we assume that the signal set $\mathcal{X}$ is *a basic ellitope*. Recall that, by definition [25,28], a basic ellitope in $\mathbf{R}^n$ is a set of the form

$$\mathcal{X} = \{x \in \mathbf{R}^n :\ \exists t \in \mathcal{T} :\ x^T T_k x \leq t_k,\ k \leq K\}, \tag{7}$$

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 $0 \leq t' \leq t \in \mathcal{T}$ one has $t' \in \mathcal{T}$.

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 $\{x \in \mathbf{R}^n : x^T T_k x \leq 1\}$ $[T_k \succeq 0]$ is a basic ellitope:

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

In particular, the unit box $\{x \in \mathbf{R}^n : \|x\|_\infty \leq 1\}$ is a basic ellitope.
B. A $\|\cdot\|_p$-ball in $\mathbf{R}^n$ with $p \in [2, \infty]$ is a basic ellitope:

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

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

$$\mathrm{Opt}_*(C) = \max_{x\in\mathcal{X}} x^T Cx,\ C \in \mathbf{S}^n \tag{8}$$

of maximizing a homogeneous quadratic form over $\mathcal{X}$. As it is shown in [28], when $\mathcal{X}$ is an ellitope, (8) admits "reasonably tight" efficiently computable upper bound. Specifically,

Theorem 2.1 [28, Proposition 4.6] *Given ellitope* (7) *and matrix* C*, consider the quadratic maximization problem* (8) *along with its relaxation*

$$\mathrm{Opt}(C) = \min_{\lambda}\left\{\phi_{\mathcal{T}}(\lambda) : \lambda \geq 0, \sum\nolimits_k \lambda_k T_k - C \succeq 0\right\}$$

(here and in what follows, $\phi_S(t) = \sup_{s\in S} t^T s$ *stands for the support function of nonempty set* $S \subset \mathbf{R}^N$*). 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:*

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

2.2 Tight Upper Bounding the Risk of Linear Estimate

Consider a linear estimate

$$\widehat{w}^H(\omega) = H^T\omega \quad [H \in \mathbf{R}^{m\times\nu}]$$

Proposition 2.1 *In the setting of this section, synthesis of a presumably good linear estimate reduces to solving the convex optimization problem*

$$\min_{H\in\mathbf{R}^{m\times\nu}} \mathfrak{R}[H] \tag{9}$$

where

$$\mathfrak{R}[H] = \min_{\substack{\lambda_\ell,\mu^\ell,\kappa^\ell,\\ \varkappa^\ell,\rho,\varrho}} \left\{ \left[1+\sqrt{2\ln(2L/\epsilon)}\right]\left[\sigma\max_{\ell\leq L}\|HR_\ell^{1/2}\|_{\mathrm{Fro}}+\rho\right]+\varrho : \right. \tag{10}$$
$$\left.\begin{array}{l}
\mu^\ell \geq 0, \varkappa^\ell \geq 0,\ \lambda_\ell + \phi_{\mathcal{T}}(\mu_\ell) \leq \rho, \kappa_\ell + \phi_{\mathcal{T}}(\varkappa^\ell) \leq \varrho,\ \ell \leq L \\
\left[\begin{array}{c|c}\lambda_\ell I_{\nu q} & \frac{1}{2}\left[R_\ell^{1/2}H^TA_1;\ldots;R_\ell^{1/2}H^TA_q\right] \\ \hline \frac{1}{2}\left[A_1^THR_\ell^{1/2},\ldots,A_q^THR_\ell^{1/2}\right] & \sum_k \mu_k^\ell T_k\end{array}\right] \succeq 0,\ \ell \leq L \\
\left[\begin{array}{c|c}\kappa^\ell I_\nu & \frac{1}{2}R_\ell^{1/2}[B-H^TA] \\ \hline \frac{1}{2}[B-H^TA]^TR_\ell^{1/2} & \sum_k \varkappa_k^\ell T_k\end{array}\right] \succeq 0,\ \ell \leq L
\end{array}\right\}.$$

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* $\mathfrak{R}[H]$*.*

When the ellitope $\mathcal{X}$—as simple as possible—is the unit $\|\cdot\|_2$-ball, the right hand side of (10) can be computed in closed analytic form, resulting in

$$\mathfrak{R}[H] = \left[1+\sqrt{2\ln(2L/\epsilon)}\right]\left[\sigma \max_{\ell\le L}\|HR_\ell^{1/2}\|_{\rm Fro} + \max_{\ell\le L}\left\|\begin{bmatrix} R_\ell^{1/2}H^TA_1\\ \cdots \\ R_\ell^{1/2}H^TA_q\end{bmatrix}\right\|_{2,2}\right]$$
$$+\max_{\ell\le L}\|R_\ell^{1/2}[B-H^TA]\|_{2,2},$$

where $\|\cdot\|_{2,2}$ stands for the spectral norm of a matrix.

2.3 A Modification

Let us assume that an Υ-repeated version of observation (2) is available, i.e., we observe

$$\omega^\Upsilon = \{\omega_k = A[\eta_k]x + \xi_k,\ k=1,...,\Upsilon\} \tag{11}$$

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

$$\mathbf{E}\{\xi\xi^T\}\preceq\sigma^2 I_m,\quad \mathbf{E}\left\{\eta\eta^T\right\}\preceq I_q. \tag{12}$$

Let us consider the following construction. For each $\ell\le L$, given $H\in\mathbf{R}^{m\times\nu}$ we denote

$$\widetilde{\mathfrak{R}}_\ell[H] = \min_{\lambda,\mu,\kappa,\varkappa}\Bigg\{\sigma\|HR_\ell^{1/2}\|_{\rm Fro}+\lambda+\phi_{\mathcal{T}}(\mu)+\kappa+\phi_{\mathcal{T}}(\varkappa): \tag{13}$$
$$\mu\ge 0,\varkappa\ge 0, \left[\begin{array}{c|c}\kappa I_\nu & \frac12 R_\ell^{1/2}[B-H^TA]\\ \hline \frac12[B-H^TA]^TR_\ell^{1/2} & \sum_k\varkappa_kT_k\end{array}\right]\succeq 0$$
$$\left[\begin{array}{c|c}\lambda I_{\nu q} & \frac12\left[R_\ell^{1/2}H^TA_1;...;R_\ell^{1/2}H^TA_q\right]\\ \hline \frac12\left[A_1^THR_\ell^{1/2},...,A_q^THR_\ell^{1/2}\right] & \sum_k\mu_kT_k\end{array}\right]\succeq 0\Bigg\}$$

and consider the convex optimization problem

$$\widetilde{H}_\ell\in\operatorname*{Argmin}_H\widetilde{\mathfrak{R}}_\ell[H].$$

We define the "reliable estimate" $\widehat{w}^{(r)}(\omega^\Upsilon)$ of $w=Bx$ as follows.

1. Given $\widetilde{H}_\ell\in\mathbf{R}^{m\times\nu}$ and observations ω_k we compute linear estimates $w_\ell(\omega_k)=\widetilde{H}_\ell\omega_k,\ \ell=1,...,L,\ k=1,...,\Upsilon$;
2. We define vectors $z_\ell\in\mathbf{R}^\nu$ as geometric medians of $w_\ell(\omega_k)$:

$$z_\ell(\omega^\Upsilon)\in\operatorname*{Argmin}_z\sum_{k=1}^{\Upsilon}\|R_\ell^{1/2}(w_\ell(\omega_k)-z)\|_2,\ \ell=1,...,L.$$

3. Finally, we select as $\widehat{w}^{(r)}(\omega^\Upsilon)$ any point of the set

$$\mathcal{W}(\omega^\Upsilon) = \bigcap_{\ell=1}^{L} \left\{ w \in \mathbf{R}^\nu : \|R_\ell^{1/2}(z_\ell(\omega^\Upsilon) - w)\|_2 \leq 4\widetilde{\mathfrak{R}}_\ell[\widetilde{H}_\ell] \right\}.$$

   or set $\widehat{w}^{(r)}(\omega^\Upsilon) = 0$ if $\mathcal{W}(\omega^\Upsilon) = \emptyset$.

We have the following analog of Proposition 2.1.

