Math. Stat. Learn. 1 (2018), 171–225
DOI 10.4171/MSL/1-2-2

Mathematical Statistics and Learning

Near-optimality of linear recovery from indirect observations

Anatoli Juditsky* and Arkadi Nemirovski**

Abstract. We consider the problem of recovering linear image Bx of a signal x known to belong to a given convex compact set $\mathcal{X}$ from indirect observation $\omega = Ax + \xi$ of x corrupted by random noise ξ with finite covariance matrix. It is shown that under some assumptions on $\mathcal{X}$ (satisfied, e.g. when $\mathcal{X}$ is the intersection of K concentric ellipsoids/elliptic cylinders, or the unit ball of the spectral norm in the space of matrices) and on the norm $\|\cdot\|$ used to measure the recovery error (satisfied, e.g. by $\|\cdot\|_p$-norms, $1 \leq p \leq 2$, on $\mathbf{R}^m$ and by the nuclear norm on the space of matrices), one can build, in a computationally efficient manner, a "seemingly good" *linear in observations estimate*. Further, in the case of zero mean Gaussian observation noise and general mappings A and B, this estimate is near-optimal among all (linear and nonlinear) estimates in terms of the maximal over $x \in \mathcal{X}$ expected $\|\cdot\|$-loss. These results form an essential extension of classical results [7, 24] and of the recent work [13], where the assumptions on $\mathcal{X}$ were more restrictive, and the norm $\|\cdot\|$ was assumed to be the Euclidean one.

Mathematics Subject Classification (2010). 62G08, 62J05; 65K10, 90C22, 90C90.

Keywords. Nonparametric estimation, linear estimation, minimax estimation, conic optimization.

1. Introduction

Broadly speaking, what follows contributes to a long line of research (see, e.g. [4, 7–9, 11, 30, 31] and references therein) started by the pioneering works [15, 16] and [24] and aimed at building efficiently and analysing performance of linear estimates of signals from noisy observations. Specifically, we consider the classical estimation problem as follows: given a "sensing matrix" $A \in \mathbf{R}^{m\times n}$ and an indirect noisy observation

$$\omega = Ax + \xi \tag{1.1}$$

of unknown deterministic "signal" x known to belong to a given "signal set" $\mathcal{X} \subset \mathbf{R}^n$, we are interested to recover the linear image Bx of the signal, where $B \in \mathbf{R}^{\nu\times n}$ is a

*The first author was supported by the LabEx PERSYVAL-Lab (ANR-11-LABX-0025) and PGMO grant 2016-2032H.
**Research of the second author was supported by NSF grantsCCF-1523768 and CMMI-1262063.

given matrix. We assume that the observation noise ξ is random with unknown (and perhaps depending on x) distribution belonging to some family $\mathcal{P}$ of Borel probability distributions on $\mathbf{R}^m$ associated with a given nonempty convex compact subset $\mathcal{Q}$ of the set of positive *definite* $m\times m$ matrices. In this context, "associated" means that the non-centered covariance matrix[1] $\mathrm{Cov}[P]:=\mathbf{E}_{\xi\sim P}\{\xi\xi^T\}$ of a distribution $P\in\mathcal{P}$ is $\succeq$-dominated by some matrix from $\mathcal{Q}$:

$$P\in\mathcal{P}\Rightarrow\exists Q\in\mathcal{Q}:\mathrm{Cov}[P]\preceq Q.^{2} \tag{1.2}$$

We quantify the risk of a candidate estimate — a Borel function $\widehat{x}(\cdot):\mathbf{R}^m\to\mathbf{R}^\nu$ — by its worst-case, under the circumstances, expected $\|\cdot\|$-error defined as

$$\mathrm{Risk}_{\mathcal{Q},\|\cdot\|}[\widehat{x}|\mathcal{X}]=\sup_{x\in\mathcal{X},\,P\in\mathcal{P}}\mathbf{E}_{\xi\sim P}\left\{\|Bx-\widehat{x}(Ax+\xi)\|\right\};$$

here $\|\cdot\|$ is a given norm on $\mathbf{R}^\nu$.

