ISSN 0005-1179 (print), ISSN 1608-3032 (online), Automation and Remote Control, 2025, Vol. 86, No. 8, pp. 718–739.

SPECIAL ISSUE

On Robust Recovery of Signals from Indirect Observations

Ya. Bekri[*,a], A. Nemirovski[,b], and A. Juditsky[*,c]**

[*] *LJK, Université Grenoble Alpes, Campus de Saint-Martin-d'Hères, 38401 France*
[**] *Georgia Institute of Technology, Atlanta, Georgia, 30332, USA*
e-mail: [a] *yannis.bekri@univ-grenoble-alpes.fr*, [b] *nemirovs@isye.gatech.edu*,
[c] *anatoli.juditsky@univ-grenoble-alpes.f*

Received March 3, 2025
Revised May 20, 2025
Accepted June 27, 2025

Abstract—We consider an *uncertain linear inverse problem* as follows. Given observation $\omega = Ax_* + \zeta$ where $A \in \mathbf{R}^{m\times p}$ and $\zeta \in \mathbf{R}^m$ is observation noise, we want to recover unknown signal x_*, known to belong to a convex set $\mathcal{X} \subset \mathbf{R}^n$. As opposed to the "standard" setting of such a problem, we suppose that the model noise ζ is "corrupted"—contains an uncertain (deterministic dense or singular) component. Specifically, we assume that ζ decomposes into $\zeta = N\nu_* + \xi$ where ξ is the random noise and $N\nu_*$ is the "adversarial contamination" with known $\mathcal{N} \subset \mathbf{R}^n$ such that $\nu_* \in \mathcal{N}$ and $N \in \mathbf{R}^{m\times n}$. We consider two "uncertainty setups" in which $\mathcal{N}$ is either a convex bounded set or is the set of sparse vectors (with at most s nonvanishing entries). We analyse the performance of "uncertainty-immunized" *polyhedral estimates*—a particular class of nonlinear estimates as introduced in [19, 20]—and show how "presumably good" estimates of the sort may be constructed in the situation where the signal set is an *ellitope* (essentially, a symmetric convex set delimited by quadratic surfaces) by means of efficient convex optimization routines.

Keywords: robust estimation, linear inverse problems with contaminated observations, signal estimation in singular noise

DOI: 10.31857/S0005117925080038

1. SITUATION AND GOALS

1.1. Introduction

Since the term was coined in the 1950s, the problem of robust estimation has received much attention in the classical statistical literature. It is impossible to give an overview of the existing literature on robust estimation, and we do not try to do it here; for the "classical" framework one may refer to early references in [39], the foundational manuscript [16], or a recent survey [41].[1]

In this paper, our focus is on robust estimation of a signal from indirect linear observations. Specifically, suppose that our objective is to recover a linear image $w_* = Bx_*$ of unknown signal x_*, known to belong to a given convex set $\mathcal{X} \subset \mathbf{R}^p$, given $B \in \mathbf{R}^{q\times p}$, $A \in \mathbf{R}^{m\times p}$, and a noisy observation

$$\omega = Ax_* + \eta_* + \xi \in \mathbf{R}^m \tag{1}$$

of x_*, perturbed by a mixed—random-and-deterministic noise $\xi + \eta_*$. Here ξ is the random noise component, while η_* is the "adversarial" deterministic noise. Recently, this problem attracted much attention in the context of robust recovery of sparse (with at most $s \ll p$ nonvanishing entries) signal x_*. In particular, robust sparse regression with an emphasis on contaminated design was

[1] An important contribution to the development of robust statistics has been the line of work on distributionally robust algorithms of stochastic optimization and system identification by B. Polyak and Ya. Tsypkin, see [31–36].

investigated in [1, 5, 9, 25, 29]; methods based on penalizing the vector of outliers were studied in [7, 13], see also [3, 37]. We refer to the monograph [8] for the description of the present state of the art.

In this paper, our emphasis is on rather different assumptions about the structure of the signal x_* to be recovered and on the contamination η_* what precludes direct comparison with the cited results. Namely, we assume that the set $\mathcal{X}$ of signals is an *ellitope*—a convex compact symmetric w.r.t. the origin subset of $\mathbf{R}^p$ delimited by quadratic surfaces.[2] Our interest in ellitopes is motivated by the fact that these signal sets are well suited for the problem of estimating unknown signal x_* from observation (1) in the Gaussian no-nuisance case ($\eta_* = 0$, $\xi \sim \mathcal{N}(0, \sigma^2 I_m)$). Specifically, let us consider linear estimate $\widehat{w}_{\rm lin}(\omega) = G_{\rm lin}^T \omega$ and *polyhedral estimate* $\widehat{w}_{\rm poly}(\omega) = B\widehat{x}_{\rm poly}(\omega)$ where

$$\widehat{x}_{\rm poly}(\omega) \in \mathop{\rm Argmin}_{x \in \mathcal{X}} \|G_{\rm poly}^T(Ax - \omega)\|_\infty$$

of w_*. Let $\mathcal{X}$ be an ellitope, and let the estimation error be measured in a co-ellitopic norm $\|\cdot\|$ (i.e., such that the unit ball $\mathcal{B}_*$ of the norm $\|\cdot\|_*$ conjugate to $\|\cdot\|$ is an ellitope). In this situation, one can point out (cf. [17, 19, 20]) efficiently computable *contrast matrices* $G_{\rm lin} \in \mathbf{R}^{m\times q}$ and $G_{\rm poly} \in \mathbf{R}^{m\times m}$ such that estimates $\widehat{w}_{\rm lin}(\cdot)$ and $\widehat{w}_{\rm poly}(\cdot)$ attain nearly minimax-optimal performance.

We suppose that the adversarial perturbation η_* has a special structure: we are given a "nuisance set" $\mathcal{N} \subset \mathbf{R}^n$ such that $\nu_* \in \mathcal{N}$ and a $m \times n$ matrix N such that $\eta_* = N\nu_*$. We consider two types of assumptions about $\mathcal{N}$: $\mathcal{N}$ is either 1) a (nonempty) compact convex set, or, more conventionally, 2) $\mathcal{N}$ is the set of sparse disturbances (with at most $s \leqslant n$ nonvanishing components). Our focus is on the design and performance analysis of the "uncertainty immunized" polyhedral estimate $\widehat{w}(\omega)$ of $w_* = Bx_*$ in the presence of the contaminating signal, and solving the problem in the first case leads to a "presumably good" solution for the second.

We would like to emphasize the principal feature of the approach we promote: in this paper, A, B, and N are "general" matrices of appropriate dimensions, while $\mathcal{X}$ and $\mathcal{N}$ are rather general sets. As a consequence, we adopt here an "operational" approach[3] initiated in [10] and further developed in [18–20, 22], within which both the estimates and their risks are yielded by efficient computation, rather than by an explicit analytical analysis, seemingly impossible under the circumstances. The term "efficient" in the above is essential and is also responsible for the principal limitations of the results to follow. First of all, it imposes restrictions on the structure of the set of signals of interest and on the norm quantifying the estimation error. As it is shown in [20], the maximum of a quadratic form over an ellitope admits a "reasonably tight" efficiently computable upper bound, leading to tight bounds on the risk of linear and polyhedral estimates when the signal set is an ellitope. Furthermore, while in the case of convex compact set $\mathcal{N}$ of contaminations, constructing risk bounds for the polyhedral estimate $\widehat{w}_G(\omega)$ associated with a *given* contrast matrix G is possible under rather weak assumptions about the nuisance set $\mathcal{N}$ (essentially, the computational tractability[4] of this set is sufficient), the fundamental problem of *contrast synthesis*—minimizing these bounds over contrast matrices—allows for efficiently computable solution only when $\mathcal{N}$ is either an ellitope itself, or is a "co-ellitope" (the polar of an ellitope).

To complete this section, we would like to mention another line of research on the problem of estimating signal x_* from observation (1) under purely deterministic disturbance (case of $\xi = 0$),

[2] See [20, Section 4.2.1] or Section 1.3 below; as of now, an instructive example of ellitope is an intersection of a finite family of ellipsoids/elliptic cylinders with a common center.

[3] As opposed to the classical "descriptive" approach to solving the estimation problem in question via deriving and optimizing, w.r.t. estimate parameters, closed-form analytical expressions for the risk of a candidate estimate.

[4] For most practical purposes, computational tractability of a est means that we can model the set constraint using the CVX [15]. For an "executive summary" of what these words actually mean, we refer the reader to [20, Appendix A].

the standard problem of optimal recovery [26, 27] and guaranteed estimation in dynamical systems under uncertain-but-bounded perturbation [6, 12, 14, 23, 24, 28, 38]. The present work may be seen as an attempt to extend the corresponding framework to the case in which both deterministic and random observation noises are present.

Organization of the paper. We introduce the exact statement of the estimation problem to be considered and the entities that are relevant for the analysis to follow in Section 1.2. Analysis and design of the polyhedral estimate in the case of uncertain-but-bounded contamination are presented in Section 2. Then in Section 3 we describe the application of the proposed framework to the case of (unbounded) singular contamination using the sparse model of the nuisance vector. Finally, we recall some results on ℓ_1 recovery used in the paper in Appendix.

Notation. In the sequel, order relations between vectors are understood entry-wise; e.g., $t \geqslant t'$ for $t, t' \in \mathbf{R}^n$ means that vector $t - t'$ has nonnegative entries. $[A; B]$ and $[A, B]$ stand for vertical and horizontal concatenation of matrices A and B of appropriate dimensions. We denote $\mathbf{S}^m$ the space of symmetric $m \times m$ matrices, $\mathbf{S}^m_+$ denotes the positive semidefinite cone of $\mathbf{S}^m$; notation $A \succeq B$ ($A \succ B$) means that matrix $A - B$ is positive semidefinite (respectively, positive definite). In what follows, for a nonempty compact set $\mathcal{Z} \subset \mathbf{R}^N$

$$\phi_{\mathcal{Z}}(\zeta) := \max_{z \in \mathcal{Z}} \zeta^T t$$

is the support function of $\mathcal{Z}$. We denote $\mathrm{n}[G]$ the maximum of Euclidean norms of the columns of a matrix G.

1.2. The Problem

The estimation problem we are interested in is as follows:

Recall that we are given observation (cf. (1))

$$\omega = Ax_* + N\nu_* + \xi \in \mathbf{R}^m \tag{2}$$

where

- $N \in \mathbf{R}^{m \times n}$, $A \in \mathbf{R}^{m \times p}$ are given matrices,
- $\nu_* \in \mathbf{R}^n$ is unknown *nuisance signal*, $\nu_* \in \mathcal{N}$, a known subset of $\mathbf{R}^n$,
- x_* is an unknown signal of interest known to belong to a given convex compact set $\mathcal{X} \subset \mathbf{R}^p$ symmetric w.r.t. the origin,
- $\xi \sim P_\xi$ is a random observation noise.

Given ω, *our objective* is to recover the linear image $w_* = Bx_*$ of x_*, B being a given $q \times p$ matrix.

Given $\epsilon \in (0, 1)$, we quantify the quality of the recovery $\widehat{w}(\cdot)$ by its ϵ-risk[5]

$$\mathrm{Risk}_\epsilon[\widehat{w}] = \inf \left\{ \rho : \mathrm{Prob}_{\xi \sim P_\xi}\{\xi : \|Bx_* - \widehat{w}(\omega)\| > \rho\} \leqslant \epsilon \quad \forall (\nu_* \in \mathcal{N} \text{ and } x_* \in \mathcal{X}) \right\}$$

where $\|\cdot\|$ is a given norm.

Observation noise assumption. In the sequel, we assume that the observation noise ξ is sub-Gaussian with parameters $(0, \sigma^2 I_m)$, that is,

$$\mathbf{E}\left\{ e^{h^T \xi} \right\} \leqslant \exp\left(\frac{1}{2} \sigma^2 \|h\|_2^2 \right).$$

[5] The ϵ-risk of an estimate depends, aside of ϵ and the estimate, on the "parameters" $\|\cdot\|$, $\mathcal{X}$, $\mathcal{N}$; these entities will always be specified by the context, allowing us to omit mentioning them in the notation $\mathrm{Risk}_\epsilon[\cdot]$.

1.3. Ellitopes

Risk analysis of a candidate polyhedral estimate heavily depends on the geometries of the signal set $\mathcal{X}$ and norm $\|\cdot\|$. In the sequel, we restrict ourselves to the case where $\mathcal{X}$ and the *polar* $\mathcal{B}_*$ of the unit ball of $\|\cdot\|$ are *ellitopes*.

By definition [17, 20], *a basic ellitope* in $\mathbf{R}^n$ is a set of the form

$$\mathcal{X} = \{x \in \mathbf{R}^n : \exists, t \in \mathcal{T} : x^T T_\ell x \leqslant t_\ell, \ell \leqslant L\}, \tag{3}$$

where $T_\ell \in \mathbf{S}^k$, $T_\ell \succeq 0$, $\sum_\ell T_\ell \succ 0$, and $\mathcal{T} \subset \mathbf{R}^L_+$ is a convex compact set with a nonempty interior which is monotone: whenever $0 \leqslant t' \leqslant t \in \mathcal{T}$ one has $t' \in \mathcal{T}$. An ellitope is an image of a basic ellitope under a linear mapping. We refer to L as *ellitopic dimension* of $\mathcal{X}$.

Clearly, every ellitope is a convex compact set symmetric w.r.t. the origin; a *basic ellitope*, in addition, has a nonempty interior.