Proposition 2.2 *In the situation of this section, it holds*

$$\sup_{x\in\mathcal{X}} \mathbf{E}_{\eta_k,\xi_k} \left\{ \|R_\ell^{1/2}(w_\ell(\omega_k) - Bx)\|_2^2 \right\} \leq \widetilde{\mathfrak{R}}_\ell^2[\widetilde{H}_\ell],\ \ell \leq L \tag{14}$$

and

$$\mathrm{Prob}\left\{ \|R_\ell^{1/2}(z_\ell(\omega^\Upsilon) - Bx)\|_2 \geq 4\widetilde{\mathfrak{R}}_\ell[\widetilde{H}_\ell] \right\} \leq e^{-0.1070\Upsilon},\ \ell \leq L. \tag{15}$$

As a consequence, whenever $\Upsilon \geq \ln[L/\epsilon]/0.1070$*, the* ϵ*-risk of the aggregated estimate* $\widehat{w}^{(r)}(\omega^\Upsilon)$ *satisfies*

$$\mathrm{Risk}_\epsilon[\widehat{w}^{(r)}(\omega^\Upsilon)|\mathcal{X}] \leq \overline{\mathfrak{R}},\ \ \overline{\mathfrak{R}} = 8\max_{\ell\leq L} \widetilde{\mathfrak{R}}_\ell[\widetilde{H}_\ell].$$

Remark. Proposition 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 (2) in this situation reads

$$\omega = A[\eta]x + \xi,\ A[\eta] = (1-\gamma)\overline{A} + \sum\nolimits_{\alpha=1}^{n} \eta_\alpha A_\alpha$$

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, $1 \leq \alpha \leq n$, 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 2.1 or Proposition 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 (9), (10) responsible for the design of the linear estimate.

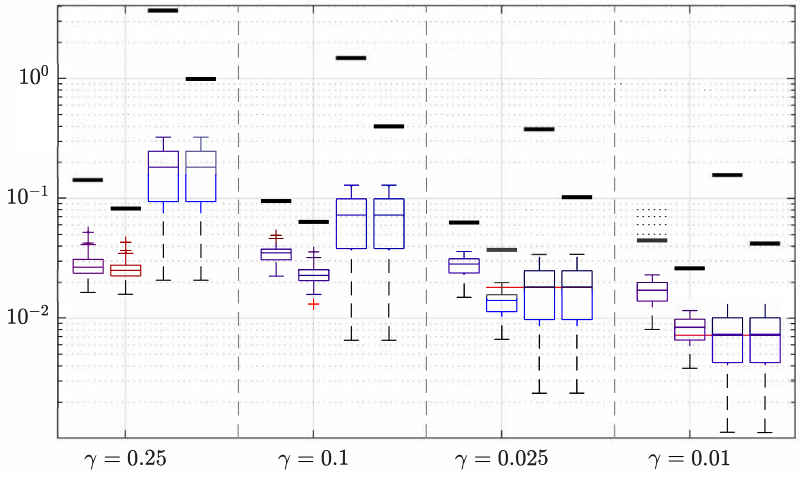

Fig. 1 Distributions of ℓ_2-recovery errors and upper bounds of the robust and "nominal" estimates for different values of γ parameter.

2.4 Numerical Illustration

In Figure 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 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 (13), and upper bound $\widetilde{\mathfrak{R}}_1[H_1]$ on its *expected error risk*,[2] (3) "nominal" estimate—estimate by Proposition 2.1 as applied to the "no uncertainty" case where all A_α in (3) are set to 0 and upper bound $\mathfrak{R}$ from (10) 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

[2] We define expected error risk of a Υ-observation estimate $\widehat{x}(\omega^\Upsilon)$ of Bx as $\sup_{x\in\mathcal{X}} \mathbf{E}_{\omega^\Upsilon \sim P_x^\Upsilon}\{\|\widehat{x}(\omega^\Upsilon) - Bx\|\}$, where P_x^Υ is the distribution of ω^Υ stemming from x.

in the "no uncertainty" case and upper bound $\widetilde{\mathfrak{R}}_1[\widetilde{H}_1]$ on its "actual"—with uncertainty present—expected error risk.

3 Design of Presumably Good Polyhedral Estimate

3.1 Preliminaries on Polyhedral Estimates

Consider a slightly more general than (2), (3) observation scheme

$$\omega = Ax + \zeta,$$

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 (7), and ζ is observation noise with probability distribution P_x which can depend on x. For example, when observation ω is given by (2), (3), we have

$$\zeta = \sum\nolimits_{\alpha=1}^{q} \eta_\alpha A_\alpha x + \xi$$

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

$$\mathrm{Prob}_{\zeta\sim P_x}\{|h^T\zeta| > 1\} \le \epsilon/M \quad \forall x \in \mathcal{X}. \tag{16}$$

A polyhedral estimate $\widehat{w}^H(\cdot)$ is specified by *contrast matrix* $H \in \mathbf{R}^{m\times M}$ restricted to have all columns in $\mathcal{H}$ according to

$$\omega \mapsto \widehat{x}^H(\omega) \in \mathop{\mathrm{Argmin}}_{u\in\mathcal{X}} \left\{\|H^T[Au-\omega]\|_\infty\right\},\ \widehat{w}^H_{\mathrm{poly}}(\omega) = B\widehat{x}^H(\omega). \tag{17}$$

It is easily seen (cf. [28, Proposition 5.1.1]) that the ϵ-risk (6) of the above estimate is upper-bounded by the quantity

$$\mathfrak{p}[H] = \sup_y \left\{\|By\| : y \in 2\mathcal{X}, \|H^TAy\|_\infty \le 2\right\}. \tag{18}$$

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 $\|H^TA[x-\widehat{x}^H(\omega)]\|_\infty \le 2$, and besides this, $x - \widehat{x}^H(\omega) \in 2\mathcal{X}$, whence $\|Bx - \widehat{w}^H_{\mathrm{poly}}(\omega)\| \le \mathfrak{p}[H]$ 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 $\|Bx - \widehat{w}^H_{\mathrm{poly}}(\omega)\| \le \mathfrak{p}[H]$, whence the ϵ-risk of the estimate $\widehat{w}^H_{\mathrm{poly}}$ 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 [27] goes via minimizing an efficiently computable upper bound on $\mathfrak{p}[H]$. It is shown in [28, Section 5.1.5] that when $\mathcal{X}$ is ellitope (7) and $\|u\|=\|Ru\|_2$, a reasonably tight upper bound on $\mathfrak{p}[H]$ is given by the efficiently computable function

$$\mathfrak{p}_+[H]=2\min_{\lambda,\mu,\upsilon}\Big\{\lambda+\phi_{\mathcal{T}}(\mu)+\sum\nolimits_i\upsilon_i:\ \mu\geq 0,\upsilon\geq 0$$
$$\left[\begin{array}{c|c}\lambda I_\nu & \frac{1}{2}RB\\ \hline \frac{1}{2}B^TR^T & A^TH\mathrm{Diag}\{\upsilon\}H^TA+\sum_k\mu_kT_k\end{array}\right]\succeq 0\Big\}.$$

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 is given by (2), (3).

3.2 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

$$h\in\mathcal{H}_\delta, x\in\mathcal{X}\Rightarrow \mathrm{Prob}_{\zeta\sim P_x}\{|h^T\zeta|>1\}\leq\delta. \tag{19}$$

Note that a "tight" sufficient condition for the validity of (19) is

$$\mathrm{Prob}_\xi\{|h^T\xi|>1/2\}\leq\delta/2, \tag{20a}$$
$$\mathrm{Prob}_\eta\left\{\left|\sum\nolimits_{\alpha=1}^q[h^TA_\alpha x]\eta_\alpha\right|>1/2\right\}\leq\delta/2,\ \forall x\in\mathcal{X}. \tag{20b}$$

Note that under the sub-Gaussian assumption (4), $h^T\xi$ is itself sub-Gaussian, $h^T\xi\sim\mathcal{SG}(0,\sigma^2\|h\|_2^2)$; thus, a tight sufficient condition for (20a) is

$$\|h\|_2\leq[\sigma\chi(\delta)]^{-1},\ \ \chi(\delta)=2\sqrt{2\ln(2/\delta)}. \tag{21}$$

Furthermore, by (5), 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 (20b) for a given x whenever

$$\|[h^TA_1x;...;h^TA_qx]\|_2\leq\chi^{-1}(\delta).$$

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

$$x\mapsto\mathcal{A}[h]x,\ \ \mathcal{A}[h]=[h^TA_1;h^TA_2;...;h^TA_q]$$

induced by the norm $\|\cdot\|_{\mathcal{X}}$ on the argument and the norm $\|\cdot\|_2$ on the image space to be upper-bounded by $\chi^{-1}(\delta)$:

$$\|\mathcal{A}[h]\|_{\mathcal{X},2} \leq \chi^{-1}(\delta). \tag{22}$$

Invoking [23, Theorem 3.1] (cf. also the derivation in the proof of Proposition 2.1 in Section A.1.2), a tight sufficient condition for the latter relation is

$$\mathrm{Opt}[h] := \min_{\lambda,\mu}\left\{\lambda + \phi_{\mathcal{T}}(\mu) : \begin{array}{c}\mu \geq 0,\\ \left[\begin{array}{c|c}\lambda I_q & \frac{1}{2}\mathcal{A}[h]\\ \hline \frac{1}{2}\mathcal{A}^T[h] & \sum_k \mu_k T_k\end{array}\right] \succeq 0\end{array}\right\} \leq \chi^{-1}(\delta), \tag{23}$$

tightness meaning that Opt$[h]$ is within factor $O(1)\sqrt{\ln(K+1)}$ of $\|\mathcal{A}[h]\|_{\mathcal{X},2}$. $\|\overline{A}[h]\|_{\mathcal{X},2} \leq \chi(\delta)$.