We assume that signal set $\mathcal{X}$ is a special type symmetric w.r.t. the origin convex compact set (a *spectratope* to be defined in Section 2.1), and require from the norm $\|\cdot\|_*$ conjugate to $\|\cdot\|$ to have a spectratope as the unit ball.[3] This allows, e.g. for $\mathcal{X}$ to be the (bounded) intersection of finitely many centered at the origin ellipsoids/elliptic cylinders/$\|\cdot\|_p$-balls ($p\in[2,\infty]$), or the (bounded) solution set of a system of two-sided Linear Matrix Inequalities

$$\left\{x\in\mathbf{R}^n:-L_k\preceq R_k[x]\preceq L_k,k\leq K\right\}\quad\left[R_k[x]\text{: linear in }x\text{ symmetric matrices}\right]$$

As for the norm $\|\cdot\|$, it can be $\|\cdot\|_p$-norm on $\mathbf{R}^\nu$, $1\leq p\leq 2$, or the nuclear norm on the space $\mathbf{R}^\nu=\mathbf{R}^{u\times v}$ of matrices.

An important property of spectratopes is that they allow for precise concentration inequalities for random (Rademacher and Gaussian) vectors, see [3, 17, 18, 25, 29] and references therein. It plays a crucial role in what follows due to several important implications:

– It allows for a tight computationally efficient upper bounding of the maximum of a quadratic form over a spectratope (Proposition 2.2). The latter allow to efficiently upper-bound the maximal over a spectratope risk of linear estimation (i.e. estimate of the form $\widehat{x}_H(\omega)=H^T\omega$), and thus leads to a computationally efficient scheme for building "presumably good" *linear* estimates with guaranteed risk (Proposition 3.5).

[1]For the sake of brevity and with some terminology abuse, in the sequel, we refer to $\mathbf{E}_{\xi\sim P}\{\xi\xi^T\}$ as to covariance matrix of $\xi\sim P$. Note that within the proposed approach we do not need the observation noise to be centered, except for the case of repeated observations, where we explicitly request for the expectation of the noise to vanish (cf. Section 3.4 and Remark 3.6).

[2]Here and below, $U\succeq V$ ($U\succ V$) means that U,V are symmetric matrices of the same size and $U-V$ is positive semidefinite (resp., positive definite); $V\preceq U$ ($V\prec U$) means exactly the same as $U\succeq V$, (resp., $U\succ V$).

[3]Obviously, any result of this type should impose some restrictions on $\mathcal{X}$ — it is well known that linear estimates are "heavily sub-optimal" on some simple signal domains [5,6,23] (e.g. $\|\cdot\|_1$-ball).

– It is also decisive in the implementation of the extended “Pinsker program” [24]: when the family $\mathcal{P}$ of distributions contains all normal distributions $\{\mathcal{N}(0, Q) : Q \in \mathcal{Q}\}$, it allows to tightly lower-bound the minimax risk of estimation over spectratopes via the Bayesian risk of estimating a random Gaussian signal, and to show that the presumably good linear estimates are “near-optimal” (optimal up to logarithmic factors) among *all* estimates, linear and nonlinear alike.

An “executive summary” of our main result, Proposition 3.8, is as follows:

Given a spectratope $\mathcal{X}$ and assuming that the unit ball $\mathcal{B}_*$ of the norm conjugate to $\|\cdot\|$ is a spectratope as well, the efficiently computable optimal solution H_* to an explicitly posed convex optimization problem yields a near-optimal linear estimate $\widehat{x}_{H_*}(\omega) = H_*^T\omega$, specifically,

$$\mathrm{Risk}_{\mathcal{Q},\|\cdot\|}[\widehat{x}_{H_*}|\mathcal{X}] \leq C\sqrt{\ln\left(\mathcal{L}\,\mathrm{RiskOpt}^{-1}_{\mathcal{Q},\|\cdot\|}[\mathcal{X}]\right)}\,\mathrm{RiskOpt}_{\mathcal{Q},\|\cdot\|}[\mathcal{X}], \qquad (*)$$

where C is an absolute constant, and $\mathcal{L}$ is polynomial in the (naturally defined) sizes and magnitudes of the data specifying B and the spectratopes $\mathcal{X}$, $\mathcal{B}_*$, and $\mathrm{RiskOpt}_{\mathcal{Q},\|\cdot\|}$ is the “true” minimax optimal risk associated with zero mean Gaussian observation noises with covariance matrices from $\mathcal{Q}$:

$$\mathrm{RiskOpt}_{\mathcal{Q},\|\cdot\|}[\mathcal{X}] = \sup_{Q\in\mathcal{Q}}\inf_{\widehat{x}}\max_{x\in\mathcal{X}}\mathbf{E}_{\xi\sim\mathcal{N}(0,Q)}\big\{\|Bx-\widehat{x}(Ax+\xi)\|\big\},$$

the infimum being taken over all estimates, linear and nonlinear alike.