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

$$\mathcal{X} = \{x \in \mathbf{R}^n : \exists t \in \mathcal{T} := [0,1]^L : x^T T_\ell x \leqslant t_\ell, \ell \leqslant L\}$$

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

$$\{x \in \mathbf{R}^n : \|x\|_p \leqslant 1\} = \left\{ x : \exists t \in \mathcal{T} = \{t \in \mathbf{R}^n_+, \|t\|_{p/2} \leqslant 1\} : \underbrace{x_\ell^2}_{x^T T_\ell x} \leqslant t_\ell, \ell \leqslant n \right\}.$$

Ellitopes admit fully algorithmic "calculus"—this family is closed with respect to basic operations preserving convexity and symmetry w.r.t. the origin, e.g., taking finite intersections, linear images, inverse images under linear embedding, direct products, arithmetic summation (for details, see [20, Section 4.6]).

Main assumption. We assume from now on that the signal set $\mathcal{X}$ and the *polar* $\mathcal{B}_*$ of the unit ball of the norm $\|\cdot\|$ are *basic ellitopes*:[6]

$$\mathcal{X} = \{x \in \mathbf{R}^n : \exists t \in \mathcal{T} : x^T T_k x \leqslant t_k, k \leqslant K\}, \tag{4a}$$

$$\mathcal{B}_* = \{y \in \mathbf{R}^q : \exists s \in \mathcal{S} : y^T S_\ell y \leqslant s_\ell, \ell \leqslant L\} \tag{4b}$$

where $\mathcal{T} \subset \mathbf{R}^K_+$, $\mathcal{S} \in \mathbf{R}^L_+$ are monotone convex compact sets with nonempty interiors, $T_k \succeq 0$, $\sum_k T_k \succ 0$, $S_\ell \succeq 0$, and $\sum_\ell S_\ell \succ 0$.

2. UNCERTAIN-BUT-BOUNDED NUISANCE

In this section, we consider the case of *uncertain-but-bounded* nuisance. Specifically, we assume that $\mathcal{N} \subset \mathbf{R}^n$ is a convex compact set, symmetric w.r.t. the origin, and specify $\pi(\cdot)$ as the semi-norm on $\mathbf{R}^m$ given by

$$\pi(h) = \sup_u \left\{ (Nu)^T h : u \in \mathcal{N} \right\}.$$

[6] The results to follow straightforwardly extend to the case where $\mathcal{X}$ and $\mathcal{B}_*$ are "general" ellitopes; we assume them to be basic to save notation.

2.1. Bounding the ϵ-Risk of Polyhedral Estimate

In this section, a polyhedral estimate is specified by $m \times I$ *contrast matrix* G and is as follows: given observation ω (see (2)), we find an optimal solution $\widehat{x}_G(\omega)$ to the (clearly solvable) optimization problem

$$\min_{x,\nu} \left\{ \|G^T(Ax + N\nu - \omega)\|_\infty : x \in \mathcal{X}, \nu \in \mathcal{N} \right\}. \tag{5}$$

Given a $m \times I$ matrix $G = [g_1, \ldots, g_I]$, let $\Xi_\epsilon[G]$ be the set of all realizations of ξ such that

$$|[g_i^T \xi]_i| \leqslant \underbrace{\sigma\sqrt{2\ln[2I/\epsilon]}}_{=:\varkappa(\epsilon)} \|g_i\|_2, \quad \forall i \leqslant I. \tag{6}$$

Note that

$$\mathrm{Prob}_{\xi \sim \mathcal{SG}(0,\sigma^2 I_m)} \{\xi \notin \Xi_\epsilon(G)\} \leqslant \epsilon. \tag{7}$$

Indeed, we have $\mathbf{E}_\xi \left\{ e^{\gamma g^T \xi} \right\} \leqslant e^{\frac{1}{2}\gamma^2 \|g\|_2^2 \sigma^2}$, implying that for all $a \geqslant 0$,

$$\mathrm{Prob}\{g^T \xi > a\} \leqslant \inf_{\gamma > 0} \exp\left\{ \frac{1}{2}\gamma^2 \|g\|_2^2 \sigma^2 - \gamma a \right\} = \exp\left\{ -\frac{1}{2} a^2 \sigma^2 \|g\|_2^2 \right\},$$

so that

$$\mathrm{Prob}\left\{ \exists i \leqslant I : |g_i^T \xi| > \varkappa(\epsilon)\|g_i\|_2 \right\} \leqslant 2I \exp\left\{ -\frac{\varkappa^2(\epsilon)}{2\sigma^2} \right\} \leqslant \epsilon.$$

Given a $m \times I$ contrast matrix $G = [g_1, \ldots, g_I]$, consider the optimization problem

$$\mathrm{Opt}[G] = \min_{\lambda,\mu,\gamma} \left\{ f_G(\lambda,\mu,\gamma) : \lambda \geqslant 0, \mu \geqslant 0, \gamma \geqslant 0, \right. \tag{8}$$
$$\left. \left[\begin{array}{c|c} \sum_\ell \lambda_\ell S_\ell & \frac{1}{2}B \\ \hline \frac{1}{2}B^T & \sum_k \mu_k T_k + A^T \left[\sum_i \gamma_i g_i g_i^T \right] A \end{array} \right] \succeq 0 \right\}$$

where

$$f_G(\lambda,\mu,\gamma) = \phi_{\mathcal{S}}(\lambda) + 4\phi_{\mathcal{T}}(\mu) + 4\psi^2[G] \sum_i \gamma_i$$

with

$$\psi[G] = \max_i \pi(g_i) + \varkappa(\epsilon)\mathrm{n}[G].$$

Proposition 1. *Let* (λ,μ,γ) *be a feasible solution to the problem in* (8). *Then*

$$\mathrm{Risk}_\epsilon[\widehat{w}_G] \leqslant f_G(\lambda,\mu,\gamma),$$

i.e., the ϵ*-risk of the estimate* $\widehat{w}_G$ *is upper bounded with* $f_G(\lambda,\mu,\gamma)$.

Proof. Let us fix $\xi \in \Xi_\epsilon[G]$, $x_* \in \mathcal{X}$, and $\eta_* \in \mathcal{N}$. Let also $\widehat{x} = \widehat{x}_G(\omega)$ be the x-component of some optimal solution $[\widehat{x};\widehat{\nu}]$, $\widehat{\nu} \in \mathcal{N}$, to (5) and, finally, let $\Delta = \widehat{x} - x_*$. Observe that $[x,\nu] = [x_*,\nu_*]$ is feasible for (5) and $\|G^T[Ax_* + N\nu_* - \omega]\|_\infty = \|G^T\xi\|_\infty \leqslant \varkappa(\epsilon)\mathrm{n}[G]$, implying that $\|G^T[A\widehat{x} + N\widehat{\nu} - \omega]\|_\infty \leqslant \varkappa(\epsilon)\mathrm{n}[G]$ as well. Therefore,

$$\|G^T A\Delta\|_\infty \leqslant 2\varkappa(\epsilon)\mathrm{n}[G] + \|G^T N[\widehat{\nu} - \nu_*]\|_\infty.$$

Taking into account that $\widehat{\nu}, \nu_* \in \mathcal{N}$, we have $\|G^T N[\widehat{\nu} - \nu_*]\|_\infty \leqslant 2\max_i \pi(g_i)$, and we arrive at

$$|g_i^T A\Delta| \leqslant 2\psi[G],\ i = 1, \ldots, I. \tag{9}$$

Now, we have $\Delta \in 2\mathcal{X}$, that is, for some $t \in \mathcal{T}$ and all k it holds $\Delta^T T_k \Delta \leqslant 4t_k$, and let $v \in \mathcal{B}_*$, so that for some $s \in \mathcal{S}$ for all ℓ it holds $v^T S_\ell v \leqslant s_\ell$. By the semidefinite constraint of (8) we have

$$\begin{aligned} v^T B\Delta &\leqslant v^T \left[\sum_\ell \lambda_\ell S_\ell\right] v + \Delta^T \left[\sum_k \mu_k T_k\right] \Delta + [A\Delta]^T \sum_i \gamma_i g_i g_i^T A\Delta \\ &\leqslant \sum_\ell \lambda_\ell s_\ell + 4\sum_k \mu_k t_k + \sum_i \gamma_i (g_i^T A\Delta)^2 \\ &\leqslant \phi_{\mathcal{S}}(\lambda) + 4\phi_{\mathcal{T}}(\mu) + \sum_i \gamma_i (g_i^T A\Delta)^2 \\ &\leqslant \phi_{\mathcal{S}}(\lambda) + 4\phi_{\mathcal{T}}(\mu) + 4\psi^2[G] \sum_i \gamma_i. \end{aligned}$$

Maximizing the left-hand side of the resulting inequality over $v \in \mathcal{B}_*$, we arrive at $\|B\Delta\| \leqslant f_G(\lambda, \mu, \gamma)$. □

Note that the optimization problem in the right-hand side of (8) is an explicit convex optimization problem, so that Opt$[G]$ is efficiently computable, provided that $\phi_{\mathcal{S}}, \phi_{\mathcal{T}}$ and π are so. Thus, Proposition 1 provides us with an efficiently computable upper bound on the ϵ-risk of the polyhedral estimate stemming from a given contrast matrix G and as such gives us a computation-friendly tool to *analyse* the performance of a polyhedral estimate. Unfortunately, this tool does not allow to *design* a "presumably good" estimate, since an attempt to make G a variable, rather than a parameter, in the right-hand side problem in (8) results in a nonconvex, and thus difficult to solve, optimization problem. We now look at two situations in which this difficulty can be overcome.

2.2. Synthesis of "Presumably Good" Contrast Matrices

We consider here two types of assumptions about the set $\mathcal{N}$ of nuisances which allow for a computationally efficient design of "presumably good" contrast matrices. Namely,

1) "ellitopic case:" $\mathcal{N}$ is a basic ellitope;
2) "co-ellitopic case:" the set $N\mathcal{N} = \{N\nu : \nu \in \mathcal{N}\}$ is the *polar* of the ellitope

$$\begin{gathered} \mathcal{N}_* = \{w \in \mathbf{R}^m : \exists \overline{r} \in \overline{\mathcal{R}} : w^T \overline{R}_j w \leqslant \overline{r}_j, j \leqslant \overline{J}\} \\ \left[\ \overline{R}_j \succeq 0,\ \textstyle\sum_j \overline{R}_j \succ 0;\ \overline{\mathcal{R}} \subset \mathbf{R}^{\overline{J}}_+,\ \mathrm{int}\overline{\mathcal{R}} \neq \emptyset, \text{ is a monotone convex compact}\right] \end{gathered}$$

Note that $\mathcal{N}_*$ is exactly the unit ball of the norm $\pi(g) = \max_{\nu \in \mathcal{N}} g^T N\nu$.

<u>2.2.1. Ellitopic case.</u> An immediate observation is that *the ellitopic case can be immediately reduced to the no-nuisance case.* Indeed, when $\mathcal{N}$ is an ellitope, so is the direct product $\overline{\mathcal{X}} = \mathcal{X} \times \mathcal{N}$. Thus, setting $\overline{A}[x;\nu] = Ax + N\nu$, $\overline{B}[x;\nu] = Bx$, observation (2) becomes

$$\omega = \overline{A}\overline{x}_* + \xi, \qquad [\overline{x_*} = [x_*;\nu_*] \in \overline{\mathcal{X}}]$$

and our objective is to recover from this observation the linear image $w_* = \overline{B}\overline{x}_*$ of the new signal $\overline{x}_*$. Design of presumably good (and near-minimax-optimal when $\xi \sim \mathcal{N}(0, \sigma^2 I_m)$) polyhedral estimates in this setting is considered in [20]. It makes sense to sketch the construction here since it explains the idea used throughout the rest of the paper.