The bottom line is that with $\mathcal{H}_\delta$ specified by constraints (21) and (22) (or by the latter replaced with its tight relaxation (23)) we do ensure (19).

3.3 Bounding the Risk of the Polyhedral Estimate $\widehat{w}^H$

Proposition 3.1 *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 $j \leq ML$ and $\delta = \epsilon/ML$. Consider optimization problem*

$$\begin{array}{l}\mathfrak{p}_+[H] = 2 \min\limits_{\lambda_\ell,\mu^\ell,\upsilon^\ell,\rho}\Bigg\{\rho : \mu^\ell \geq 0, \upsilon^\ell \geq 0,\ \lambda_\ell + \phi_{\mathcal{T}}(\mu^\ell) + \sum_{j=1}^M \upsilon_j^\ell \leq \rho,\ \ell \leq L\\ \qquad\left[\begin{array}{c|c}\lambda_\ell I_\nu & \frac{1}{2}R_\ell^{1/2}B\\ \hline \frac{1}{2}B^T R_\ell^{1/2} & A^T H_\ell \mathrm{Diag}\{\upsilon^\ell\} H_\ell^T A + \sum_k \mu_k^\ell T_k\end{array}\right] \succeq 0,\ \ell \leq L\Bigg\}.\end{array} \tag{24}$$

Then $\mathrm{Risk}_\epsilon[\widehat{w}^H|\mathcal{X}] \leq \mathfrak{p}_+[H]$.

3.4 Optimizing $\mathfrak{p}_+[H]$—the Strategy

Proposition 3.1 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 (24) results in a nonconvex problem. A remedy, as proposed in [28, 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 $\varkappa \geq 1$ 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 $\varkappa \geq 1$, we can efficiently build collections of vectors $h_j \in \mathcal{H}$, and reals $\upsilon_j \geq 0$, $j \leq J$, such that $\Theta = \sum_{j=1}^{J} \upsilon_j h_j h_j^T$ and $\sum_j \upsilon_j \leq \varkappa \varrho$.

Example. Let $\mathcal{H}$ be a centered at the origin Euclidean ball of radius $R > 0$ in $\mathbf{R}^J$. When setting

$$\mathbf{K} = \{(\Theta, \varrho) : \Theta \succeq 0, \mathrm{Tr}(\Theta) \leq R^2 \varrho\},$$

we obtain a cone $(M, 1)$-compatible with $\mathcal{H}$. Indeed, for $h_j \in \mathcal{H}$ and $\upsilon_j \geq 0$ we have

$$\mathrm{Tr}\left(\sum_j \upsilon_j h_j h_j^T\right) \leq R^2 \sum_j \upsilon_j,$$

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 $\varrho \geq \mathrm{Tr}(\Theta)/R^2$ and specifying $f_1, ..., f_m$ as the orthonormal system of eigenvectors of Θ, and λ_j as the corresponding eigenvalues and setting $h_j = R f_j$, $\upsilon_j = R^{-2}\lambda_j$), we get $h_j \in \mathcal{H}$, $\Theta = \sum_j \upsilon_j h_j h_j^T$ and $\sum_j \upsilon_j = \mathrm{Tr}(\Theta)/R^2 \leq \varrho$.

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

$$\min_{H=[H^1,\ldots,H^L]} \{\mathfrak{p}_+[H] : \mathrm{Col}_j[H^\ell]_j \in \mathcal{H}_\delta\} \tag{25}$$

with the problem

$$\begin{array}{l}\min\limits_{\substack{\bar\lambda_\ell, \bar\mu^\ell,\\ \Theta_\ell, \varrho_\ell, \bar\rho}} \Big\{\bar\rho : (\Theta_\ell, \varrho_\ell) \in \mathbf{K}, \bar\mu^\ell \geq 0,\ \bar\lambda_\ell + \phi_{\mathcal{T}}(\bar\mu^\ell) + \varrho_\ell \leq \bar\rho,\ \ell \leq L, \\ \qquad \left[\begin{array}{c|c}\bar\lambda_\ell I_\nu & \frac{1}{2} R_\ell^{1/2} B \\ \hline \frac{1}{2} B^T R_\ell^{1/2} & A^T \Theta_\ell A + \sum_k \bar\mu_k^\ell T_k\end{array}\right] \succeq 0, \ell \leq L\Big\}.\end{array} \tag{26}$$

Unlike (25), 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 (25). Namely,

– Given a feasible solution $H_\ell, \lambda_\ell, \mu^\ell, \upsilon^\ell, \rho$ to the problem of interest (25), 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 (26), 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 (26) 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 $\mathcal{H}_\delta$, $\Theta_\ell = H_\ell \mathrm{Diag}\{\bar\upsilon^\ell\} H_\ell^T$,

$\sum_j \bar{\upsilon}_j^\ell \leq \varkappa\varrho_\ell$. 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 (25). Indeed, all we need is to verify that this solution satisfies, for every $\ell \leq L$, constraints of (24). To check the semidefinite constraint, note that

$$\left[\begin{array}{c|c} \lambda_\ell I_\nu & \frac{1}{2}R_\ell^{1/2}B \\ \hline \frac{1}{2}B^TR_\ell^{1/2} & A^TH_\ell\mathrm{Diag}\{\upsilon^\ell\}H_\ell^TA + \sum_k \mu_k^\ell T_k \end{array}\right]$$
$$= \left[\begin{array}{c|c} \gamma_\ell\bar{\lambda}_\ell I_\nu & \frac{1}{2}R_\ell^{1/2}B \\ \hline \frac{1}{2}B^TR_\ell^{1/2} & \gamma_\ell^{-1}\left[A^TH_\ell\mathrm{Diag}\{\bar{\upsilon}^\ell\}H_\ell^TA + \sum_k \bar{\mu}_k^\ell T_k\right] \end{array}\right],$$

and the matrix in the right-hand side is $\succeq 0$ by the semidefinite constraint of (26) combined with $\Theta_\ell = \sum_j \bar{\upsilon}_j^\ell \mathrm{Col}_j[H_\ell]\mathrm{Col}_j^T[H_\ell]$. Furthermore, note that by construction $\sum_j \bar{\upsilon}_j^\ell \leq \varkappa\varrho_\ell$, whence

$$\lambda_\ell + \phi_{\mathcal{T}}(\mu^\ell) + \sum\nolimits_j \upsilon_j^\ell = \gamma_\ell\bar{\lambda}_\ell + \gamma_\ell^{-1}[\phi_{\mathcal{T}}(\bar{\mu}^\ell) + \varkappa\varrho_\ell] = 2\sqrt{\bar{\lambda}_\ell[\phi_{\mathcal{T}}(\bar{\mu}^\ell) + \varkappa\varrho_\ell]}$$
$$\leq 2\sqrt{\varkappa}\sqrt{\bar{\lambda}_\ell[\phi_{\mathcal{T}}(\bar{\mu}^\ell) + \varrho_\ell]} \leq \sqrt{\varkappa}\left[\bar{\lambda}_\ell + \phi_{\mathcal{T}}(\bar{\mu}^\ell) + \varrho_\ell\right] \leq \sqrt{\varkappa}\bar{\rho} = \rho$$

(we have taken into account that $\varkappa \geq 1$).

We conclude that the (efficiently computable) optimal solution to the relaxed problem (26) can be efficiently converted to a feasible solution to problem (25) 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 (25) 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 (25) associated with $\mathcal{H} = \mathcal{H}_\delta$ given by (21) and (23). 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 [28, 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 [29]. 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 [29] from ellitopes to *spectratopes*.

3.5 Estimate Synthesis Utilizing Cones Compatible with Spectratopes

Let for $S^{ij} \in \mathbf{S}^{d_i}$, $1 \leq i \leq I$, $1 \leq j \leq N$, 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

$$\mathcal{H} = \{g \in \mathbf{R}^N : \exists r \in \mathcal{R} : S_i^2[g] \preceq r_i I_{d_i}, i \leq I\}; \tag{27}$$

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 $g \neq 0$. A spectratope, by definition, is a linear image of a basic spectratope.