It should be stressed that the “nonoptimality factor” in $(*)$ is logarithmic and is completely independent of the sensing matrix A — the entity “primarily responsible” for the minimax optimal risk.

The above result constitutes an important extension to the approach developed in [13], the progress as compared to [13] being as follows:

– [13] dealt with the case of $\mathcal{P} = \{\mathcal{N}(0, Q)\}$, i.e. the observation noise was assumed to be zero mean Gaussian with known covariance matrix, while now we allow for $\mathcal{P}$ to be a general family of probability distributions with covariance matrices $\succeq$-dominated by matrices from a given convex compact set $\mathcal{Q} \subset \mathrm{int}\,\mathbf{S}^m_+$; [4]

– Present results apply to an essentially wider family of signal sets: spectratopes as compared to *ellitopes* considered in [13]; ellitopes are also spectratopes, see Section 2.1, but not vice versa. For instance, the intersection of centered at the origin ellipsoids/elliptic cylinders/$\|\cdot\|_p$-balls, $p \in [2,\infty]$, is an ellitope (and thus a spectratope), whereas the (bounded) solution set of a finite system of two-sided LMI’s is a spectratope, but not an ellitope;

[4]From now on, $\mathbf{S}^k$ stands for the space of symmetric $k\times k$ matrices, and $\mathbf{S}^k_+$ is the cone of positive semidefinite matrices from $\mathbf{S}^k$.

– The analysis in [13] was limited to the case of $\|\cdot\|_2$-losses, while now we allow for a much wider family of norms quantifying the recovery error.

Note that, in addition to observations with random noise, in what follows we also address observations with "uncertain-but-bounded" and "mixed" (combined) noise. In the latter case ξ, instead of being random, is selected, perhaps in adversarial manner, from a given spectratope — the situation which was not considered in [13] at all.

These results can also be considered as a new contribution to the line of research initiated by [7], where it is proved (Proposition 4–Theorem 7) that *if* $\mathcal{X}$ *is convex, orthosymmetric and quadratically convex* (that is, $\mathcal{X}=\{x\in\mathbf{R}^n: \exists t\in\mathcal{T}: x_i^2\leq t_i, i\leq n\}$ with convex compact $\mathcal{T}\subset\mathbf{R}^n_+$), *observations are direct:* $\omega=x+\xi$, $\xi\sim\mathcal{N}(0,I_n)$, $Bx=x$, *and* $\|\cdot\|=\|\cdot\|_2$, *the risk of an efficiently computable linear estimate is within factor 1.25 of the minimax optimal risk*. The suboptimality guarantees provided by the latter result are *essentially better* than those of Proposition 3.8 in the current paper. However, it is also *essentially more restrictive* in its scope — an orthosymmetric convex and quadratically convex set is a very special case of an ellitope, the observations should be direct, and $\|\cdot\|$ should be $\|\cdot\|_2$.

Note that linear estimators can be efficiently built and optimized for some signal domains which are not spectratopes, e.g. when $\mathcal{X}$ is given as a convex hull of a finite set, e.g. $\mathcal{X}$ is $\|\cdot\|_1$-ball (in this case, the smallest risk linear estimate can be "heavily nonoptimal" among all estimates). In general, however, optimizing risk over just linear estimates in a computationally efficient fashion can be problematic. Beyond the scope of spectratopes in the role of signal sets and unit balls of the norms conjugate to those in which the recovery error is measured, the only known to us general situation where "presumably good" linear estimation is computationally tractable and results in (nearly) minimax optimal estimates is that where the recovery error is measured in $\|\cdot\|_\infty$. In the latter case, the breakthrough papers [4, 12] (see also [14]) imply that whenever $\mathcal{X}$ is a computationally tractable convex compact set and the observation noise is Gaussian, an efficiently computable linear estimate is $\|\cdot\|_\infty$-minimax optimal within the factor $O(1)\sqrt{\ln(\nu)}$.

The main body of the paper is organized as follows. We start with describing the family of sets we work with – the spectratopes (Section 2.1), and derive the crucial for the rest of the paper result on tight upper-bounding the maximum of a quadratic form over a spectratope (Section 2.2). Next we explain how to build in a computationally efficient fashion "presumably good" linear estimates in the case of stochastic (Section 3) and uncertain-but-bounded (Section 4) observation noise and establish near-optimality of these estimates. All technical proofs are relegated to Section 5. Appendix A lists principal rules of calculus of spectratopes. Appendix B contains implementation details for the illustrative example (covariance matrix estimation) presented in Section 3.4. Finally, Appendix C contains an "executive summary" of conic duality, our principal working horse.