Thus, consider the case when $\mathcal{N} = \{0\}$, and let the signal set and the norm $\|\cdot\|$ still be given by (4). In this situation problem (8) becomes

$$\mathrm{Opt}[G] = \min_{\lambda,\mu,\gamma}\left\{\phi_{\mathcal{S}}(\lambda) + 4\phi_{\mathcal{T}}(\mu) + 4\varkappa^2(\epsilon)\mathrm{n}^2[G]\sum_i \gamma_i :\ \lambda \geqslant 0,\ \mu \geqslant 0,\ \gamma \geqslant 0,\right. \tag{10}$$
$$\left.\left[\begin{array}{c|c}\sum_\ell \lambda_\ell S_\ell & \frac{1}{2}B \\ \hline \frac{1}{2}B^T & \sum_k \mu_k T_k + A^T\left[\sum_i \gamma_i g_i g_i^T\right]A\end{array}\right] \succeq 0\right\}.$$

Note that when $\theta > 0$, we have $\mathrm{Opt}[G] = \mathrm{Opt}[\theta G]$. Indeed, (λ, μ, γ) is a feasible solution to the problem specifying $\mathrm{Opt}[G]$ if and only if $\lambda, \mu, \theta^2\gamma)$ is a feasible solution to the problem specifying $\mathrm{Opt}[\theta G]$, and the values of the respective objectives at these solutions are the same. It follows that as far as optimization of $\mathrm{Opt}[G]$ in G is concerned, we lose nothing when restricting ourselves to contrast matrices G with $\varkappa(\epsilon)\mathrm{n}[G] = 1$. In other words, by setting

$$\theta(g) = \varkappa(\epsilon)\|g\|_2 \tag{11}$$

and augmenting variables λ, μ, and γ in (10) by variables g_i, $\theta(g_i) \leqslant 1$, $i = 1, \ldots, I$ (recall that we want to make G variable rather than parameter and to minimize $\mathrm{Opt}[G]$ over G), we arrive at the problem

$$\mathrm{Opt} = \min_{\lambda,\mu,\gamma,\{g_i\},\rho}\left\{\phi_{\mathcal{S}}(\lambda) + 4\phi_{\mathcal{T}}(\mu) + 4\rho :\ \lambda \geqslant 0,\ \mu \geqslant 0,\ \gamma \geqslant 0,\right. \tag{12}$$
$$\left.\theta(g_i) \leqslant 1,\ \sum_i \gamma_i \leqslant \rho,\ \left[\begin{array}{c|c}\sum_\ell \lambda_\ell S_\ell & \frac{1}{2}B \\ \hline \frac{1}{2}B^T & \sum_k \mu_k T_k + A^T\left[\sum_i \gamma_i g_i g_i^T\right]A\end{array}\right] \succeq 0\right\}.$$

Now, aggregating variables γ, $g_1, \ldots, g_I$ into the matrix $\Theta = \sum_i \gamma_i g_i g_i^T$ and denoting by $\mathfrak{T}$ the set of the pairs $(\Theta \in \mathbf{S}^m_+, \rho)$ for which there exists decomposition $\Theta = \sum_{i\leqslant I}\gamma_i g_i g_i^T$ with $\theta(g_i) \leqslant 1$ and $\gamma_i \geqslant 0$, $\sum_i \gamma_i \leqslant \rho$, (12) can be rewritten as the optimization problem

$$\mathrm{Opt} = \min_{\lambda,\mu,\Theta,\rho}\left\{\phi_{\mathcal{S}}(\lambda) + 4\phi_{\mathcal{T}}(\mu) + 4\rho :\ \lambda \geqslant 0,\ \mu \geqslant 0,\right. \tag{13}$$
$$\left.(\Theta, \rho) \in \mathfrak{T},\ \left[\begin{array}{c|c}\sum_\ell \lambda_\ell S_\ell & \frac{1}{2}B \\ \hline \frac{1}{2}B^T & \sum_k \mu_k T_k + A^T\Theta A\end{array}\right] \succeq 0\right\}.$$

Observe that when $I \geqslant m$, $\mathfrak{T}$ is a simple convex cone:

$$\mathfrak{T} = \{(\Theta, \rho) :\ \Theta \succeq 0,\ \rho \geqslant \varkappa^2(\epsilon)\mathrm{Tr}(\Theta)\},$$

so that (13) is an explicit (and clearly solvable) convex optimization program. To convert an optimal solution $(\lambda^*, \mu^*, \Theta^*, \rho^*)$ to (13) into an optimal solution to (12), it suffices to subject Θ^* to the eigenvalue decomposition $\Theta^* = \sum_{i=1}^I \upsilon_i e_i e_i^T$ with $\|e_i\|_2 = 1$ and $\upsilon_i \geqslant 0$, $i \in \{1, \ldots, m\}$, and $e_i = 0$, $\upsilon_i = 0$, $i \in \{m, \ldots, I\}$, and set $g_i^* = \varkappa^{-1}(\epsilon)e_i$, $\gamma_i^* = \varkappa^2(\epsilon)\upsilon_i$, thus arriving at an optimal solution $(\lambda^*,\ \mu^*,\ \{g_i^*, \gamma_i^*\}_{i\leqslant I},\ \rho^*)$ to problem (12).

2.2.2. Co-ellitopic case. The just outlined approach to reducing the nonconvex problem (12) responsible for the design of the best, in terms of $\mathrm{Opt}[G]$, contrast matrix G to an explicit convex optimization problem heavily utilizes the fact that the unit ball of the norm $\theta(\cdot)$ (cf. (11)) is a simple ellitope—a multiple of the unit Euclidean ball; this was the reason for $\mathfrak{T}$ to be a computationally tractable convex cone. Our future developments are built on the fact that when the unit ball of $\theta(\cdot)$ is a basic ellitope, something similar takes place: the associated set $\mathfrak{T}$, while not necessarily convex and computationally tractable, can be tightly approximated by a computationally tractable convex cone. The underlying result (which is [21, Proposition 3.2], up to notation) is as follows:

Proposition 2. *Let $I \geqslant m$, and let $\mathcal{W} \subset \mathbf{R}^m$ be a basic ellitope:*

$$\mathcal{W} = \{w \in \mathbf{R}^m : \exists r \in \mathcal{R} : w^T R_j w \leqslant r_j,\ j \leqslant J\}$$
$$\Big[R_j \succeq 0,\ \textstyle\sum_j R_j \succ 0;\ \mathcal{R} \subset \mathbf{R}^J_+,\ \mathrm{int}\mathcal{R} \neq \emptyset, \text{ is a monotone convex compact}\Big]$$

Let us associate with $\mathcal{W}$ the closed convex cone[7]

$$\mathbf{K} = \{(\Theta, \rho) : \exists r \in \mathcal{R} : \mathrm{Tr}(\Theta R_j) \leqslant \rho r_j,\ 1 \leqslant j \leqslant J,\ \Theta \succeq 0,\ \rho \geqslant 0\}.$$

Whenever a matrix $\Theta \in \mathbf{S}^m_+$ is representable as $\sum_i \gamma_i w_i w_i^T$ with $\gamma_i \geqslant 0$ and $w_i \in \mathcal{W}$, one has $(\Theta, \sum_{i=1}^I \gamma_i) \in \mathbf{K}$, and nearly vice versa: whenever $(\Theta, \rho) \in \mathbf{K}$, one can find efficiently (via a randomized algorithm) vectors $w_i \in \mathcal{W}$, and reals $\gamma_i \geqslant 0$, $i \leqslant I$, such that $\Theta = \sum_i \gamma_i w_i w_i^T$ and and

$$\sum_i \gamma_i \leqslant 2\sqrt{2}\ln(4m^2 J)\rho.$$

We are now ready to outline a "presumably good" contrast design in the co-ellitopic case. Let us put $R_j = \frac{1}{4}\overline{R}_j$, $j \leqslant \overline{J}$, and $R_{\overline{J}+1} = \frac{\varkappa^2(\epsilon)}{4} I_m$ and consider the ellitope

$$\begin{aligned}\mathcal{W} &= 2\Big[\mathcal{N}_* \cap \{w : \varkappa(\epsilon)\|w\|_2 \leqslant 1\}\Big]\\ &= \Big\{w \in \mathbf{R}^m : \exists r \in \mathcal{R} = \overline{\mathcal{R}} \times [0,1] : w^T R_j w \leqslant r_j,\ j \leqslant J = \overline{J} + 1\Big\},\end{aligned} \tag{14}$$

and let $\theta(\cdot)$ be the norm on $\mathbf{R}^n$ with the unit ball $\mathcal{W}$. Note that $\theta(\cdot) = 2\max\left[\pi(\cdot), \varkappa(\epsilon)\|\cdot\|_2\right]$, so that for every $G = [g_1, .., g_I]$, the quantity $\psi[G]$, see (8), is upper-bounded by $\max_i \theta(g_i)$, and this bound is tight within the factor 2. Consequently, Proposition 1 states that the ϵ-risk of the polyhedral estimate with contrast matrix G is upper-bounded by the quantity

$$\begin{aligned}\overline{\mathrm{Opt}}[G] = \min_{\lambda,\mu,\gamma}\Bigg\{&\phi_{\mathcal{S}}(\lambda) + 4\phi_{\mathcal{T}}(\mu) + 4\Big[\max_i \theta(g_i)\Big]^2 \sum_i \gamma_i :\ \lambda \geqslant 0,\ \mu \geqslant 0,\ \gamma \geqslant 0,\\ &\left[\begin{array}{c|c}\sum_\ell \lambda_\ell S_\ell & \frac{1}{2}B\\ \hline \frac{1}{2}B^T & \sum_k \mu_k T_k + A^T\left[\sum_i \gamma_i g_i g_i^T\right]A\end{array}\right] \succeq 0\Bigg\}\end{aligned} \tag{15}$$

and $\mathrm{Opt}[G] \leqslant \overline{\mathrm{Opt}}[G] \leqslant 2\mathrm{Opt}[G]$. As in the previous section, the problem of minimizing $\overline{\mathrm{Opt}}[G]$ over G can be reformulated in the form (13). A computationally efficient way to get a tight approximation to the optimal solution of the latter problem is given by the following result.

Let $I \geqslant m$, $\alpha = 2\sqrt{2}\ln(4m^2 J)$, and let

$$\mathbf{K} = \{(\Theta, \rho) : \exists r \in \mathcal{R} : \mathrm{Tr}(\Theta R_j) \leqslant \rho r_j,\ 1 \leqslant j \leqslant J,\ \Theta \succeq 0,\ \rho \geqslant 0\}$$

(see (14)). Consider the convex optimization problem

$$\begin{aligned}\mathrm{Opt}_* = \min_{\lambda,\mu,\gamma,\Theta,\rho}\Bigg\{&\phi_{\mathcal{S}}(\lambda) + 4\phi_{\mathcal{T}}(\mu) + 4\alpha\rho :\ \lambda \geqslant 0,\ \mu \geqslant 0,\\ &(\Theta,\rho) \in \mathbf{K},\ \left[\begin{array}{c|c}\sum_\ell \lambda_\ell S_\ell & \frac{1}{2}B\\ \hline \frac{1}{2}B^T & \sum_k \mu_k T_k + A^T\Theta A\end{array}\right] \succeq 0\Bigg\}.\end{aligned} \tag{16}$$

[7] This indeed is a closed convex cone—the conic hull of the convex compact set $\{\Theta \succeq 0 : \exists r \in \mathcal{R} : \mathrm{Tr}(\Theta R_j) \leqslant r_j,\ 1 \leqslant j \leqslant J\} \times \{1\}$.

Theorem 1. *One can convert, in a computationally efficient way, the* Θ*-component* Θ^* *of an optimal solution to the (clearly solvable) problem* (16) *into the contrast matrix* G^* *such that*

$$\overline{\mathrm{Opt}}[G^*] \leqslant \sqrt{\alpha}\min_G \overline{\mathrm{Opt}}[G] \leqslant 2\sqrt{\alpha}\min_G \mathrm{Opt}[G].$$

In particular, the ϵ*-risk of the polyhedral estimate with contrast matrix* G^* (*this risk is upper-bounded by* $\overline{\mathrm{Opt}}[G^*]$) *does not exceed* $2\sqrt{\alpha}\min_G \mathrm{Opt}[G]$.

Proof. When repeating the reasoning in the previous section, we conclude that $\overline{\mathrm{Opt}} := \inf_G \overline{\mathrm{Opt}}[G]$ is equal to

$$\inf_{g_1,\ldots,g_I}\left\{\overline{\mathrm{Opt}}([g_1,\ldots,g_I]) :\ \max_i \theta(g_i) = 1\right\}.$$

The latter inf is clearly attained at certain collection $g_i^+,\ldots,g_I^+$ with $\max_i \theta(g_i^+) = 1$. Let $G^+ = [g_1^+,\ldots,g_I^+]$, let $\lambda^+,\mu^+,\gamma_i^+$, $i \leqslant I$, be an optimal solution to the problem in the right-hand side of (15) associated with $g_i = g_i^+$, $i \leqslant I$, and let $\Theta^+ = \sum_i \gamma_i^+[g_i^+][g_i^+]^T$ and $\rho^+ = \sum_i \gamma_i^+$. We clearly have

$$\overline{\mathrm{Opt}} = \overline{\mathrm{Opt}}[G^+] = \phi_{\mathcal{S}}(\lambda^+) + 4\phi_{\mathcal{T}}(\mu^+) + 4\rho^+.$$

Besides this, we are in the case where $\theta(g) \leqslant 1$ is equivalent to $g \in \mathcal{W}$, and therefore, by the first claim in Proposition 2, $(\Theta^+,\rho^+) \in \mathbf{K}$, implying that $(\lambda^+,\mu^+,\Theta^+,\rho^+)$ is a feasible solution to the optimization problem in (16). Due to the structure of the latter problem, for $\kappa > 0$ the collection $(\kappa^{-1}\lambda^+,\kappa\mu^+,\kappa\Theta^+,\kappa\rho^+)$ is feasible for (16) with the corresponding value of the objective $\kappa^{-1}\phi_{\mathcal{S}}(\lambda^+) + \kappa[\phi_{\mathcal{T}}(\mu^+) + 4\alpha\rho^+]$. It follows that

$$\begin{aligned}
\mathrm{Opt}_* &\leqslant \inf_{\kappa>0}\left[\kappa^{-1}\phi_{\mathcal{S}}(\lambda^+) + \kappa[4\phi_{\mathcal{T}}(\mu^+) + 4\alpha\rho^+]\right] \\
&= 2\Big(\phi_{\mathcal{S}}(\lambda^+)\underbrace{[4\phi_{\mathcal{T}}(\mu^+) + 4\alpha\rho^+]}_{\leqslant\alpha[4\phi_{\mathcal{T}}(\mu^+)+4\rho^+]}\Big)^{1/2} \leqslant 2\sqrt{\phi_{\mathcal{S}}(\lambda^+)[4\phi_{\mathcal{T}}(\mu^+) + 4\rho^+]}\sqrt{\alpha} \\
&\leqslant \sqrt{\alpha}[\phi_{\mathcal{S}}(\lambda^+) + 4\phi_{\mathcal{T}}(\mu^+) + 4\rho^+] = \sqrt{\alpha}\,\overline{\mathrm{Opt}}.
\end{aligned}$$