As shown in [28], 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 (21) and (22) 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 (7), so that $\mathcal{X}$ is the unit Euclidean ball,

$$\mathcal{X} = \{x \in \mathbf{R}^n : x^T x \leq 1\}.$$

In this situation, denoting by $\|\cdot\|_{2,2}$ the spectral norm of a matrix, constraints (21) and (22) specify the set

$$\begin{aligned} \mathcal{H}_\delta &= \left\{h \in \mathbf{R}^m : \|h\|_2 \leq (\sigma\chi(\delta))^{-1}, \|\mathcal{A}[h]\|_{2,2} \leq \chi^{-1}(\delta)\right\} \\ &= \left\{h \in \mathbf{R}^m : \exists r \in \mathcal{R} : S_j^2[h] \preceq r_j I_{d_j}, j \leq 2\right\}, \end{aligned} \tag{28}$$

where $\mathcal{R} = \{[r_1; r_2] : 0 \leq r_1, r_2 \leq 1\}$,

$$S_1[h] = \sigma\chi(\delta) \left[\begin{array}{c|c} & h \\ \hline h^T & \end{array}\right] \in \mathbf{S}^{m+1},\; S_2[h] = \chi(\delta) \left[\begin{array}{c|c} & \mathcal{A}[h] \\ \hline \mathcal{A}[h]^T & \end{array}\right] \in \mathbf{S}^{m+q}$$

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 (27), 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

$$G \preceq G' \Rightarrow \mathcal{S}_i[G] \preceq \mathcal{S}_i[G'], \tag{29a}$$

$$\{G \succeq 0 \ \&\ \mathcal{S}_i[G] = 0\,\forall \ell\} \Rightarrow G = 0. \tag{29b}$$

A cone "tightly compatible" with a basic spectratope is given by the following

Proposition 3.2 *Let $\mathcal{H} \subset \mathbf{R}^N$ be a basic spectratope*

$$\mathcal{H} = \{g \in \mathbf{R}^N : \exists r \in \mathcal{R} : S_i^2[g] \preceq r_i I_{d_i},\ i \leq I\}$$

with "spectratopic data" $\mathcal{R}$ and $S_i[\cdot]$, $i \leq I$, satisfying the requirements in the above definition.

Let us specify the closed convex cone $\mathbf{K} \subset \mathbf{S}_+^N \times \mathbf{R}_+$ as

$$\mathbf{K} = \left\{(\Sigma, \rho) \in \mathbf{S}_+^N \times \mathbf{R}_+ : \exists r \in \mathcal{R} : \mathcal{S}_i[\Sigma] \preceq \rho r_i I_{d_i}, i \leq I\right\}.$$

Then

(i) *whenever $\Sigma = \sum_j \lambda_j g_j g_j^T$ with $\lambda_j \geq 0$ 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) $\lambda_j \geq 0$ and g_j, $j \leq N$, such that*

$$\Sigma = \sum\nolimits_j \lambda_j g_j g_j^T \ \text{ with } \sum\nolimits_j \lambda_j \leq \varkappa\rho \ \text{ and } \ g_j \in \mathcal{H},\ j \leq N.$$

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 A.2.2 of the appendix.

3.6 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 3.4, 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

$$\mathrm{Opt} = \min_{\substack{\Theta_\ell \in \mathbf{S}^m,\\ \varrho_\ell, \bar{\lambda}_\ell, \bar{\mu}_\ell}} \left\{ \bar{\rho} : \bar{\mu}_\ell \geq 0,\ \bar{\lambda}_\ell + \bar{\mu}_\ell + \varrho_\ell \leq \bar{\rho},\ \Theta_\ell \succeq 0,\ \ell \leq L, \right. \tag{30}$$
$$\left. \begin{array}{l} \left[\begin{array}{c|c} \left[\mathrm{Tr}(A_\alpha^T \Theta_\ell A_\beta)\right]_{\alpha,\beta=1}^q & \\ \hline & \sum_{\alpha,\beta} A_\alpha^T \Theta_\ell A_\beta \end{array}\right] \preceq \chi^{-2}(\delta)\varrho_\ell I_{q+n},\ \ell \leq L, \\ \left[\begin{array}{c|c} \bar{\lambda}_\ell I_\nu & \frac{1}{2} R_\ell^{1/2} B \\ \hline \frac{1}{2} B^T R_\ell^{1/2} & A^T \Theta_\ell A + \bar{\mu}_\ell I_n \end{array}\right] \succeq 0,\ \sigma^2 \chi^2(\delta) \mathrm{Tr}(\Theta_\ell) \leq \varrho_\ell,\ \ell \leq L \end{array} \right\}$$

—this is what under the circumstances becomes problem (26) with the cone **K** given by Proposition 3.2 as applied to the spectratope $\mathcal{H}_\delta$ given by (28). Note that by Proposition 3.2, **K** is $\varkappa$-compatible with $\mathcal{H}_\delta$, with

$$\varkappa = 4 \ln(4m(m + n + q + 1)).$$

For instance, in the case of rank 1 matrices $A_\alpha = f_\alpha g_\alpha^T$ and $\|\cdot\| = \|\cdot\|_2$ (30) becomes

$$\mathrm{Opt} = \min_{\substack{\Theta \in \mathbf{S}^m,\\ \varrho, \bar{\lambda}, \bar{\mu}}} \left\{ \bar{\rho} : \bar{\mu} \geq 0,\ \Theta \succeq 0,\ \sigma^2 \chi^2(\delta) \mathrm{Tr}(\Theta) \leq \varrho,\ \bar{\lambda} + \bar{\mu} + \varrho \leq \bar{\rho} \right.$$
$$\left. \begin{array}{l} \left[\begin{array}{c|c} \left[(f_\alpha^T \Theta f_\beta) g_\alpha^T g_\beta\right]_{\alpha,\beta=1}^q & \\ \hline & \left[\sum_{\alpha,\beta=1}^q [f_\alpha^T \Theta f_\beta]\, g_\alpha g_\beta^T\right] \end{array}\right] \preceq \chi^{-2}(\delta) \varrho I_{q+n} \\ \left[\begin{array}{c|c} \bar{\lambda}_\ell I_\nu & \frac{1}{2} B \\ \hline \frac{1}{2} B^T & A^T \Theta A + \bar{\mu} I_n \end{array}\right] \succeq 0 \end{array} \right\} \tag{31}$$

• use the randomized algorithm described in the proof of Proposition 3.2 to convert the Θ_ℓ-components of the optimal solution to (30) 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 (21), (22), 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 (25), and the ϵ-risk of the associated polyhedral estimate is upper-bounded by $2\sqrt{\varkappa}\mathrm{Opt}$ with Opt given by (30). In Figure 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 2.3. In the plot cells, from left to right: (1) robust linear estimate by Proposition 2.1 and upper bound $\mathfrak{R}$ on its 0.05-risk, (2) robust

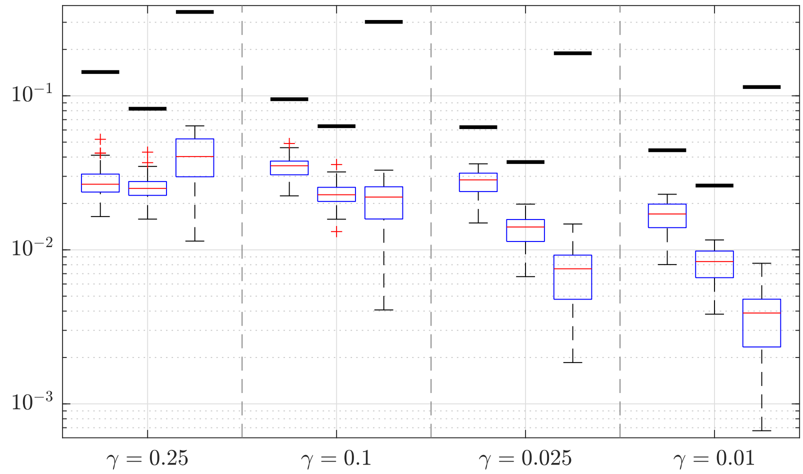

Fig. 2 Distributions of ℓ_2-recovery errors and upper bounds of the robust linear and robust polyhedral estimates for different values of γ parameter.

linear estimate $w_1(\omega_1)$ yielded by Proposition 2.2 and upper bound $\widetilde{\mathfrak{R}}_1$ on its expected error risk, (3) robust polyhedral estimate by Proposition 3.2 and upper bound on its 0.05-risk.