2. Preliminaries

We start with describing the main geometric object we intend to work with — a *spectratope*.

2.1. Spectratopes.

A basic spectratope is a set $\mathcal{X} \subset \mathbf{R}^n$ given by *basic spectratopic representation* — a representation of the form

$$\mathcal{X} = \left\{x \in \mathbf{R}^n : \exists t \in \mathcal{T} : R_k^2[x] \preceq t_k I_{d_k},\ 1 \le k \le K\right\}, \tag{2.1}$$

where

(S$_1$) $R_k[x] = \sum_{i=1}^n x_i R^{ki}$ are symmetric $d_k \times d_k$ matrices linearly depending on $x \in \mathbf{R}^n$ (i.e. "matrix coefficients" R^{ki} belong to $\mathbf{S}^{d_k}$);

(S$_2$) $\mathcal{T} \in \mathbf{R}^K_+$ is a *monotonic* set, meaning that $\mathcal{T}$ is a convex compact subset of $\mathbf{R}^K_+$ which contains a positive vector and is monotone:

$$0 \le t' \le t \in \mathcal{T} \Rightarrow t' \in \mathcal{T}.$$ [5]

(S$_3$) Whenever $x \neq 0$, it holds $R_k[x] \neq 0$ for at least one $k \le K$.

Remark 2.1. By the Schur Complement lemma, the set (2.1) given by the data satisfying (S$_1$), (S$_2$) can be represented as

$$\mathcal{X} = \left\{x \in \mathbf{R}^n : \exists t \in \mathcal{T} : \left[\begin{array}{c|c} t_k I_{d_k} & R_k[x] \\ \hline R_k[x] & I_{d_k} \end{array}\right] \succeq 0,\ k \le K\right\}.$$

By the latter representation, $\mathcal{X}$ is nonempty, closed, convex, symmetric w.r.t. the origin, and contains a neighbourhood of the origin (the latter is due to the fact that $\mathcal{T}$ contains a strictly positive vector). This set is bounded if and only if the data, in addition to (S$_1$), (S$_2$), satisfies (S$_3$).

A spectratope. $\mathcal{X} \subset \mathbf{R}^p$ is a set represented as a linear image of a basic spectratope:

$$\mathcal{X} = \left\{x \in \mathbf{R}^p : \exists (y \in \mathbf{R}^n,\ t \in \mathcal{T}) : x = Py,\ R_k^2[y] \preceq t_k I_{d_k},\ 1 \le k \le K\right\}, \tag{2.2}$$

where P is a $p \times n$ matrix, and $R_k[\cdot]$, $\mathcal{T}$ are as in (S$_1$)–(S$_3$). We call the quantity

$$D = \sum_{k=1}^K d_k$$

the *size* of spectratope $\mathcal{X}$.

[5] The inequalities between vectors are understood componentwise.

Example 1 (Ellitopes). An *ellitope* was defined in [13] as a set $\mathcal{X} \subset \mathbf{R}^n$ representable as

$$\mathcal{X} = \{x \in \mathbf{R}^n : \exists (y \in \mathbf{R}^N, t \in \mathcal{T}) : x = Py, \; y^T S_k y \le t_k, \; k \le K\}, \tag{2.3}$$

where $S_k \succeq 0$, $\sum_K S_k \succ 0$, and $\mathcal{T}$ satisfies (S_2). Basic examples of ellitopes are:

– bounded intersections of centered at the origin ellipsoids/elliptic cylinders: whenever $S_k \succeq 0$ and $\sum_k S_k \succ 0$,

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

– $\|\cdot\|_p$-balls, $2 \le p \le \infty$:

$$\{x \in \mathbf{R}^n : \|x\|_p \le 1\} \\ = \{x \in \mathbf{R}^n : \exists t \in \mathcal{T} := \{t \ge 0, \; \|t\|_{p/2} \le 1\} : x^T S_k x := x_k^2 \le t_k, \; k \le n\}.$$