Finally, let $\overline{\lambda},\overline{\mu},\overline{\Theta},\overline{\rho}$ be an optimal solution to (16). As $(\overline{\Theta},\overline{\rho}) \in \mathbf{K}$, the second claim in Proposition 2 states that there exists (and can be efficiently found) decomposition $\overline{\Theta} = \sum_i \overline{\gamma}_i[\overline{g}_i][\overline{g}]_i^T$ with $\overline{g}_i \in \mathcal{W}$ (i.e., $\theta(\overline{g}_i) \leqslant 1$), $i \leqslant I$, $\overline{\gamma}_i \geqslant 0$, and $\sum_i \overline{\gamma}_i \leqslant \alpha\overline{\rho}$. The ϵ-risk of the polyhedral estimate with the contrast matrix $\overline{G} = [\overline{g}_1,\ldots,\overline{g}_I]$ is then upper-bounded by $\overline{\mathrm{Opt}}[\overline{G}]$. However, $\overline{\lambda},\overline{\mu}$, and $\{\overline{\gamma}_i\}$ form a feasible solution to the problem specifying $\overline{\mathrm{Opt}}[\overline{G}]$, and the value of the objective at this solution is upper bounded with

$$\phi_{\mathcal{S}}(\overline{\lambda}) + 4\phi_{\mathcal{T}}(\overline{\mu}) + 4[\max_i \theta(\overline{g}_i)]\sum_i \overline{\gamma}_i \leqslant \phi_{\mathcal{S}}(\overline{\lambda}) + 4\phi_{\mathcal{T}}(\overline{\mu}) + 4\alpha\overline{\rho} = \mathrm{Opt}_*.$$

Thus, the ϵ-risk of the polyhedral estimate with contrast matrix $\overline{G}$ does not exceed

$$\mathrm{Opt}_* \leqslant \sqrt{\alpha}\,\overline{\mathrm{Opt}} \leqslant 2\sqrt{\alpha}\min_G \mathrm{Opt}[G]. \qquad \square$$

3. OBSERVATIONS WITH OUTLIERS

In this section, we consider the estimation problem posed in Section 1.2 in the situation where the nuisance ν_* in (2) is sparse—has at most a given number s of nonzero entries.

Estimate construction. Let $\epsilon \in (0,1)$ be a given reliability tolerance. We consider the polyhedral estimate specified by two contrast matrices $H = [h_1,\ldots,h_n] \in \mathbf{R}^{m\times n}$ and $G = [g_1,\ldots,g_I] \in \mathbf{R}^{n\times I}$ which is as follows. Given observation ω (see (2)) we solve the optimization problem

$$\min_{\nu,x}\left\{\|\nu\|_1 :\ x \in \mathcal{X},\ \begin{array}{ll} |h_k^T[N\nu + Ax - \omega]| \leqslant \overline{\varkappa}(\epsilon)\|h_k\|_2, & k = 1,\ldots,n, \\ |g_i^T[N\nu + Ax - \omega]| \leqslant \overline{\varkappa}(\epsilon)\|g_i\|_2, & i = 1,\ldots,I \end{array}\right\}, \tag{17}$$

where

$$\overline{\varkappa}(\epsilon) = \sigma\sqrt{2\ln[2(n+I)/\epsilon]}.$$

Let $(\widehat{\nu}, \widehat{x}) = (\widehat{\nu}(\omega), \widehat{x}(\omega))$ be an optimal solution to the problem when the problem is feasible, otherwise we put $(\widehat{\nu}, \widehat{x}) = (0, 0)$. Vector

$$\widehat{w}_{G,H}(\omega) = B\widehat{x}(\omega)$$

is the estimate of $w_* = Bx_*$ we intend to use.

3.1. Risk Analysis

Let us denote $\Xi_\epsilon(G, H)$ the set of realizations of ξ such that

$$|h_k^T\xi| \leqslant \overline{\varkappa}(\epsilon)\|h_k\|_2,\ k = 1, \ldots, n, \quad |g_i^T\xi| \leqslant \overline{\varkappa}(\epsilon)\|g_i\|_2,\ i = 1, \ldots, I, \quad \forall\xi \in \Xi_\epsilon(G, H). \tag{18}$$

For the same reasons as in (7), one has

$$\mathrm{Prob}_{\xi\sim\mathcal{SG}(0,\sigma^2 I_m)}(\Xi_\epsilon(G, H)) \geqslant 1 - \epsilon.$$

Let us now fix $x_* \in \mathcal{X}$, s-sparse ν_*, and $\xi \in \Xi_\epsilon(G, H)$, so that our observation is $\omega = Ax_* + N\nu_* + \xi$.
A. By (18) we have $|h_k^T\xi| \leqslant \overline{\varkappa}(\epsilon)\|h_k\|_2$ and $|g_i^T\xi| \leqslant \overline{\varkappa}(\epsilon)\|g_i\|_2$ for all $k \leqslant n$ and $i \leqslant I$, while (17) becomes the problem

$$\min_{\nu,x}\left\{\|\nu\|_1 :\ x \in \mathcal{X},\ \begin{array}{ll} |h_k^T[N[\nu-\nu_*] + A[x-x_*] - \xi]| \leqslant \overline{\varkappa}(\epsilon)\|h_k\|_2, & k = 1, \ldots, n, \\ |g_i^T[N[\nu-\nu_*] + A[x-x_*] - \xi]| \leqslant \overline{\varkappa}(\epsilon)\|g_i\|_2, & i = 1, \ldots, I \end{array}\right\}. \tag{19}$$

We conclude that $(\nu, x) = (\nu_*; x_*)$ is a feasible solution to (17). Thus, we are in the case where $\widehat{\nu}$, $\widehat{x}$ are feasible for (19), and

$$\|\widehat{\nu}\|_1 \leqslant \|\nu_*\|_1.$$

B. Assume from now on that $(H, \|\cdot\|_\infty)$ satisfies Condition $Q_\infty(s, \kappa)$ of Section 3.5 with $\kappa < \frac{1}{2}$, that is,[8]

$$\|w\|_\infty \leqslant \|H^TNw\|_\infty + \frac{\kappa}{s}\|w\|_1 \quad \forall w \in \mathbf{R}^n. \tag{20}$$

Since $\widehat{\nu}$ and $\widehat{x}$ are feasible for (19), we have

$$|h_k^T[N[\widehat{\nu} - \nu_*] + A[\widehat{x} - x_*] - \xi]| \leqslant \overline{\varkappa}(\epsilon)\|h_k\|_2,\ \forall k \leqslant n.$$

Invoking (18) and the fact that $A[\widehat{x} - x_*] \in 2A\mathcal{X}$ (since $\mathcal{X}$ is symmetric w.r.t. the origin), we conclude that

$$\|H^TN[\widehat{\nu} - \nu_*]\|_\infty \leqslant \max_k\left[\overline{\varkappa}(\epsilon)\|h_k\|_2 + 2\max_{x\in\mathcal{X}}|h_k^TAx|\right],$$

and besides this, ν_* is s-sparse and $\|\widehat{\nu}\|_1 \leqslant \|\nu_*\|_1$. Now Proposition 5 with ν_* in the role of ν implies that

$$\|\widehat{\nu} - \nu_*\|_q \leqslant \frac{(2s)^{\frac{1}{q}}}{1 - 2\kappa}\max_k\left[\overline{\varkappa}(\epsilon)\|h_k\|_2 + 2\max_{x\in\mathcal{X}}|h_k^TAx|\right], \quad 1 \leqslant q \leqslant \infty, \tag{21}$$

[8] Condition $Q_\infty(s, \kappa)$ is the simplest (and the most restrictive) member of the family $Q_q(s, \kappa)$, $q \in [1, \infty]$ of conditions used to establish the properties of ℓ_1-recovery of sparse signals. The property of this condition crucial here is that it can be efficiently verified. We refer to [20, Section 1.3] for the discussion of efficiently verifiable conditions in sparse recovery and their relation to other conditions used (Restricted Isometry Property (RIP) [4], Restricted Eigenvalue (RE) [2], Mutual Incoherence (MI) [11], and Compatibility [40]).

in particular, that

$$\|\widehat{\nu}-\nu_*\|_\infty \leqslant \frac{1}{1-2\kappa}\max_k\left[\overline{\varkappa}(\epsilon)\|h_k\|_2+2\max_{x\in\mathcal{X}}|h_k^TAx\|\right] =: \rho_H, \tag{22a}$$

$$\|\widehat{\nu}-\nu_*\|_1 \leqslant \frac{2s}{1-2\kappa}\max_k\left[\overline{\varkappa}(\epsilon)\|h_k\|_2+2\max_{x\in\mathcal{X}}|h_k^TAx\|\right] = 2s\rho_H. \tag{22b}$$

In addition, [20, Proposition 1.10] states that the set $\mathcal{H}$ of the pairs (H,κ) with $m\times n$ matrices H satisfying Condition $Q_\infty(s,\kappa)$ is the computationally tractable convex set

$$\mathcal{H}=\left\{(H,\kappa)\in\mathbf{R}^{m\times n}\times\mathbf{R}:\ |[I_n-N^TH]_{ij}|\leqslant s^{-1}\kappa,\ 1\leqslant i,j\leqslant n\right\}. \tag{23}$$

C. Since $\widehat{\nu}$ and $\widehat{x}$ are feasible for (19), we have

$$|g_i^T\left(N[\widehat{\nu}-\nu_*]+A[\widehat{x}-x_*]-\xi\right)|\leqslant\overline{\varkappa}(\epsilon)\|g_i\|_2,\quad i=1,\ldots,I,$$

while $|g_i^T\xi|\leqslant\overline{\varkappa}(\epsilon)\|g_i\|_2\ \forall i$ due to $\xi\in\Xi_\epsilon(G,H)$. We conclude that

$$|g_i^TA[\widehat{x}-x_*]|\leqslant 2\overline{\varkappa}(\epsilon)\|g_i\|_2+|g_i^TN[\widehat{\nu}-\nu_*]|,\ i\leqslant I. \tag{24}$$

Let $\|z\|_{k,1}$, $z\in\mathbf{R}^n$, be the sum of $\min[k,n]$ largest magnitudes of entries in z; note that $\|\cdot\|_{k,1}$ is the norm conjugate to the norm with the unit ball $\{u:\|u\|_\infty\leqslant 1,\|u\|_1\leqslant k\}$. Consequently, (22) implies that

$$|g_i^TN[\widehat{\nu}-\nu_*]|\leqslant\rho_H\|N^Tg_i\|_{2s,1}, \tag{25}$$

and, therefore, by (24)

$$|g_i^TA[\widehat{x}-x_*]|\leqslant\psi_H[G],\ \ \psi_H[G]=\max_i\left[2\overline{\varkappa}(\epsilon)\|g_i\|_2+\rho_H\|N^Tg_i\|_{2s,1}\right]. \tag{26}$$

Let

$$f_{G,H}(\lambda,\mu,\gamma)=\phi_{\mathcal{S}}(\lambda)+4\phi_{\mathcal{T}}(\mu)+\psi_H^2[G]\sum_i\gamma_i,$$

and let us consider the optimization problem (cf. (8))

$$\mathrm{Opt}[G,H]=\min_{\lambda,\mu,\gamma}\left\{f_{G,H}(\lambda,\mu,\gamma):\right.$$
$$\left.\lambda\geqslant 0,\ \mu\geqslant 0,\ \gamma\geqslant 0,\ \left[\begin{array}{c|c}\sum_\ell\lambda_\ell S_\ell & \frac{1}{2}B\\ \hline \frac{1}{2}B^T & \sum_k\mu_kT_k+A^T\left[\sum_i\gamma_ig_ig_i^T\right]A\end{array}\right]\succeq 0.\right\} \tag{27}$$

Applying the same argument as in the proof of Proposition 1, with (26) in the role of (9), we arrive at the following result:

Proposition 3. *In the situation of this section given* $\kappa\in(0,1/2)$ *and* $m\times n$ *matrix* H *satisfying* $(H,\kappa)\in\mathcal{H}$, *see* (23), *let* (λ,μ,γ) *be a feasible solution to* (27). *Then*

$$\mathrm{Risk}_\epsilon[\widehat{w}_{G,H}]\leqslant f_{G,H}(\lambda,\mu,\gamma).$$

3.2. Synthesis of Contrast Matrices

Our present objective is to design contrast matrices H and G with a small value of the bound $\mathrm{Opt}[G,H]$ for the ϵ-risk of the estimate $\widehat{w}_{G,H}$.

D. Building the contrast matrix $H \in \mathbf{R}^{m\times n}$ is straightforward: the risk bound $\mathrm{Opt}[G,H]$, depends on $H=[h_1,\dots,h_n]$ solely through the quantity

$$\rho_H = \frac{1}{1-2\kappa}\max_{k\leqslant n}\left[\overline{\varkappa}(\epsilon)\|h_k\|_2 + 2\max_{x\in\mathcal{X}}\|h_k^TAx\|_\infty\right]$$

and is smaller the smaller is ρ_H. For a fixed $\kappa\in(0,1/2)$, a presumably good choice of $H=[h_1,\dots,h_n]$ is then given by optimal solutions to n optimization problems

$$h_k = \operatorname*{argmin}_h\left\{\overline{\varkappa}(\epsilon)\|h\|_2 + 2\max_{x\in\mathcal{X}}\|h^TAx\|_\infty:\ h\in\mathbf{R}^m,\ \|\mathrm{Col}_i[I_n-N^Th]\|_\infty\leqslant s^{-1}\kappa\right\} \tag{28}$$

which, when recalling what $\mathcal{X}$ is, by conic duality, are equivalent to the problems

$$h_k = \operatorname*{argmin}_{h,v,\chi}\left\{\overline{\varkappa}(\epsilon)\|h\|_2 + v + \phi_{\mathcal{T}}(\chi):\ h\in\mathbf{R}^m, \chi\geqslant 0,\right.$$
$$\left.\left[\begin{array}{c|c} v & h^TA\\ \hline A^Th & \sum_k\chi_kT_k\end{array}\right]\succeq 0,\ \|\mathrm{Col}_i[I_n-N^Th]\|_\infty\leqslant s^{-1}\kappa\right\},\quad 1\leqslant k\leqslant n.$$

E. The proposed construction of G is less straightforward. We proceed as follows. Let $G=[G_1,G_2]$ where $G_2,G_1\in\mathbf{R}^{m\times m}$ (so that $I=2m$).