3.7 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 2.3, we are about to show that passing from observation (2) to its K-repeated, with "moderate" K, version (cf. (11))

$$\omega^K = \{\omega_k = A[\eta_k]x + \xi_k, \quad k = 1, ..., K\}$$

with pairs (η_k, ξ_k) independent across k, we can relax the sub-Gaussianity assumption replacing it with moment condition (12). Specifically, let us set

$$\mathcal{H} = \left\{h \in \mathbf{R}^m : \sigma\|h\|_2 \le \tfrac{1}{8}, \|\mathcal{A}[h]\|_{\mathcal{X},2} \le \tfrac{1}{8}\right\}, \ \mathcal{A}[h]x = [h^T A_1; ...h^T A_q]$$

(cf. (21) and (22)).

Given tolerance an $m \times M$ contrast matrix H with columns $h_j \in \mathcal{H}$, and observation (11), we build the polyhedral estimate as follows.[3]

1. For $j = 1, ..., M$ we compute empirical medians y_j of the data $h_j^T \omega_k$, $k = 1, ..., K$,
$$y_j = \text{median}\{h_j^T \omega_k, \ 1 \le k \le K\}.$$

[3] Readers acquainted with the literature on robust estimation will immediately recognize that the proposed construction is nothing but a reformulation of the "median-of-means" estimate [42] (see also [36, 19, 39, 35]) for our purposes.

2. We specify $\widehat{x}^H(\omega^K)$ as a point from $\mathrm{Argmin}_{u\in\mathcal{X}}\|y-H^TAu\|_\infty$ and use, as the estimate of Bx, the vector $\widehat{w}^H_{\rm poly}(\omega^K)=B\widehat{x}^H(\omega^K)$.

Lemma 3.1 *In the situation of this section, let ξ_k and η_k satisfy moment constraint of (12), and let $K\geq\overline{\kappa}=2.5\ln[M/\epsilon]$. Then estimate $\widehat{w}^H_{\rm poly}(\omega^K)$ satisfies (cf. (18))*

$$\mathrm{Risk}_\epsilon[\widehat{w}^H_{\rm poly}(\omega^K)|\mathcal{X}]\leq\mathfrak{p}[H].$$

As an immediate consequence of the result of Lemma 3.1, the constructions and results of Sections 3.3–3.6 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 (12) 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.

4 Conclusions

In this paper, we develop an approach to the design and performance analysis of two classes of statistical techniques—linear and polyhedral estimates—for recovering signal x from noisy observations of its linear image Ax. We assume that a priori information about the signal localizes it in a known in advance and "well-structured" convex set belonging to a family rich enough to cover a wide spectrum of potential applications—the ellitopes and the spectratopes. We focus on the situation where the sensing matrix A is affected by random perturbations. Assuming that these perturbations and observation noises are sub-Gaussian, we develop computationally efficient routines of design of "presumably good" linear and polyhedral estimates and evaluating their performance.

A natural alternative to the model of observation subject to "stochastic uncertainty" in the sensing matrix is the model in which perturbations of A are uncertain-but-bounded; this alternative is the subject of our forthcoming paper.

Acknowledgements This work was supported by Multidisciplinary Institute in Artificial intelligence MIAI @ Grenoble Alpes (ANR-19-P3IA-0003). Data sets generated during the current study are available from the corresponding author on reasonable request.

A Proofs

A.1 Proofs for Section 2

A.1.1 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., [18,46,47]).

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:

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

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 $\xi \sim \mathcal{N}(h, S^T S)$ one has

$$\mathbf{E}_\eta\left\{e^{h^T\eta + \frac{1}{2}\eta^T S^T S\eta}\right\} = \mathbf{E}_\eta\left\{\mathbf{E}_\xi\left\{e^{\eta^T\xi}\right\}\right\} = \mathbf{E}_\xi\left\{\mathbf{E}_\eta\left\{e^{\eta^T\xi}\right\}\right\} \le \mathbf{E}_\xi\left\{e^{\frac{1}{2}\xi^T\Sigma\xi}\right\},$$

so that

$$\ln\left(\mathbf{E}_\eta\left\{e^{h^T\eta + \frac{1}{2}\eta^T S^T S\eta}\right\}\right) \le \ln\left(\mathbf{E}_\xi\left\{e^{\frac{1}{2}\xi^T\Sigma\xi}\right\}\right)$$
$$= -\tfrac{1}{2}\ln\mathrm{Det}(I - S\Sigma S^T) + \tfrac{1}{2}h^T\Sigma h + \tfrac{1}{2}h^T\Sigma S^T(I - S\Sigma S^T)^{-1}S\Sigma h$$
$$= -\tfrac{1}{2}\ln\mathrm{Det}(I - S\Sigma S^T) + \tfrac{1}{2}h^T\Sigma^{1/2}(I - S\Sigma S^T)^{-1}\Sigma^{1/2}h.$$

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. [34, Lemma 1])

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

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

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

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

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

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

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

$$\mathrm{Prob}\left\{\zeta^T Q\zeta \ge \min_{\alpha > 2\lambda_{\max}(Q)} \alpha[\Phi(2Q/\alpha) + \ln\epsilon^{-1}]\right\} \le \epsilon$$

Clearly, similar bounds hold with Φ replaced with $\widetilde{\Phi}$ and $\overline{\Phi}$. For instance,

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

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

$$\mathrm{Prob}\left\{\zeta^T Q\zeta \ge \mathrm{Tr}(Q) + 2\|Q\|_{\mathrm{Fro}}\sqrt{\ln\epsilon^{-1}} + 2\lambda_{\max}(Q)\ln\epsilon^{-1}\right\} \le \epsilon. \qquad (32)$$

Corollary A.1 *Let $\epsilon \in (0,1)$, $W_1, ..., W_L$ be matrices from $\mathbf{S}^d_+$, and let $\upsilon \sim \mathcal{SG}(0,V)$ be a d-dimensional sub-Gaussian random vector. Then*

$$\mathrm{Prob}\left\{\max_{\ell\leq L} \upsilon^T W_\ell \upsilon \geq \left[1+\sqrt{2\ln(L/\epsilon)}\right]^2 \max_{\ell\leq L}\mathrm{Tr}(W_\ell V)\right\} \leq \epsilon.$$

Proof. Let $R^2 = \max_{\ell\leq L}\mathrm{Tr}(W_\ell V)$. W.l.o.g. we may assume that $\upsilon = V^{1/2}\zeta$ where $\zeta \sim \mathcal{SG}(0,I)$. Let us fix $\ell \leq L$. Applying (32) with $Q = V^{1/2}W_\ell V^{1/2}$ and ϵ replaced with ϵ/L, when taking into account that $\upsilon^T W_\ell \upsilon = \zeta^T Q \zeta$ with

$$\lambda_{\max}(Q) \leq \|Q\|_{\mathrm{Fro}} \leq \mathrm{Tr}(Q) \leq R^2,$$

we get

$$\mathrm{Prob}\left\{\upsilon^T W_\ell \upsilon \geq \left[1+\sqrt{2\ln(L/\epsilon)}\right]^2 R^2\right\} \leq \frac{\epsilon}{L},$$

and the claim of the corollary follows. □

A.1.2 Proof of Proposition 2.1

Let H be a candidate contrast matrix.
1^o. Observe that

$$\|\widehat{w}^H(\omega) - Bx\| \leq \|H^T\xi\| + \left\|H^T\sum\nolimits_{\alpha=1}^q \eta_\alpha A_\alpha x\right\| + \|[B-H^TA]x\|. \tag{33}$$

Clearly,

$$\|[B-H^TA]x\| \leq \max_{\ell\leq L}\left\{\max_{x\in\mathcal{X}} x^T[B-H^TA]^T R_\ell [B-H^TA]x\right\}^{1/2},$$

so that by Theorem 2.1,

$$\forall x\in\mathcal{X} \quad \|[B-H^TA]x\| \leq \max_{\ell\leq L}\mathfrak{r}_\ell(H) \tag{34}$$

where

$$\mathfrak{r}^2_\ell(H) = \min_\upsilon\left\{\phi_{\mathcal{T}}(\upsilon):\ \upsilon\geq 0,\ \left[\begin{array}{c|c} I_\nu & R_\ell^{1/2}[B-H^TA] \\ \hline [B-H^TA]^T R_\ell^{1/2} & \sum_k \upsilon_k T_k\end{array}\right] \succeq 0\right\}.$$

Taking into account that $\sqrt{u} = \min_{\lambda\geq 0}\{\frac{u}{4\lambda}+\lambda\}$ for $u>0$, we get

$$\mathfrak{r}_\ell(H) = \min_{\upsilon,\lambda}\left\{\lambda + \frac{\phi_{\mathcal{T}}(\upsilon)}{4\lambda}:\ \upsilon\geq 0, \lambda\geq 0,\ \left[\begin{array}{c|c} I_\nu & R_\ell^{1/2}[B-H^TA] \\ \hline [B-H^TA]^T R_\ell^{1/2} & \sum_k \upsilon_k T_k\end{array}\right] \succeq 0\right\}.$$