It is immediately seen that an ellitope (2.3) is a spectratope as well. Indeed, let $S_k = \sum_{j=1}^{r_k} s_{kj} s_{kj}^T$, $r_k = \mathrm{Rank}(S_k)$, be a dyadic representation of the positive semidefinite matrix S_k, so that

$$y^T S_k y = \sum_j (s_{kj}^T y)^2, \quad \forall y,$$

and let

$$\widehat{\mathcal{T}} = \Big\{\{t_{kj} \ge 0, \; 1 \le j \le r_k, \; 1 \le k \le K\} : \exists t \in \mathcal{T} : \sum_j t_{kj} \le t_k, \; k \le K\Big\}, \\ R_{kj}[y] = s_{kj}^T y \in \mathbf{S}^1 = \mathbf{R}.$$

We clearly have

$$\mathcal{X} = \{x \in \mathbf{R}^n : \exists (\{t_{kj}\} \in \widehat{\mathcal{T}}, y) : x = Py, \; R_{kj}^2[y] \preceq t_{kj} I_1, \; \forall k, j\}$$

and the right hand side is a valid spectratopic representation of $\mathcal{X}$. Note that the spectratopic size of $\mathcal{X}$ is $D = \sum_{k=1}^K r_k$.

Example 2 ("Matrix box"). Let L be a positive definite $d \times d$ matrix. Then the "matrix box"

$$\mathcal{X} = \{X \in \mathbf{S}^d : -L \preceq X \preceq L\} = \{X \in \mathbf{S}^d : -I_d \preceq L^{-1/2} X L^{-1/2} \preceq I_d\} \\ = \{X \in \mathbf{S}^d : R^2[X] := [L^{-1/2} X L^{-1/2}]^2 \preceq I_d\}$$

is a basic spectratope (augment $R_1[\cdot] := R[\cdot]$ with $K = 1$, $\mathcal{T} = [0, 1]$). As a result, a *bounded* set $\mathcal{X} \subset \mathbf{R}^n$ given by a system of "two-sided" Linear Matrix Inequalities, specifically,

$$\mathcal{X} = \left\{x \in \mathbf{R}^n : \exists t \in \mathcal{T} : -\sqrt{t_k} L_k \preceq S_k[x] \preceq \sqrt{t_k} L_k,\ 1 \le k \le K\right\}$$

where $S_k[x]$ are symmetric $d_k \times d_k$ matrices linearly depending on x, $L_k \succ 0$ and $\mathcal{T}$ satisfies (S_2), is a basic spectratope:

$$\mathcal{X} = \left\{x \in \mathbf{R}^n : \exists t \in \mathcal{T} : R_k^2[x] \preceq t_k I_{d_k},\ 1 \le k \le K\right\}$$
$$[R_k[x] = L_k^{-1/2} S_k[x] L_k^{-1/2}].$$

More examples of spectratopes can be built using their "calculus." It turns out that nearly all basic operations with sets preserving convexity, symmetry w.r.t. the origin, and boundedness (these are "built-in" properties of spectratopes), such as taking finite intersections, direct products, arithmetic sums, linear images, and inverse linear images under linear embeddings, as applied to spectratopes, yield spectratopes as well. Furthermore, a spectratopic representation of the result of such an operation is readily given by spectratopic representations of the operands; see Appendix A for principal calculus rules.

2.2. Upper-bounding quadratic form on a spectratope. We are about to establish the first crucial in our context fact about spectratopes — the possibility to tightly upper-bound an (indefinite) quadratic form over a spectratope. To proceed, we need some definitions.

Linear maps associated with a spectratope. We associate with a basic spectratope (2.1), (S_1)–(S_3) the following entities:

1. Linear mappings

$$Q \mapsto \mathcal{R}_k[Q] = \sum_{i,j} Q_{ij} R^{ki} R^{kj} : \mathbf{S}^n \to \mathbf{S}^{d_k}.$$

As is immediately seen, we have

$$\mathcal{R}_k[yy^T] \equiv R_k^2[y], \tag{2.4}$$

implying that $\mathcal{R}_k[Q] \succeq 0$ whenever $Q \succeq 0$, whence $\mathcal{R}_k[\cdot]$ is $\succeq$-monotone:

$$Q' \succeq Q \Rightarrow \mathcal{R}_k[Q'] \succeq \mathcal{R}_k[Q]. \tag{2.5}$$

Besides this, if ξ is a random vector taking values in $\mathbf{R}^n$ with covariance matrix Q, we have

$$\mathbf{E}_\xi\left\{R_k^2[\xi]\right\} = \mathbf{E}_\xi\left\{\mathcal{R}_k[\xi\xi^T]\right\} = \mathcal{R}_k\left[\mathbf{E}_\xi\{\xi\xi^T\}\right] = \mathcal{R}_k[Q], \tag{2.6}$$

where the first equality is given by (2.4).