E.1 Notice that as $\xi\in\Xi_\epsilon(G,H)$, problem (19) is feasible, and $(\widehat{x},\widehat{\nu})$ is its feasible solution. For a column g of G, by the constraints of the problem, we have

$$|g^TA[\widehat{x}-x_*]|\leqslant 2\overline{\varkappa}(\epsilon)\|g\|_2 + |g^TN[\widehat{\nu}-\nu_*]|\leqslant 2\overline{\varkappa}(\epsilon)\|g\|_2 + 2s\rho_H\|N^Tg\|_\infty, \tag{29}$$

(we have used (24) and (25)), implying that

$$\left(g^TA[\widehat{x}-x_*]\right)^2\leqslant 2\left(4\overline{\varkappa}^2(\epsilon)\|g\|_2^2 + 4s^2\rho_H^2\|N^Tg\|_\infty^2\right),\quad i=1,\dots,m. \tag{30}$$

Note that the set

$$\mathcal{M}=\left\{g\in\mathbf{R}^m:\ 8\overline{\varkappa}^2(\epsilon)\|g\|_2^2+8s^2\rho_H^2\|N^Tg\|_\infty^2\leqslant 1\right\}$$

is an ellitope: when denoting $N=[\mathrm{n}_1,\dots,\mathrm{n}_n]$ we have

$$\mathcal{M}=\left\{g\in\mathbf{R}^m:\ \exists r\in[0,1]^n:\ g^T\underbrace{\left(8\overline{\varkappa}^2(\epsilon)I_m+8s^2\rho_H^2\mathrm{n}_j\mathrm{n}_j^T\right)}_{M_j}g\leqslant r_j,\ j=1,\dots,n\right\}.$$

E.2 Next, observe that when $\xi\in\Xi_\epsilon(G,H)$, by (21) one has

$$\|\widehat{\nu}-\nu_*\|_2\leqslant\frac{\sqrt{2s}}{1-2\kappa}\max_{k\leqslant n}\left[\overline{\varkappa}(\epsilon)\|h_k\|_2+\max_{x\in\mathcal{X}}|h_k^TAx|\right]=\sqrt{2s}\rho_H.$$

Then by (29), for a column g of G it holds

$$\begin{aligned}\left(g^TA[\widehat{x}-x_*]\right)^2&\leqslant\left(2\overline{\varkappa}(\epsilon)\|g\|_2+|g^TN[\widehat{\nu}-\nu_*]|\right)^2\leqslant\left(2\overline{\varkappa}(\epsilon)\|g\|_2+\sqrt{2s}\rho_H\|N^Tg\|_2\right)^2\\&\leqslant g^T\left(8\overline{\varkappa}^2(\epsilon)I_m+4s\rho_H^2NN^T\right)g.\end{aligned} \tag{31}$$

Now, let us put

$$Q = \left(8\overline{\varkappa}^2(\epsilon)I_m + 4s\rho_H^2 NN^T\right)^{-1/2}, \tag{32}$$

and consider the optimization problem

$$\text{Opt} = \min_{\lambda,\mu,\Theta_1,\Theta_2,\rho} \Bigg\{ f_H(\lambda,\mu,\Theta_1,\Theta_2,\rho):\ \lambda \geqslant 0,\ \mu \geqslant 0,\ \Theta_1 \succeq 0,\ \Theta_2 \succeq 0, \tag{33a}$$

$$\text{Tr}(M_j\Theta_1) \leqslant \rho,\ j = 1,\ldots,n,\ \left[\begin{array}{c|c} \sum_\ell \lambda_\ell S_\ell & \frac{1}{2}B \\ \hline \frac{1}{2}B^T & \sum_k \mu_k T_k + A^T(\Theta_1 + Q\Theta_2 Q^T)A \end{array}\right] \succeq 0 \Bigg\}$$

where

$$f_H(\lambda,\mu,\Theta_1,\Theta_2,\rho) = \phi_{\mathcal{S}}(\lambda) + 4\phi_{\mathcal{T}}(\mu) + \text{Tr}(\Theta_2) + 2\sqrt{2}\ln(4m^2n)\rho. \tag{33b}$$

Note that the constraints on Θ_1 and ρ of the problem (33a) say exactly that (Θ_1,ρ) belongs to the cone $\mathbf{K}$ associated, as explained in Proposition 2, with the ellitope $\mathcal{M}$ in the role of $\mathcal{W}$.

Theorem 2. *Given a feasible solution $(\lambda,\mu,\tau,\Theta_1,\Theta_2)$ to* (33)*, let us build $m\times m$ contrast matrices G_1, G_2 as follows.*

- *To build G_1, we apply the second part of Proposition 2 to $\Theta_1,\rho,\mathcal{M}$ in the roles of $\Theta,\rho,\mathcal{W}$, to get, in a computationally efficient way, a decomposition $\Theta_1 = \sum_{i=1}^m \gamma_i g_{1,i}g_{1,i}^T$ with $g_{1,i} \in \mathcal{M}$ and $\gamma_i \geqslant 0$, $\sum_i \gamma_i \leqslant 2\sqrt{2}\ln(4m^2n)\rho$. We set $G_1 = [g_{1,1},\ldots,g_{1,m}]$.*
- *To build G_2, we subject Θ_2 to eigenvalue decomposition $\Theta_2 = \Gamma\text{Diag}\{\chi\}\Gamma^T$ and set $G_2 = [g_{2,1},\ldots,g_{2,m}] = Q\Gamma$.*

Note that $\Theta_1 + Q\Theta_2 Q = \sum_i \gamma_i g_{1,i}g_{1,i}^T + \sum_i \chi_i g_{2,i}g_{2,i}^T$.

For the resulting polyhedral estimate $\widehat{w}_{G,H}$ and for all $x_* \in \mathcal{X}$, s-sparse ν_* and $\xi \in \Xi_\epsilon(G,H)$ it holds

$$\|\widehat{w}_{G,H}(Ax_* + N\nu_* + \xi) - Bx_*\| \leqslant f_H(\lambda,\mu,\Theta_1,\Theta_2,\rho) \tag{34}$$

implying that the ϵ-risk of the estimate is upper-bounded by $f_H(\lambda,\mu,\Theta_1,\Theta_2,\rho)$ (due to $\xi \in \Xi_\epsilon(G,H)$ with probability $\geqslant 1-\epsilon$).

Proof. Let us fix $x_* \in \mathcal{X}$, s-sparse ν_*, $\xi \in \Xi_\epsilon(G,H)$, and let $w = Ax_* + N\nu_* + \xi$. By **A**, problem (17) is feasible, so that $(\widehat{x},\widehat{\nu}) = (\widehat{x}(\omega),\widehat{\nu}(\omega))$ is its optimal solution, and $\widehat{w} = B\widehat{x}$ is the estimate $\widehat{w}_{G,H}(\omega)$. Let $\Delta = \widehat{x} - x_*$, and let $e_1,\ldots,e_m$ be the columns of the orthonormal matrix Γ. By construction of G_2, we have for all $j \leqslant m$ (see (31))

$$(g_{2,j}^T A\Delta)^2 \leqslant g_{2,j}^T\left(8\overline{\varkappa}^2(\epsilon)I_m + 4s\rho_H^2 NN^T\right)g_{2,j} = e_j^T\left[Q\left(8\overline{\varkappa}^2(\epsilon)I_m + 4s\rho_H^2 NN^T\right)Q\right]e_j = e_j^T e_j = 1.$$

Furthermore, due to $g_{1,i} \in \mathcal{M}$ one has (see (30))

$$\left(g_{1,i}^T A\Delta\right)^2 \leqslant 8\overline{\varkappa}^2(\epsilon)\|g\|_2^2 + 8s^2\rho_H^2\|N^T g\|_\infty^2 \leqslant 1 \quad \forall i \leqslant m.$$

Now, by the semidefinite constraint of (33a) and due to $\Theta_1 + Q\Theta_2 Q = \sum_i \gamma_i g_{1,i}g_{1,i}^T + \sum_i \chi_i g_{2,i}g_{2,i}^T$, for every $v \in \mathcal{B}_*$ we have

$$\begin{aligned}
v^T B\Delta &\leqslant v^T\left[\sum_\ell \lambda_\ell S_\ell\right]v + \Delta^T\left[\sum_k \mu_k T_k\right]\Delta + [A\Delta]^T\left[\sum_i \gamma_i g_{1,i}g_{1,i}^T + \sum_i \chi_i g_{2,i}g_{2,i}^T\right]A\Delta \\
&\leqslant \phi_{\mathcal{S}}(\lambda) + 4\phi_{\mathcal{T}}(\mu) + \sum_i \chi_i (g_{1,i}^T A\Delta)^2 + \sum_j \gamma_j (g_{2,j}^T A\Delta)^2 \\
&\quad \Big[\text{as } [v^T S_1 v;\ldots;v^T S_L v] \in \mathcal{S} \text{ due to } v \in \mathcal{B}_* \text{ and } [\Delta^T T_1\Delta;\ldots;\Delta^T T_L\Delta] \in 4\mathcal{T} \text{ due to } \Delta \in 2\mathcal{X}\Big] \\
&\leqslant \phi_{\mathcal{S}}(\lambda) + 4\phi_{\mathcal{T}}(\mu) + \sum_i \chi_i + \sum_j \gamma_j \leqslant f_H(\lambda,\mu,\tau,\Theta_1,\Theta_2)
\end{aligned}$$

due to $\sum_i \gamma_i \leqslant 2\sqrt{2}\ln(4m^2n)\rho$ and $\sum_i \chi_i = \mathrm{Tr}(\Theta_2)$. Taking the supremum over $v \in \mathcal{B}_*$ in the resulting inequality, we arrive at (34). □

3.3. An Alternative

Our objective in this section is to refine risk bounds (27) and (33a) to produce more efficient contrasts. Our course of action is as follows. First, to extend the possible choice of H-contrasts" responsible" for the perturbation recovery, we refine the bounds (22) for the accuracy of sparse recovery, notably, to allow using contrasts not satisfying Condition $Q_\infty(s,\kappa)$. Second, we improve the bounding of the risk of the estimate $\widehat{w}(\omega)$ by taking into account the contribution of the H-component of the "complete" contrast matrix $[H, G]$ when optimizing the G-component of the contrast.

In the sequel, we consider the estimate described at the beginning of Section 3, the only difference being in the sizes of contrast matrices G and H: now $H = [h_1, \ldots, h_M] \in \mathbf{R}^{m\times M}$, and $G = [g_1, \ldots, g_{2m}]$. Thus, in our present setting, given observation ω, we solve the optimization problem

$$\min_{\nu,x}\left\{\|\nu\|_1 : x \in \mathcal{X},\ \begin{array}{ll} |h_k^T(N\nu + Ax - \omega)| \leqslant \overline{\varkappa}(\epsilon)\|h_k\|_2, & k = 1,\ldots,M, \\ |g_i^T(N\nu + Ax - \omega)| \leqslant \overline{\varkappa}(\epsilon)\|g_{2,i}\|_2, & i = 1,\ldots,2m, \end{array}\right\} \tag{35}$$

with

$$\overline{\varkappa}(\epsilon) = \sigma\sqrt{2\ln[(2M + 4m)/\epsilon]},$$

specify $\widehat{x}(\omega), \widehat{\nu}(\omega)$ as an optimal solution to the problem when the problem is feasible, otherwise set $(\widehat{x}(\omega), \widehat{\nu}(\omega)) = (0, 0)$, and take $\widehat{w}_{G,H}(\omega) = B\widehat{x}(\omega)$ as the estimate of Bx_*.