Setting $\mu = \upsilon/(4\lambda)$, by the homogeneity of $\phi_{\mathcal{T}}(\cdot)$ we obtain

$$\mathfrak{r}_\ell(H) = \min_{\mu,\lambda}\left\{\lambda + \phi_{\mathcal{T}}(\mu):\ \mu\geq 0,\ \left[\begin{array}{c|c} \lambda I_\nu & \frac{1}{2}R_\ell^{1/2}[B-H^TA] \\ \hline \frac{1}{2}[B-H^TA]^T R_\ell^{1/2} & \sum_k \mu_k T_k\end{array}\right] \succeq 0\right\}. \tag{35}$$

2^o. Next, by Corollary A.1 of the appendix,

$$\mathrm{Prob}\left\{\|H^T\xi\| \geq [1+\sqrt{2\ln(2L/\epsilon)}]\sigma\max_{\ell\leq L}\sqrt{\mathrm{Tr}(HR_\ell H^T)}\right\} \leq \epsilon/2. \tag{36}$$

Similarly, because

$$\left\|H^T\sum\nolimits_{\alpha=1}^q \eta_\alpha A_\alpha x\right\| = \max_{\ell\leq L}\left\|R_\ell^{1/2}H^T[A_1x,...,A_qx]\eta\right\|_2,$$

we conclude that for any $x \in \mathcal{X}$

$$\text{Prob}\left\{\left\|H^T\sum\nolimits_{\alpha=1}^q \eta_\alpha A_\alpha x\right\| \geq [1+\sqrt{2\ln(2L/\epsilon)}]\max_{\ell\leq L} s_\ell(H)\right\} \leq \epsilon/2$$

where $s_\ell(H) = \left\{\max_{x\in\mathcal{X}} x^T\left[\sum_\alpha A_\alpha^T H R_\ell H^T A_\alpha\right] x\right\}^{1/2}$. Again, by Theorem 2.1, $s_\ell(H)$ may be tightly upper-bounded by the quantity $\bar{\mathfrak{s}}_\ell(H)$ such that

$$\bar{\mathfrak{s}}_\ell^2(H) = \min_\upsilon \{\phi_\mathcal{T}(\upsilon) :\, \upsilon \geq 0,$$
$$\left.\left[\begin{array}{c|c} I_{\nu q} & [R_\ell^{1/2}H^TA_1;...;R_\ell^{1/2}H^TA_q] \\ \hline [A_1^THR_\ell^{1/2},...,A_q^THR_\ell^{1/2}] & \sum_k \upsilon_k T_k \end{array}\right] \succeq 0\right\}.$$

Now, repeating the steps which led to (35) above, we conclude that

$$\bar{\mathfrak{s}}_\ell(H) = \min_{\mu',\lambda'}\Big\{\lambda' + \phi_\mathcal{T}(\mu') :\, \mu' \geq 0,$$
$$\left.\left[\begin{array}{c|c} \lambda' I_{\nu q} & \frac{1}{2}[R_\ell^{1/2}H^TA_1;...;R_\ell^{1/2}H^TA_q] \\ \hline \frac{1}{2}[A_1^THR_\ell^{1/2},...,A_q^THR_\ell^{1/2}] & \sum_k \mu'_k T_k \end{array}\right] \succeq 0\right\}.$$

3^o. When substituting the above bounds into (33), we conclude that for every feasible solution $\lambda_\ell, \mu^\ell, \kappa^\ell, \varkappa^\ell, \rho, \varrho$ to problem (10) associated with H, the ϵ-risk of the linear estimate $\widehat{w}_{\rm lin}^H(\cdot)$ may be upper-bounded by the quantity

$$[1+\sqrt{2\ln(2L/\epsilon)}]\left[\sigma\max_{\ell\leq L}\|HR_\ell^{1/2}\|_{\rm Fro} + \rho\right] + \varrho. \qquad \square$$

A.1.3 Proof of Proposition 2.2

1^o. Let $\ell \leq L$ and $k \leq K$ 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 (13). One has

$$\begin{aligned}\mathbf{E}_{\xi_k}\left\{\|R_\ell^{1/2}H^T\xi_k\|_2^2\right\} &= \text{Tr}\left(\mathbf{E}_{\xi_k}\left\{R_\ell^{1/2}H^T\xi_k\xi_k^THR_\ell^{1/2}\right\}\right)\\ &\leq \sigma^2\text{Tr}(HR_\ell H^T) = \sigma^2\|HR_\ell^{1/2}\|_{\rm Fro}^2. \end{aligned} \tag{37}$$

Next, for any $x \in \mathcal{X}$ fixed we have

$$\begin{aligned}&\mathbf{E}_{\eta_k}\left\{\left\|R_\ell^{1/2}H^T[\sum\nolimits_\alpha [\eta_k]_\alpha A_\alpha]x\right\|_2^2\right\} = \mathbf{E}_{\eta_k}\left\{\left\|R_\ell^{1/2}H^T[A_1x,...,A_qx]\eta_k\right\|_2^2\right\}\\ &= x^T\left[\sum\nolimits_\alpha A_\alpha^T HR_\ell H^TA_\alpha\right]x = \|[R_\ell^{1/2}H^TA_1;...;R_\ell^{1/2}H^TA_q]x\|_2^2 \leq (\lambda+\phi_\mathcal{T}(\mu))^2, \end{aligned} \tag{38}$$

where the concluding inequality follows from the constraints in (13) (cf. item 2^o of the proof of Proposition 2.1). Next, similarly to item 1^o of the proof of Proposition 2.1 we have

$$\|R_\ell^{1/2}(B-H^TA)x\|_2^2 \leq (\kappa+\phi_\mathcal{T}(\varkappa))^2.$$

Put together, the latter bound along with (37) and (38) imply (14).
2^o. By the Chebyshev inequality,

$$\forall \ell, k \quad \text{Prob}\left\{\|R_\ell^{1/2}(w_\ell(\omega_k) - Bx)\|_2 \geq 2\widetilde{\mathfrak{R}}_\ell[H_\ell]\right\} \leq \tfrac{1}{4};$$

applying [39, Theorem 3.1] we conclude that

$$\forall \ell \quad \text{Prob}\left\{\|R_\ell^{1/2}(z_\ell(\omega^K) - Bx)\|_2 \geq 2C_\alpha\widetilde{\mathfrak{R}}_\ell[H_\ell]\right\} \leq e^{-K\psi(\alpha,\frac{1}{4})},$$

where

$$\psi(\alpha,\beta)=(1-\alpha)\ln\frac{1-\alpha}{1-\beta}+\alpha\ln\frac{\alpha}{\beta}$$

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 $\ell\leq L$

$$\mathrm{Prob}\left\{\|R_\ell^{1/2}(z_\ell(\omega^K)-Bx)\|_2\geq 4\widetilde{\mathfrak{R}}_\ell[H_\ell]\right\}\leq e^{-0.1070K},$$

what is (15).

3^o. Now, let $K\geq\ln(L/\epsilon)/0.1070$. In this case, for all $\ell\leq L$

$$\mathrm{Prob}\left\{\|R_\ell^{1/2}(z_\ell(\omega^K)-Bx)\|_2\geq 4\widetilde{\mathfrak{R}}_\ell[H_\ell]\right\}\leq \epsilon/L,$$

so that with probability $\geq 1-\epsilon$ the set $\mathcal{W}(\omega^K)$ is not empty (it contains Bx), and for all $v\in\mathcal{W}(\omega^K)$ one has

$$\|R_\ell^{1/2}(v-Bx)\|_2\leq\|R_\ell^{1/2}(z_\ell(\omega^K)-v)\|_2+\|R_\ell^{1/2}(z_\ell(\omega^K)-Bx)\|_2\geq 8\widetilde{\mathfrak{R}}_\ell[H_\ell].\qquad\square$$

A.2 Proofs for Section 3

A.2.1 Proof of Proposition 3.1

All we need to prove is that if $\lambda_\ell,\mu^\ell,\upsilon^\ell,\rho$ is a feasible solution to the optimization problem (24), then the inequality

$$\mathrm{Risk}_\epsilon[\widehat{w}^H_{\mathrm{poly}}|\mathcal{X}]\leq 2\rho$$

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

$$\mathcal{Z}^c=\{\zeta:\|H^T\zeta\|_\infty>1\}\qquad\qquad[\zeta=\textstyle\sum_\alpha\eta_\alpha A_\alpha x+\xi]$$