3.3.1. Risk analysis. The above problem can be rewritten equivalently as

$$\min_{\nu,x}\left\{\|\nu\|_1 : x \in \mathcal{X},\ \begin{array}{ll} |h_k^T(N[\nu - \nu_*] + A[x - x_*] - \xi)| \leqslant \overline{\varkappa}(\epsilon)\|h_k\|_2, & k = 1,\ldots,M, \\ |g_i^T(N[\nu - \nu_*] + A[x - x_*] - \xi)| \leqslant \overline{\varkappa}(\epsilon)\|g_i\|_2, & i = 1,\ldots,2m, \end{array}\right\} \tag{36}$$

and when setting

$$\Xi_\epsilon(G, H) := \left\{\xi \in \mathbf{R}^m :\ \begin{array}{ll} |h_k^T\xi| \leqslant \overline{\varkappa}(\epsilon)\|h_k\|_2, & k = 1,\ldots,M, \\ |g_i^T\xi| \leqslant \overline{\varkappa}(\epsilon)\|g_i\|_2, & i = 1,\ldots,2m, \end{array}\right\} \tag{37}$$

we have

$$\mathrm{Prob}_{\xi\sim\mathcal{SG}(0,\sigma^2 I_m)}(\Xi_\epsilon(G, H)) \geqslant 1 - \epsilon.$$

Let us fix $\xi \in \Xi_\epsilon(G, H)$ and set $\omega = Ax_* + N\nu_* + \xi$. As $(\widehat{\nu}, \widehat{x})$ is a feasible for (36), $\widehat{x} := \widehat{x}(\omega)$, $\widehat{\nu} := \widehat{\nu}(\omega)$ is feasible as well, $\|\widehat{\nu}\|_1 \leqslant \|\nu_*\|_1$. Thus, same as in the proof of Proposition 5, for $z = \widehat{\nu} - \nu_*$ it holds

$$\|z\|_1 \leqslant 2\|z\|_{s,1}$$

implying that

$$\|z\|_1 \leqslant 2s\|z\|_\infty, \quad \|z\|_2 \leqslant \sqrt{2s}\|z\|_\infty. \tag{38}$$

Now denote $\Delta = \widehat{x} - x_*$, and consider n pairs of convex optimization problems

$$\mathrm{Opt}_2[i] = \max_{v,t,w}\left\{\sqrt{w_i t} : v \in 2\mathcal{X},\ \begin{array}{l} \|w\|_\infty \leqslant w_i,\ \|w\|_1 \leqslant t,\ t \leqslant 2sw_i, \\ |h_k^T(Nw + Av)| \leqslant 2\overline{\varkappa}(\epsilon)\|h_k\|_2,\ k = 1,\ldots,M \end{array}\right\} \tag{$P_2[i]$}$$

$$\mathrm{Opt}_\infty[i] = \max_{v,w}\left\{w_i : v \in 2\mathcal{X},\ \begin{array}{l} \|w\|_\infty \leqslant w_i,\ \|w\|_1 \leqslant 2sw_i, \\ |h_k^T(Nw + Av)| \leqslant 2\overline{\varkappa}(\epsilon)\|h_k\|_2,\ k = 1,\ldots,M \end{array}\right\}. \tag{$P_\infty[i]$}$$

Observe that a feasible solution (v,t,w) to $(P_2[i])$ satisfies $\|w\|_\infty \leqslant w_i$ and $\|w\|_1 \leqslant t$, whence

$$\|w\|_2 \leqslant \sqrt{w_i t} \leqslant \mathrm{Opt}_2[i]. \tag{39}$$

Now, let $\iota = \iota_z$ be the index of the largest in magnitude entry in z. Taking into account that $\xi \in \Xi_\epsilon(G,H)$ and recalling that $\Delta \in 2\mathcal{X}$, we conclude that when $z_\iota \geqslant 0$, $(v,t,w) = (\Delta, \|z\|_1, z)$ is feasible for $(P_2[\iota])$ and $(v,w) = (\Delta, z)$ is feasible for $(P_\infty[\iota])$, while when $z_\iota < 0$ the same holds true for $(v,t,w) = (-\Delta, \|z\|_1, -z)$ and $(v,w) = (-\Delta, -z)$. Indeed in the first case $v = \Delta \in \mathcal{X}$, $|h_k^T[A\widehat{x} + N\widehat{\nu} - \omega]| \leqslant \overline{\varkappa}(\epsilon)\|h_k\|_2$ and $|h_k^T[Ax_* + N\nu_* - \omega]| \leqslant \overline{\varkappa}(\epsilon)\|h_k\|_2$ as both pairs $(\widehat{x}, \widehat{\nu})$ and (x_*, ν_*) are feasible for (35), implying the second line constraints of $(P_2[i])$. Note that we are in the case of $z_\iota = \|z\|_\infty$, that is, constraints in the first line of $(P_2[i])$ are satisfied for $w = z$ due to (38). Thus, $(\Delta, \|z\|_1, z)$ indeed is feasible for $(P_2[i])$. As a byproduct of our reasoning, (Δ, z) is feasible for $(P_\infty[i])$. In the second case, the reasoning is completely similar.

Next, setting

$$\mathrm{Opt}_2 = \max_i \mathrm{Opt}_2[i], \quad \mathrm{Opt}_\infty = \max_i \mathrm{Opt}_\infty[i], \tag{40}$$

and recalling that $(\Delta, \|z\|_1, z)$ or $(-\Delta, \|z\|_1, -z)$ is feasible for some of the problems $(P_2[i])$, and (Δ, z) or $(-\Delta, -z)$ is feasible for some of the problems $(P_\infty[i])$, when invoking (39) we get for all $\xi \in \Xi_\epsilon(G,H)$

$$\|z\|_\infty \leqslant \mathrm{Opt}_\infty, \quad \|z\|_2 \leqslant \mathrm{Opt}_2, \quad \|z\|_1 \leqslant 2s\mathrm{Opt}_\infty.$$

Consequently, for all $d \in \mathbf{R}^m$

$$\begin{aligned} |d^T Nz| &\leqslant \max_z \left\{ d^T Nz : \|z\|_\infty \leqslant \mathrm{Opt}_\infty, \|z\|_2 \leqslant \mathrm{Opt}_2, \|z\|_1 \leqslant 2s\mathrm{Opt}_\infty \right\} \\ &= \underbrace{\min_{u,v,w} \left\{ \|u\|_1 \mathrm{Opt}_\infty + \|v\|_2 \mathrm{Opt}_2 + 2s\|w\|_\infty \mathrm{Opt}_\infty, \, u+v+w = N^T d \right\}}_{=:\overline{\pi}(N^T d)}. \end{aligned} \tag{41}$$

Now, recalling that $\widehat{x}, \widehat{\nu}$ is feasible for (36) and that $\xi \in \Xi_\epsilon(G,H)$, we conclude that columns d_i, $i = 1, \ldots, M+2m$ of the *"aggregated" contrast matrix* $D = [G,H]$ satisfy

$$|d_i^T A\Delta| \leqslant |d_i^T Nz| + |d_i^T \xi| + \overline{\varkappa}(\epsilon)\|g\|_2,$$

whence

$$|{d_i}^T A\Delta| \leqslant \underbrace{\overline{\pi}(N^T d_i) + 2\overline{\varkappa}(\epsilon)\|d_i\|_2}_{=:\psi_H(d_i)}, \quad i \leqslant M+2m. \tag{42}$$

Next, let us put

$$\bar{f}_{G,H}(\lambda,\mu,\gamma) = \phi_{\mathcal{S}}(\lambda) + 4\phi_{\mathcal{T}}(\mu) + \sum_i \gamma_i \psi_H^2(d_i),$$

and consider optimization problem (cf. (27))

$$\begin{aligned} \mathrm{Opt}[G,H] = \min_{\lambda,\mu,\gamma} \Bigg\{ & \bar{f}_{G,H}(\lambda,\mu,\gamma) : \lambda \geqslant 0, \; \mu \geqslant 0, \; \gamma \geqslant 0, \\ & \left[\begin{array}{c|c} \sum_\ell \lambda_\ell S_\ell & \frac{1}{2}B \\ \hline \frac{1}{2}B^T & \sum_k \mu_k T_k + A^T \left[\sum_i \gamma_i d_i {d_i}^T \right] A \end{array} \right] \succeq 0. \Bigg\} \end{aligned} \tag{43}$$

Applying the same argument as in the proof of Proposition 1, with (42) in the role of (9), we arrive at the following result:

Proposition 4. *In the situation of this section, let* (λ,μ,γ) *be a feasible solution to* (43)*. Then*

$$\mathrm{Risk}_\epsilon[\widehat{w}_{G,H}|\mathcal{X}, \mathcal{N}] \leqslant \bar{f}_{G,H}(\lambda,\mu,\gamma).$$

3.3.2. Contrast matrix synthesis. We continue our analysis of the estimate $\widehat{w}_{G,H}$ in the situation when the observation is $\omega = Ax_* + N\nu_* + \xi$ with $\xi \in \Xi_\epsilon(G, H)$, see (37). By (41), for $z = \widehat{\nu} - \nu_*$ and all $g \in \mathbf{R}^m$ we have

$$|g^T Nz| \leqslant \min\left\{\|N^T g\|_2 \mathrm{Opt}_2,\ \sqrt{2s}\|N^T g\|_2 \mathrm{Opt}_\infty,\ 2s\|N^T g\|_\infty \mathrm{Opt}_\infty\right\}$$

what implies (cf. (42)) that for all $i \leqslant 2m$

$$|g_i^T A\Delta| \leqslant 2\overline{\varkappa}(\epsilon)\|g_i\|_2 + \min\left\{\|N^T g_i\|_2 \mathrm{Opt}_2,\ \sqrt{2s}\|N^T g_i\|_2 \mathrm{Opt}_\infty,\ 2s\|N^T g_i\|_\infty \mathrm{Opt}_\infty\right\}. \tag{44}$$

Note that the right-hand side in (44) is nonconvex in g, making our design techniques inapplicable. To circumvent this difficulty, we intend to utilize the following important feature of polyhedral estimates: one may easily "aggregate" several estimates of this type to yield an estimate with the risk which is nearly as small as the smallest of the risks of the estimates combined.

Here is how it works in the present setting. We split the $m \times 2m$ contrast G into two $m \times m$ blocks $G_\chi = [g_{\chi,1}, \ldots, g_{\chi,m}]$, $\chi = 1, 2$, and design the blocks utilizing the respective inequalities inherited from (44), specifically, the inequalities

$$\begin{aligned} |g_{1,i}^T A\Delta| &\leqslant 2\overline{\varkappa}(\epsilon)\|g_{1,i}\|_2 + 2s\|N^T g_{1,i}\|_\infty \mathrm{Opt}_\infty, \\ |g_{2,i}^T A\Delta| &\leqslant 2\overline{\varkappa}(\epsilon)\|g_{2,i}\|_2 + \|N^T g_{2,i}\|_2 \underbrace{\min\{\mathrm{Opt}_2,\ \sqrt{2s}\mathrm{Opt}_\infty\}}_{=:\varrho_{2,H}} \end{aligned}$$

where $\Delta = \widehat{x} - x_*$. We weaken these inequalities to

$$\begin{aligned} |g_{1,i}^T A\Delta|^2 &\leqslant \pi_1^2(g_{1,i}), \quad \pi_1(g) = \sqrt{8\overline{\varkappa}^2(\epsilon)\|g\|_2^2 + 8s^2 \mathrm{Opt}_\infty^2 \|N^T g\|_\infty^2}, \\ |g_{2,i}^T A\Delta|^2 &\leqslant \pi_2^2(g_{2,i}), \quad \pi_2(g) = \sqrt{8\overline{\varkappa}^2(\epsilon)\|g\|_2^2 + 2\varrho_{2,H}^2 \|N^T g\|_2^2}. \end{aligned}$$

Notice that norms π_χ, $\chi = 1, 2$, are ellitopic, so we can use in our present situation the techniques from Section 3.2, thus arriving at an analogue of Theorem 2. To this end, denote by $\mathrm{n}_1, \ldots, \mathrm{n}_n$ the columns of N and set

$$\overline{M}_j = 8\overline{\varkappa}^2(\epsilon) I_m + 8s^2 \mathrm{Opt}_\infty^2 \mathrm{n}_j \mathrm{n}_j^T,\ j \leqslant m, \ \text{ and } \ \overline{Q} = \left(8\overline{\varkappa}^2(\epsilon) I_m + 2\varrho_{2,H}^2 NN^T\right)^{-1/2}.$$

Next, observe that the unit ball of $\pi_1(\cdot)$ is the ellitope

$$\overline{\mathcal{M}} = \left\{w \in \mathbf{R}^m : \exists r \in [0,1]^M : w^T \overline{M}_j w \leqslant \rho_j,\ j = 1, \ldots, M\right\}$$

and the unit ball of π_2 is the ellipsoid $w^T \overline{Q}^{-2} w \leqslant 1$. Now, let us consider the optimization problem

$$\mathrm{Opt} = \min_{\lambda,\mu,\tau,\Theta_1,\Theta_2,\rho} \left\{ \overline{f}_H(\lambda, \mu, \tau, \Theta_1, \Theta_2, \rho) :\ \lambda \geqslant 0,\ \mu \geqslant 0,\ \tau \geqslant 0, \right. \tag{45a}$$

$$\left. \begin{array}{l} \Theta_1 \succeq 0,\ \Theta_2 \succeq 0,\ \mathrm{Tr}(\overline{M}_j \Theta_1) \leqslant \rho,\ j = 1, \ldots, n, \\ \left[\begin{array}{c|c} \sum_\ell \lambda_\ell S_\ell & \frac{1}{2}B \\ \hline \frac{1}{2}B^T & \sum_k \mu_k T_k + A^T\left[\sum_i \tau_i h_i h_i^T\right]A + A^T(\Theta_1 + Q\Theta_2 Q^T)A \end{array}\right] \succeq 0 \end{array} \right\}$$

where

$$\overline{f}_H(\lambda, \mu, \tau, \Theta_1, \Theta_2, \rho) = \phi_{\mathcal{S}}(\lambda) + 4\phi_{\mathcal{T}}(\mu) + \sum_i \tau_i \psi_H^2(h_i) + \mathrm{Tr}(\Theta_2) + 2\sqrt{2}\ln(4m^2 n)\rho. \tag{45b}$$

Note that the constraints on Θ_1 and ρ in this problem say exactly that (Θ_1, ρ) belongs to the cone $\mathbf{K}$ associated, according to Proposition 2, with the ellitope $\overline{\mathcal{M}}$ in the role of $\mathcal{W}$.