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

$$\|H^T[\omega-Ax]\|_\infty=\|H^T\zeta\|_\infty\leq 1,$$

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 $\|H^T[A\widehat{x}-\omega]\|_\infty\leq 1$, see (17). We conclude that setting $z=\frac{1}{2}[x-\overline{x}]$, we have

$$\|H_\ell^TAz\|_\infty\leq 1,\ell\leq L$$

with $z\in\mathcal{X}$, implying that $z^TT_kz\leq t_k$, $k\leq K$, for some $t\in\mathcal{T}$. Now let $u\in\mathbf{R}^\nu$ with $\|u\|_2\leq 1$. Semidefinite constraints in (24) imply that

$$\begin{aligned}u^TR_\ell^{1/2}Bz&\leq u^T\lambda_\ell I_\nu u+z^T\left[A^TH_\ell\mathrm{Diag}\{\upsilon^\ell\}H_\ell^TA+\textstyle\sum_k\mu_k^\ell T_k\right]z\\&\leq\lambda_\ell u^Tu+\textstyle\sum_j\upsilon_j^\ell[H^TAz]_j^2+\sum_k\mu_k^\ell t_k\leq\lambda_\ell+\sum_j\upsilon_j^\ell+\phi_{\mathcal{T}}(\mu^\ell)\leq\rho\end{aligned}$$

(recall that $\|u\|_2\leq 1$, $\lambda_\ell\geq 0,\mu^\ell\geq 0,\upsilon^\ell\geq 0$, $t\in\mathcal{T}$, and $\|H_\ell^TAz\|_\infty\leq 1$). We conclude that $u^TR_\ell^{1/2}Bz\leq\rho$, $\ell\leq L$, whenever $\|u\|_2\leq 1$, i.e., $\|R_\ell^{1/2}[Bz]\|_2\leq\rho^2$. The latter relation holds true for all $\ell\leq L$, implying that $\|Bz\|\leq\rho$, that is, $\|Bx-\widehat{x}(\omega)\|=2\|Bz\|\leq 2\rho$ whenever $\zeta\in\mathcal{Z}$. $\square$

A.2.2 Proof of Proposition 3.2

0^o. We need the following technical result (Noncommutative Khintchine Inequality).

Theorem A.1 [49, 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*

$$\mathrm{Prob}\left\{\left\|\sum\nolimits_{i=1}^I \xi_i Q_i\right\| \ge t\right\} \le 2n \exp\left\{-\frac{t^2}{2v_Q}\right\},$$

where $\|\cdot\|$ is the spectral norm, and $v_Q = \left\|\sum_{i=1}^I Q_i^2\right\|$.

1^o. Proof of (i). Let $\lambda_j \ge 0$, $g_j \in \mathcal{H}$, $j \le M$, and $\Sigma = \sum_j \lambda_j g_j g_j^T$. Then for every j there exists $r^j \in \mathcal{R}$ such that $S_i^2[g_j] \preceq [r^j]_i I_{d_i}$, $i \le I$. 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

$$\mathcal{S}_i\left[\sum\nolimits_j \lambda_j g_j g_j^T\right] = \sum\nolimits_j \lambda_j S_i^2[g_j] \preceq \sum\nolimits_j \lambda_j [r^j]_i I_{d_i} = \left[\sum\nolimits_j \lambda_j\right] r_i I_{d_i},$$

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 $g_j \in \mathcal{H}$, $\lambda_j \ge 0$, and $\sum_j \lambda_j \le \varkappa\rho$. There is nothing to prove when $\rho = 0$, since in this case $\Sigma = 0$ due to $(\Sigma, 0) \in \mathbf{K}$ combined with (29b). Now let $\rho > 0$, so that for some $r \in \mathcal{R}$ we have

$$\mathcal{S}_i[\Sigma] \preceq \rho r_i I_{d_i},\ i \le I, \tag{39}$$

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 $\le \sqrt{2/N}$. 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,

$$\sum\nolimits_{p=1}^N \mathrm{Col}_p[Z^\varsigma]\mathrm{Col}_p^T[Z^\varsigma] \equiv \Sigma.$$

Recall that

$$\mathrm{Col}_j[Z^\varsigma] = \sum\nolimits_p \varsigma_p O_{pj} \mathrm{Col}_p[Z],$$

and thus

$$S_i[\mathrm{Col}_j[Z^\varsigma]] = \sum\nolimits_p \varsigma_p O_{pj} S_i[\mathrm{Col}_p[Z]].$$

Now observe that

$$\begin{aligned}
\sum\nolimits_p \left(O_{pj} S_i[\mathrm{Col}_p[Z]]\right)^2 &= \sum\nolimits_p O_{pj}^2 S_i^2[\mathrm{Col}_p[Z]] = \sum\nolimits_p O_{pj}^2 \mathcal{S}_i[\mathrm{Col}_p[Z]\mathrm{Col}_p^T[Z]] \\
\text{[see (29a)]} \quad &\preceq \frac{2}{N}\sum\nolimits_p \mathcal{S}_i[\mathrm{Col}_p[Z]\mathrm{Col}_p^T[Z]] \\
&= \frac{2}{N}\mathcal{S}_i[\sum\nolimits_p \mathrm{Col}_p[Z]\mathrm{Col}_p^T[Z]] = \frac{2}{N}\mathcal{S}_i[\Sigma] \preceq \frac{2}{N}\rho s_i I_{d_i}
\end{aligned}$$

due to (39). By the Noncommutative Khintchine Inequality we have

$$\forall \gamma > 0 : \mathrm{Prob}\left\{S_i^2[\mathrm{Col}_j[Z^\varsigma]] \preceq \gamma \frac{2}{N}\rho s_i I_{d_i}\right\} \ge 1 - 2d_i \exp\{-\gamma/2\}$$

Setting

$$\gamma=2\ln(4DN),\ D=\sum\nolimits_i d_i,\quad g_j^\varsigma=\sqrt{\frac{N}{2\gamma\rho}}\mathrm{Col}_j[Z^\varsigma],\quad \lambda_j=\frac{2\gamma\rho}{N},\ 1\leq j\leq N,$$

we conclude that event

$$\Xi=\left\{\varsigma: S_i^2[g_j^\varsigma]\preceq s_iI_{d_i}, i\leq I, j\leq N\right\}\subset\left\{g_j^\varsigma\in\mathcal{H},\ j\leq N\right\}$$

satisfies $\mathrm{Prob}(\Xi)\geq\frac{1}{2}$, while

$$\sum\nolimits_j\lambda_j g_j^\varsigma[g_j^\varsigma]^T=\sum\nolimits_j\mathrm{Col}_j[Z^\varsigma]\mathrm{Col}_j^T[Z^\varsigma]\equiv\Sigma \text{ and } \sum\nolimits_j\lambda_j=\gamma\kappa\rho=2\gamma\rho=\varkappa\rho.$$

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 $\leq 2^{-k}$.

A.2.3 Proof of Lemma 3.1

The proof of the lemma is given by the standard argument underlying median-of-means construction (cf. [42, Section 6.5.3.4]). For the sake of completeness, we reproduce it here.
1^o. Observe that when (12) 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 (12) 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, $z_j=y_j-h_j^TAx$ 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[4] that the probability that at least $K/2$ of them happen is bounded with

$$\begin{aligned}\pi(K):=&\sum_{k\geq K/2}\binom{K}{k}(1/8)^k(7/8)^{K-k}\leq\sum_{k\geq K/2}\binom{K}{k}2^{-K}[(1/4)^k(7/4)^{K-k}]\\&\leq(\sqrt{7}/4)^K\leq e^{-0.4K}.\end{aligned}$$

In other words, the probability of each event $E_j=\{\omega^K:|y_j-h_j^TAx|>1\}$, $j=1,...,M$, 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 $\|y-H^TAx\|_\infty\leq 1$, and so $\|y-H^TA\widehat{x}^H(\omega^K)\|_\infty\leq 1$ as well. We conclude that for every $x\in\mathcal{X}$, the probability of the event

$$\left\{x-\widehat{x}^H(\omega^K)\in 2\mathcal{X},\ \|H^TA[x-\widehat{x}^H(\omega^K)]\|_\infty\leq 2\right\}$$

is at least $1-M\pi(K)\geq 1-\epsilon$ when $K\geq 2.5\ln[M/\epsilon]$, and when it happens, one has $\|Bx-\widehat{w}^H_{\rm poly}(\omega^K)\|\leq\mathfrak{p}[H]$. □

[4] We refer to, e.g., [15, Section 2.3.2] for the precise justification of this obvious claim.