Theorem 3. *Given a feasible solution* $(\lambda, \mu, \tau, \Theta_1, \Theta_2, \rho)$ *to* (45)*, let us build* $m \times m$ *contrast matrices* G_1*,* G_2 *as follows.*

- *To build* G_1*, we apply the second part of Proposition 2 to* $(\Theta_1,\, \rho,\, \overline{\mathcal{M}})$ *in the role of* $(\Theta,\, \rho,\, \mathcal{W})$*, to get, in a computationally efficient way, a decomposition* $\Theta_1 = \sum_{i=1}^m \gamma_i g_{1,i} g_{1,i}^T$ *with* $g_{1,i} \in \overline{\mathcal{M}}$ *and* $\gamma_i \geqslant 0$*,* $\sum_i \gamma_i \leqslant 2\sqrt{2}\ln(4m^2n)\rho$*. We set* $G_1 = [g_{1,1}, \ldots, g_{1,m}]$*.*
- *To build* G_2*, we subject* Θ_2 *to the eigenvalue decomposition* $\Theta_2 = \Gamma \mathrm{Diag}\{\chi\}\Gamma^T$ *and set* $G_2 = [g_{2,1}, \ldots, g_{2,m}] = Q\Gamma$*.*

Note that $\Theta_1 + Q\Theta_2 Q = \sum_i \gamma_i g_{1,i} g_{1,i}^T + \sum_i \chi_i g_{2,i} g_{2,i}^T$*.*

For the resulting polyhedral estimate $\widehat{w}_{G,H}$ *and for all* $x_* \in \mathcal{X}$*,* s*-sparse* ν_**, and* $\xi \in \Xi_\epsilon(G, H)$ *if holds*

$$\|\widehat{w}_{G,H}(Ax_* + N\nu_* + \xi) - Bx_*\| \leqslant \overline{f}_H(\lambda, \mu, \tau, \Theta_1, \Theta_2, \rho) \tag{46}$$

implying that the ϵ*-risk of the estimate is upper-bounded by* $f_H(\lambda, \mu, \tau, \Theta_1, \Theta_2, \rho)$ *(as* $\xi \in \Xi_\epsilon(G, H)$ *with probability* $\geqslant 1 - \epsilon$*).*

Proof of the theorem follows that of Theorem 2 and is omitted.

3.4. Putting Things Together

Finally, we can "aggregate" polyhedral estimates from Sections 3.2 and 3.3 in the following construction (cf. [20, Section 5.1.6]):

Let us put

$$\overline{\varkappa}(\epsilon) = \sigma\left(2\ln[(2n + 2M + 8m)/\epsilon]\right)^{1/2},$$

and let $\widetilde{H} = [\widetilde{h}_1, \ldots, \widetilde{h}_n] \in \mathbf{R}^{m\times n}$, $\widetilde{G} = [\widetilde{g}_1, \ldots, \widetilde{g}_{2m}] \in \mathbf{R}^{m\times 2m}$, and $\overline{H} = [\overline{h}_1, \ldots, \overline{h}_M] \in \mathbf{R}^{m\times M}$, $\overline{G} = [\overline{g}_1, \ldots, \overline{g}_{2m}] \in \mathbf{R}^{m\times 2m}$ be the contrast matrices specified according to the synthesis recipes of Sections 3.2 and 3.3, respectively. We define the aggregated estimate $\widehat{w}$ of w_* as $\widehat{w}(\omega) = B\widehat{x}(\omega)$ where $\widehat{x}(\omega)$ is the x-component of

$$(\widehat{x}(\omega), \widehat{\nu}(\omega)) \in \underset{x,\nu}{\mathrm{Argmin}}\left\{\|\nu\|_1 : x \in \mathcal{X},\ \begin{array}{ll} |\widetilde{h}_k^T[N\nu + Ax - \omega]| \leqslant \overline{\varkappa}(\epsilon)\|\widetilde{h}_k\|_2, & k = 1, \ldots, n, \\ |\overline{h}_k^T[N\nu + Ax - \omega]| \leqslant \overline{\varkappa}(\epsilon)\|\overline{h}_k\|_2, & k = 1, \ldots, M, \\ |\widetilde{g}_i^T[N\nu + Ax - \omega]|_\infty \leqslant \overline{\varkappa}(\epsilon)\|\widetilde{g}_i\|_2, & i = 1, \ldots, 2m, \\ |\overline{g}_i^T[N\nu + Ax - \omega]|_\infty \leqslant \overline{\varkappa}(\epsilon)\|\overline{g}_i\|_2, & i = 1, \ldots, 2m, \end{array}\right\}$$

when the problem is feasible, and $\widehat{x}(\omega) = 0$ otherwise.

Let us denote $G = [\widetilde{G}, \overline{G}] \in \mathbf{R}^{m\times 4m}$, let also $(\widetilde{\lambda}, \widetilde{\mu}, \widetilde{\gamma})$ be a feasible solution to the problem (27) with $H = \widetilde{H}$, and let $(\overline{\lambda}, \overline{\mu}, \overline{\gamma})$ be a feasible solution to the problem (43) with $H = \overline{H}$. Let $f_{G,H}$ and $\bar{f}_{G,H}$ be specified in (27) and (43) respectively. From Propositions 3 and 4 it immediately follows that for every s-sparse ν_* and every $x_* \in \mathcal{X}$ the error bound

$$\mathrm{Risk}_\epsilon[\widehat{w}(\cdot)|\mathcal{X}, \mathcal{N}] \leqslant \min\left[f_{G,\widetilde{H}}(\widetilde{\lambda}, \widetilde{\mu}, \widetilde{\gamma}), \bar{f}_{G,\overline{H}}(\overline{\lambda}, \overline{\mu}, \overline{\gamma})\right] \tag{47}$$

holds true.

Note that the resulting estimate can be efficiently optimized w.r.t. all parameters involved, except for $\overline{H}$, by specifying

- $\widetilde{H}$ as (near) minimizer of $\rho[H]$ over $H \in \mathcal{H}$ (23),
- $\widetilde{G}$ as a result of the decomposition of the (Θ_1, Θ_2)-component of a (near-) optimal solution to the problem (33a), (33b) (see Theorem 2) associated with $\widetilde{H}$,
- $\overline{G}$ as a result of the decomposition of the (Θ_1, Θ_2)-component of a (near-) optimal solution to the problem (45) (see Theorem 3) associated with $\overline{H}$.

3.5. Numerical Illustration

In our "proof of concept" experiment we compare three estimates of x_*: 1) estimate $\widehat{x}_{HG}$ with contrast matrix $[H, G]$ computed according to the recipe of Section 3.2, 2) estimate $\widehat{x}_{IG}$ with contrast $[\overline{H}, \overline{G}] = [I_m, \overline{G}]$ with $\overline{G}$ conceived utilizing the synthesis routine of Section 3.3.2, and 3) "aggregated" estimate $\widehat{x}_{HIG}$ with combined contrast $[H, I_m, G, \overline{G}]$. We solve adopted versions of optimization problems in (28) and (33a), (33b) to compute contrasts H and G of the estimate $\widehat{x}_{HG}$, and solve (45), to build the contrast $\overline{G}$ of $\widehat{x}_{IG}$. For instance, when computing the contrast $\overline{G}$, we set $\overline{\varkappa}(\epsilon) = \sqrt{2}\sigma\text{erfcinv}\left(\frac{\epsilon}{2n}\right)$ where erfcinv$(\cdot)$ is the inverse complementary Gaussian error function; when processing problem (45) numerically, Θ_1 was set to 0; the resulting problem can be rewritten as

$$\text{Opt} = \min_{\lambda,\mu,\gamma,\Theta} \left\{ \lambda + 4\sum_{k=1}^{p}\mu_k + \sum_{j=1}^{n}\gamma_j + \text{Tr}(\Theta) : \lambda \geqslant 0,\ \mu \geqslant 0,\ \gamma \geqslant 0,\ \Theta \succeq 0, \right.$$
$$\left. \left[\begin{array}{c|c} \lambda I_p & \frac{1}{2}I_p \\ \hline \frac{1}{2}I_p & \overline{A}^T\Theta\overline{A} + P^T\text{Diag}\{\mu\}P + \overline{A}^T\text{Diag}\{\gamma\}\overline{A} \end{array}\right] \succeq 0 \right\} \tag{48}$$

where $\overline{A} = A/\rho_2$ with $\rho_2 = 2\overline{\varkappa}(\epsilon) + \varrho_{2,\overline{H}}$, the subsequent entries in Pz being z_1, $[z_2 - z_1]/h$, $[z_{i-2} - 2z_{i-1} + z_i]/h^2$, $3 \leqslant i \leqslant p$, and $h = 2\pi/p$. The corresponding risk bounds are evaluated by computing solutions to (43). Optimization problems involved are processed using `Mosek` commercial solver [30] via `CVX` [15].[9]

In our illustration,

- $m = n = 256$, $q = p = 32$, $N = I_n$, $B = I_p$, A is a $n \times p$ random matrix with Gaussian entries such that $A^T A = I_p$;
- $\mathcal{X}$ is the restriction on the p-point equidistant grid on the segment $\Delta = [0, 2\pi]$ of functions f satisfying $|f(0)| \leqslant 4$, $|f'(0)| \leqslant 1$, $|f''(t)| \leqslant 4$, $t \in \Delta$;
- the norm $\|\cdot\|$ quantifying the recovery error is the standard Euclidean norm on $\mathbf{R}^p$;
- $\xi \sim \mathcal{N}(0, \sigma^2 I_m)$ with $\sigma = 0.1$, $\epsilon = 0.05$, and $s = 8$.

Figure 1 illustrates the results of the computation. In each experiment, we compute $n_S = 100$ recoveries $\widehat{x}_{HG}$, $\widehat{x}_{IG}$, and $\widehat{x}_{HIG}$ of randomly selected signals $x_* \in \mathcal{X}$ with generated at random

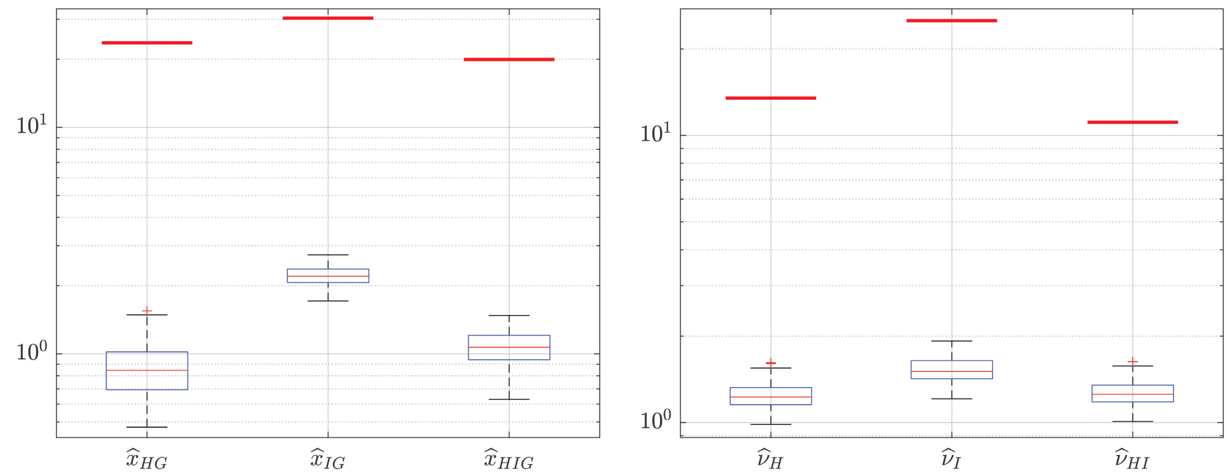

Fig. 1. Left plot: distributions of $\|\cdot\|_2$-errors of recovery of x_* and theoretical upper bounds on $\text{Risk}_{0.05}$ (red horizontal bars); right plot: distributions of $\|\cdot\|_2$-errors and theoretical upper bounds on $\text{Risk}_{0.05}$ of recovery of ν_*.

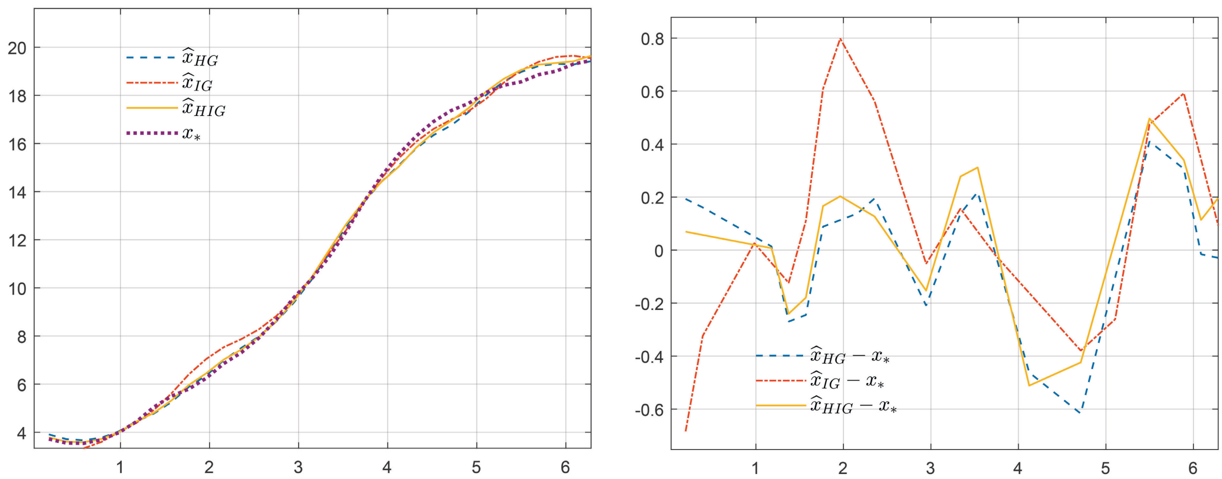

Fig. 2. A typical signal/estimates realization and recovery errors.

sparse nuisances ν_*. The results are presented in the left plot in Fig. 1. The right plot displays the boxplots of errors of recovery of the nuisance ν_* along with the upper risk bound Opt_2 of (40). Figure 2 illustrates a typical realization of the signal and the recovery errors; the values of $\|\cdot\|_2$-recovery errors are $\|\widehat{x}_{HG} - x_*\|_2 = 1.48\ldots$, $\|\widehat{x}_{IG} - x_*\|_2 = 2.02\ldots$, and $\|\widehat{x}_{HIG} - x_*\|_2 = 1.43\ldots$, the corresponding $\|x_*\|_2 = 72.2\ldots$.

APPENDIX. Error bound for ℓ_1 recovery

Condition $Q_\infty(s, \kappa)$

Given an $m \times n$ sensing matrix N, positive integer $s \leqslant n$, and $\kappa \in (0, 1/2)$, we say that $m \times p$ matrix H satisfy condition $\mathbf{Q}_\infty(s, \kappa)$ if

$$\|w\|_\infty \leqslant \|H^T N w\|_\infty + \frac{\kappa}{s}\|w\|_1 \quad \forall w \in \mathbf{R}^n. \tag{A.1}$$

For $y \in \mathbf{R}^n$, let y^s stand for the vector obtained from y by zeroing our all but the s largest in magnitude entries.

Proposition 5. *Given N and integer $s > 0$, assume that H satisfies the condition $\mathbf{Q}_\infty(s, \kappa)$ with $\kappa < \frac{1}{2}$. Then for all $\nu, \widehat{\nu} \in \mathbf{R}^n$ such that $\|\widehat{\nu}\|_1 \leqslant \|\nu\|_1$ it holds:*

$$\|\widehat{\nu} - \nu\|_q \leqslant \frac{(2s)^{\frac{1}{q}}}{1 - 2\kappa}\left[\|H^T N[\widehat{\nu} - \nu]\|_\infty + \frac{\|\nu - \nu^s\|_1}{s}\right], \; 1 \leqslant q \leqslant \infty. \tag{A.2}$$

Proof. Let us denote $\rho = \|H^T N[\widehat{\nu} - \nu]\|_\infty$, and let $z = \widehat{\nu} - \nu$.

1°. Let $I \subset \{1, \ldots, n\}$ of cardinality $\leqslant s$ and let $\overline{I}$ be its complement in $\{1, \ldots, n\}$. When denoting by x_I the vector obtained from a vector x by zeroing out the entries with indexes not belonging to I, we have

$$\|\widehat{\nu}_{\overline{I}}\|_1 = \|\widehat{\nu}\|_1 - \|\widehat{\nu}_I\|_1 \leqslant \|\nu\|_1 - \|\widehat{\nu}_I\|_1 = \|\nu_I\|_1 + \|\nu_{\overline{I}}\|_1 - \|\widehat{\nu}_I\|_1 \leqslant \|z_I\|_1 + \|\nu_{\overline{I}}\|_1,$$

and therefore

$$\|z_{\overline{I}}\|_1 \leqslant \|\widehat{\nu}_{\overline{I}}\|_1 + \|\nu_{\overline{I}}\|_1 \leqslant \|z_I\|_1 + 2\|\nu_{\overline{I}}\|_1.$$

[9] MATLAB code for this experiment is available at GitHub repository `https://github.com/ai1-fr/poly-robust`.

It follows that

$$\|z\|_1 = \|z_I\|_1 + \|z_{\overline{I}}\|_1 \leqslant 2\|z_I\|_1 + 2\|\nu_{\overline{I}}\|_1. \tag{A.3}$$

Besides this, by definition of ρ we have

$$\|H^T N z\|_\infty \leqslant \rho. \tag{A.4}$$

2^o^. Since H satisfies $\mathbf{Q}_\infty(s,\kappa)$, we have

$$\|z\|_{s,1} \leqslant s\|H^T N z\|_\infty + \kappa\|z\|_1$$

where $\|z\|_{s,1}$ is the ℓ_1-norm of the s-dimensional vector composed of the s largest in magnitude entries of z. By (A.4) it follows that $\|z\|_{s,1} \leqslant s\rho + \kappa\|z\|_1$ which combines with the evident inequality $\|z_I\| \leqslant \|z\|_{s,1}$ (recall that $\mathrm{Card}(I) = s$) and with (A.3) to imply that

$$\|z\|_1 \leqslant 2\|z_I\|_1 + 2\|\nu_{\overline{I}}\|_1 \leqslant 2s\rho + 2\kappa\|z\|_1 + 2\|\nu_{\overline{I}}\|_1,$$

hence (recall that $\kappa \leqslant \frac{1}{2}$)

$$\|z\|_1 \leqslant \frac{2s\rho + 2\|\nu_{\overline{I}}\|_1}{1 - 2\kappa}. \tag{A.5}$$

On the other hand, since H satisfies $\mathbf{Q}_\infty(s,\kappa)$, we also have

$$\|z\|_\infty \leqslant \|H^T N z\|_\infty + \frac{\kappa}{s}\|z\|_1,$$

which combines with (A.5) and (A.4) to imply that

$$\|z\|_\infty \leqslant \rho + \frac{\kappa}{s}\frac{2s\rho + 2\|\nu_{\overline{I}}\|_1}{1 - 2\kappa} = (1 - 2\kappa)^{-1}\left[\rho + \frac{\|\nu_{\overline{I}}\|_1}{s}\right]. \tag{A.6}$$

We conclude that for all $1 \leqslant q \leqslant \infty$,

$$\|z\|_p \leqslant \|z\|_\infty^{\frac{q-1}{q}} \|z\|_1^{\frac{1}{q}} \leqslant \frac{(2s)^{\frac{1}{q}}}{1 - 2\kappa}\left[\rho + \frac{\|\nu_{\overline{I}}\|_1}{s}\right]. \qquad \square$$

REFERENCES

1. Balakrishnan, S., Du S.S., Li, J., and Singh, A., Computationally efficient robust sparse estimation in high dimensions, in *Conference on Learning Theory*, pp. 169–212, PMLR, 2017.
2. Bickel, P.J., Ritov, Y., and Tsybakov F.B., Simultaneous analysis of Lasso and Dantzig selector, *The Annals of Statistics*, 2009, vol. 37(4), pp. 1705–1732.
3. Bruce, F.G., Donoho, D.L., Gao, H.-Y., and Martin, R.D., Denoising and robust nonlinear wavelet analysis, in *Wavelet Applications*, vol. 2242, pp. 325–336, SPIE, 1994.
4. Candes, E.J. and Tao, T., Decoding by linear programming, *IEEE transactions on information theory*, 2005, vol. 51(12), pp. 4203–4215.
5. Chen, Y., Caramanis, C., and Mannor, S., Robust sparse regression under adversarial corruption, in *International conference on machine learning*, pp. 774–782, PMLR, 2013.
6. Chernousko, F.L., *State estimation for dynamic systems*, CRC Press, 1993.
7. Dalalyan, A. and Thompson, P., Outlier-robust estimation of a sparse linear model using ℓ_1-penalized Huber's m-estimator, *Advances in neural information processing systems*, 2019, vol. 32.

8. Diakonikolas, I and Kane, D.M., *Algorithmic High-Dimensional Robust Statistics*, Cambridge University Press, 2023.

9. Diakonikolas, I., Kong, W., and Stewart, A., Efficient algorithms and lower bounds for robust linear regression, in *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms*, pp. 2745–2754, SIAM, 2019.

10. Donoho, D.L., Statistical estimation and optimal recovery, *The Annals of Statistics*, 1994, vol. 22(1), pp. 238–270.

11. Donoho, D.L. and Huo, X., Uncertainty principles and ideal atomic decomposition, *Information Theory, IEEE Transactions on*, 2001, vol. 47(7), pp. 2845–2862.

12. Fogel, E. and Huang, Y.-F., On the value of information in system identification-bounded noise case, *Automatica*, 1982, vol. 18(2), pp. 229–238.

13. Foygel, R. and Mackey, L., Corrupted sensing: Novel guarantees for separating structured signals, *IEEE Transactions on Information Theory*, 2014, vol. 60(2), pp. 1223–1247.

14. Granichin, O. and Polyak, P., *Randomized Algorithms of an Estimation and Optimization Under Almost Arbitrary Noises*, Nauka, 2003.

15. Grant, M. and Boyd, S., *The CVX Users' Guide. Release 2.1*, 2014. https://web.cvxr.com/cvx/doc/CVX.pdf

16. Huber, P., *Robust Statistics*, Wiley New York, 1981.

17. Juditsky, A. and Nemirovski, A., Near-optimality of linear recovery from indirect observations, *Mathematical Statistics and Learning*, 2018, vol. 1(2), pp. 171–225.

18. Juditsky, A. and Nemirovski, A., Near-optimality of linear recovery in Gaussian observation scheme under $\|\cdot\|_2^2$-loss, *The Annals of Statistics*, 2018, vol. 46(4), pp. 1603–1629.

19. Juditsky, A. and Nemirovski, A., On polyhedral estimation of signals via indirect observations, *Electronic Journal of Statistics*, 2020, vol. 14(1), pp. 458–502.

20. Juditsky, A. and Nemirovski, A., *Statistical Inference via Convex Optimization*, Princeton University Press, 2020.

21. Juditsky, A. and Nemirovski, A., On design of polyhedral estimates in linear inverse problems, *SIAM Journal on Mathematics of Data Science*, 2024, vol. 6(1), pp. 76–96.

22. Juditsky, A.B. and Nemirovski, A.S., Nonparametric estimation by convex programming, *The Annals of Statistics*, 2009, pp. 2278–2300.

23. Kurzhanski, A., *Identification–a theory of guaranteed estimates*, Springer, 1989.

24. Kurzhanski, A. and Vályi, I., *Ellipsoidal calculus for estimation and control*, Springer, 1997.

25. Liu, L., Shen, Y., Li, T., and Caramanis, C., High dimensional robust sparse regression, in *International Conference on Artificial Intelligence and Statistics*, pp. 411–421, PMLR, 2020.

26. Micchelli, C.A. and Rivlin, Y.J., *A Survey of Optimal Recovery*, pp. 1–54, Springer US, Boston, MA, 1977.

27. Micchelli, C.A. and Rivlin, T.J., Lectures on optimal recovery. In P.R. Turner, editor, *Numerical Analysis Lancaster 1984*, pp. 21–93, Springer Berlin Heidelberg, 1985.

28. Milanese, M. and Vicino, A., Optimal estimation theory for dynamic systems with set membership uncertainty: An overview, *Automatica*, 1991, vol. 27(6), pp. 997–1009.

29. Minsker, S., Ndaoud, M., and Wang, L., Robust and tuning-free sparse linear regression via square-root slope, *SIAM Journal on Mathematics of Data Science*, 2024, vol. 6(2), pp. 428–453.

30. Mosek, A., *The MOSEK optimization toolbox for MATLAB manual. Version 8.0*, 2015. http://docs.mosek.com/8.0/toolbox/

31. Polyak, B.T. and Tsypkin, Y.Z., Adaptive estimation algorithms: convergence, optimality, stability, *Avtomatika i telemekhanika*, 1979, no.3, pp. 71–84.

32. Polyak, B.T. and Tsypkin, Y.Z., Robust identification, *Automatica*, 1980, vol. 16(1), pp. 53–63.

33. Polyak, B.T. and Tsypkin, Y.Z., Robust pseudogradient adaptation algorithms, *Avtomatika i Telemekhanika*, 1980, no. 10, pp. 91–97.

34. Polyak, B.T. and Tsypkin, Y.Z., Optimal and robust methods for unconditional optimization, *IFAC Proceedings Volumes*, 1981, vol. 14(2), pp. 519–523.

35. Polyak, B.T. and Tsypkin, Y.Z., Criterial algorithms of stochastic optimization, *Avtomatika i Telemekhanika*, 1984, no. 6, pp. 95–104.

36. Polyak, B.T. and Tsypkin, Y.Z., Optimal recurrent algorithms for identification of nonstationary plants, *Computers & electrical engineering*, 1992, vol. 18(5), pp. 365–371.

37. Sardy, S., Tseng, P., and Bruce, A., Robust wavelet denoising, *IEEE transactions on signal processing*, 2001, vol. 49(6), pp 1146–1152.

38. Schweppe, F.C., *Uncertain dynamic systems*, Englewood Cliffs, NJ: Prentice-Hall, 1973.

39. Tukey, J.W., A survey of sampling from contaminated distributions, *Contributions to probability and statistics*, 1960, pp. 448–485.

40. Van de Geer, S., *Estimation and Testing under Sparsity*, Springer, 2016.

41. Yu, C. and Yao, W., Robust linear regression: A review and comparison, *Communications in Statistics-Simulation and Computation*, 2017, vol. 46(8), pp. 6261–6282.

This paper was recommended for publication by P.S. Shcherbakov, a member of the Editorial Board