References

1. M. Bennani, M.-C. Brunet, and F. Chatelin. 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(16):847–850, 1988.
2. R. Carroll and D. Ruppert. The use and misuse of orthogonal regression in linear errors-in-variables models. *The American Statistician*, 50(1):1–6, 1996.
3. L. Cavalier and N. W. Hengartner. Adaptive estimation for inverse problems with noisy operators. *Inverse Problems*, 21(4):1345–1361, 2005.
4. L. Cavalier and M. Raimondo. Wavelet deconvolution with noisy eigenvalues. *IEEE Transactions on Signal Processing*, 55(6):2414–2424, 2007.
5. A. Cohen, M. Hoffmann, and M. Reiss. Adaptive wavelet galerkin methods for linear inverse problems. *SIAM Journal on Numerical Analysis*, 42(4):1479–1501, 2004.
6. D. L. Donoho. Statistical estimation and optimal recovery. *The Annals of Statistics*, 22(1):238–270, 1994.
7. D. L. Donoho. Nonlinear solution of linear inverse problems by wavelet–vaguelette decomposition. *Applied and Computational Harmonic Analysis*, 2(2):101–126, 1995.
8. D. L. Donoho, R. C. Liu, and B. MacGibbon. Minimax risk over hyperrectangles, and implications. *The Annals of Statistics*, 18(3):1416–1437, 1990.
9. S. Efromovich. *Nonparametric curve estimation: methods, theory, and applications*. Springer Science & Business Media, 2008.
10. S. Efromovich and V. Koltchinskii. On inverse problems with unknown operators. *IEEE Transactions on Information Theory*, 47(7):2876–2894, 2001.
11. S. Efromovich and M. Pinsker. Sharp-optimal and adaptive estimation for heteroscedastic nonparametric regression. *Statistica Sinica*, 6(4):925–942, 1996.
12. J. Fan and Y. K. Truong. Nonparametric regression with errors in variables. *The Annals of Statistics*, 21(4):1900–1925, 1993.
13. L. J. Gleser. Estimation in a multivariate "errors in variables" regression model: large sample results. *The Annals of Statistics*, 9(1):24–44, 1981.
14. M. Grasmair, H. Li, A. Munk, et al. Variational multiscale nonparametric regression: Smooth functions. 54(2):1058–1097, 2018.
15. V. Guigues, A. Juditsky, and A. Nemirovski. Hypothesis testing via Euclidean separation. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 56(3):1929 – 1957, 2020.
16. P. Hall and J. L. Horowitz. Nonparametric methods for inference in the presence of instrumental variables. *The Annals of Statistics*, 33(6):2904–2929, 2005.
17. M. Hoffmann and M. Reiss. Nonlinear estimation for linear inverse problems with error in the operator. *The Annals of Statistics*, 36(1):310–336, 2008.
18. D. Hsu, S. Kakade, and T. Zhang. A tail inequality for quadratic forms of subgaussian random vectors. *Electronic Communications in Probability*, 17:1–6, 2012.
19. D. Hsu and S. Sabato. Heavy-tailed regression with a generalized median-of-means. In E. P. Xing and T. Jebara, editors, *Proceedings of the 31st International Conference on Machine Learning*, volume 32 of *Proceedings of Machine Learning Research*, pages 37–45, Bejing, China, 22–24 Jun 2014. PMLR.
20. I. A. Ibragimov and R. Z. Has'minskii. *Statistical estimation: asymptotic theory*, volume 16. Springer Science & Business Media, 2013.
21. I. M. Johnstone and B. W. Silverman. Speed of estimation in positron emission tomography and related inverse problems. *The Annals of Statistics*, 18(1):251–280, 1990.
22. I. M. Johnstone and B. W. Silverman. Discretization effects in statistical inverse problems. *Journal of complexity*, 7(1):1–34, 1991.
23. A. Juditsky, G. Kotsalis, and A. Nemirovski. Tight computationally efficient approximation of matrix norms with applications. *Open Journal of Mathematical Optimization*, 3(7):1–38, 2022.
24. A. Juditsky and A. Nemirovski. Nonparametric estimation by convex programming. *The Annals of Statistics*, 37(5a):2278–2300, 2009.
25. A. Juditsky and A. Nemirovski. Near-optimality of linear recovery from indirect observations. *Mathematical Statistics and Learning*, 1(2):171–225, 2018.

26. A. Juditsky and A. Nemirovski. Near-optimality of linear recovery in Gaussian observation scheme under $\|\cdot\|_2^2$-loss. *The Annals of Statistics*, 46(4):1603–1629, 2018.
27. A. Juditsky and A. Nemirovski. On polyhedral estimation of signals via indirect observations. *Electronic Journal of Statistics*, 14(1):458—502, 2020.
28. A. Juditsky and A. Nemirovski. *Statistical Inference via Convex Optimization*. Princeton University Press, 2020.
29. A. Juditsky and A. Nemirovski. On design of polyhedral estimates in linear inverse problems. *SIAM Journal on Mathematics of Data Science*, 6(1):76–96, 2024.
30. J. Kaipio and E. Somersalo. *Statistical and computational inverse problems*, volume 160. Springer Science & Business Media, 2006.
31. J. A. Kuks and W. Olman. Minimax linear estimation of regression coefficients (i). *Iswestija Akademija Nauk Estonskoj SSR*, 20:480–482, 1971.
32. J. A. Kuks and W. Olman. Minimax linear estimation of regression coefficients (ii). *Iswestija Akademija Nauk Estonskoj SSR*, 21:66–72, 1972.
33. A. Kukush, I. Markovsky, and S. Van Huffel. Consistency of the structured total least squares estimator in a multivariate errors-in-variables model. *Journal of Statistical Planning and Inference*, 133(2):315–358, 2005.
34. B. Laurent and P. Massart. Adaptive estimation of a quadratic functional by model selection. *The Annals of Statistics*, 28(5):1302–1338, 2000.
35. G. Lecué and M. Lerasle. Robust machine learning by median-of-means: theory and practice. *The Annals of Statistics*, 48(2):906–931, 2020.
36. M. Lerasle and R. I. Oliveira. Robust empirical mean estimators. *arXiv preprint arXiv:1112.3914*, 2011.
37. B. A. Mair and F. H. Ruymgaart. Statistical inverse estimation in Hilbert scales. *SIAM Journal on Applied Mathematics*, 56(5):1424–1444, 1996.
38. C. Marteau. Regularization of inverse problems with unknown operator. *Mathematical Methods of Statistics*, 15(4):415–443, 2006.
39. S. Minsker. Geometric median and robust estimation in Banach spaces. *Bernouilli*, 21(4):2308–2335, 2015.
40. F. Natterer. *The Mathematics of Computerized Tomography*. SIAM, 1986.
41. A. Nemirovski. Topics in non-parametric statistics. In P. Bernard, editor, *Lectures on Probability Theory and Statistics, Ecole d'Eté de Probabilités de Saint-Flour*, volume 28, pages 87–285. Springer, 2000.
42. A. S. Nemirovski and D. B. Yudin. *Problem Complexity and Method Efficiency in Optimization*. J. Wiley & Son, 1983.
43. A. Nemirovskii. Nonparametric estimation of smooth regression functions. *Soviet Journal of Computation and System Science*, 23(6):1–11, 1985.
44. M. Pinsker. Optimal filtration of square-integrable signals in gaussian noise. *Problems of Information Transmission*, 16(2):120–133, 1980.
45. K. Proksch, F. Werner, A. Munk, et al. Multiscale scanning in inverse problems. *The Annals of Statistics*, 46(6B):3569–3602, 2018.
46. M. Rudelson and R. Vershynin. Hanson-Wright inequality and sub-gaussian concentration. *Electronic Communications in Probability*, 18(82):1–9, 2013.
47. V. Spokoiny and M. Zhilova. Sharp deviation bounds for quadratic forms. *Mathematical Methods of Statistics*, 22:100–113, 2013.
48. G. W. Stewart. Stochastic perturbation theory. *SIAM Review*, 32(4):579–610, 1990.
49. J. A. Tropp. An introduction to matrix concentration inequalities. *Foundations and Trends in Machine Learning*, 8(1-2):1–230, 2015.
50. A. B. Tsybakov. *Introduction to Nonparametric Estimation*. Springer Series in Statistics. Springer, New York, 2009.
51. S. Van Huffel and P. Lemmerling. *Total least squares and errors-in-variables modeling: analysis, algorithms and applications*. Springer Science & Business Media, 2013.
52. C. R. Vogel. *Computational Methods for Inverse Problems*. SIAM, 2002.
53. L. Wasserman. *All of nonparametric statistics*. Springer Science & Business Media, 2006.