2. Linear mappings $\Lambda_k \mapsto \mathcal{R}_k^*[\Lambda_k]: \mathbf{S}^{d_k} \to \mathbf{S}^n$ given by

$$\left[\mathcal{R}_k^*[\Lambda_k]\right]_{ij} = \tfrac{1}{2}\operatorname{Tr}\left(\Lambda_k[R^{ki}R^{kj} + R^{kj}R^{ki}]\right), \quad 1 \le i, j \le n. \tag{2.7}$$

It is immediately seen that $\mathcal{R}_k^*[\cdot]$ is the adjoint of $\mathcal{R}_k[\cdot]$:

$$\begin{aligned}\forall (\Lambda_k \in \mathbf{S}^{d_k},\ Q \in \mathbf{S}^n) : \langle \Lambda_k, \mathcal{R}_k[Q]\rangle &= \operatorname{Tr}\left(\Lambda_k \mathcal{R}_k[Q]\right)\\ &= \operatorname{Tr}\left(\mathcal{R}_k^*[\Lambda_k]Q\right) = \langle \mathcal{R}_k^*[\Lambda_k], Q\rangle,\end{aligned} \tag{2.8}$$

where $\langle A, B\rangle = \operatorname{Tr}(AB)$ is the Frobenius inner product of symmetric matrices. Besides this, we have[6]

$$\Lambda_k \succeq 0 \Rightarrow \mathcal{R}_k^*[\Lambda_k] \succeq 0. \tag{2.9}$$

3. The linear space $\Lambda^K = \mathbf{S}^{d_1} \times \cdots \times \mathbf{S}^{d_K}$ of all ordered collections $\Lambda = \{\Lambda_k \in \mathbf{S}^{d_k}\}_{k \le K}$ along with the linear mapping

$$\Lambda \mapsto \lambda[\Lambda] := \left[\operatorname{Tr}(\Lambda_1); \ldots; \operatorname{Tr}(\Lambda_K)\right] : \Lambda^K \to \mathbf{R}^K.$$

Besides this, for a monotonic set $\mathcal{T} \subset \mathbf{R}^K$ we define:

– the *support function* of $\mathcal{T}$

$$\phi_{\mathcal{T}}(g) = \max_{t \in \mathcal{T}} g^T t,$$

which clearly is a convex positively homogeneous, of degree 1, nonnegative real-valued function on $\mathbf{R}^K$. Since $\mathcal{T}$ contains positive vectors, $\phi_{\mathcal{T}}$ is coercive on $\mathbf{R}^K_+$, meaning that $\phi_{\mathcal{T}}(\lambda^s) \to +\infty$ along every sequence $\{\lambda^s \ge 0\}$ such that $\|\lambda^s\| \to \infty$;

– the conic hull of $\mathcal{T}$

$$\mathbf{K}[\mathcal{T}] = \operatorname{cl}\left\{[t; s] \in \mathbf{R}^{K+1} : s > 0,\ s^{-1}t \in \mathcal{T}\right\}$$

which clearly is a regular cone in $\mathbf{R}^{K+1}$ (i.e. it is closed, convex, and pointed with a nonempty interior) such that

$$\mathcal{T} = \left\{t : [t; 1] \in \mathbf{K}[\mathcal{T}]\right\}.$$

It is immediately seen that the cone $(\mathbf{K}[\mathcal{T}])_*$ dual to $\mathbf{K}[\mathcal{T}]$ can be described as follows:

$$\begin{aligned}\left(\mathbf{K}[\mathcal{T}]\right)_* := \left\{[g; r] \in \mathbf{R}^{K+1} : [g; r]^T[t; s] \ge 0\ \forall [t; s] \in \mathbf{K}[\mathcal{T}]\right\}\\ = \left\{[g; r] \in \mathbf{R}^{K+1} : r \ge \phi_{\mathcal{T}}(-g)\right\}.\end{aligned}$$

[6]Note that when $\Lambda_k \succeq \mathbf{0}$ and $Q = yy^T$, the first quantity in (2.8) is nonnegative by (2.4), and therefore (2.8) states that $y^T \mathcal{R}_k^*[\Lambda_k] y \ge \mathbf{0}$ for every y, implying that $\mathcal{R}_k^*[\Lambda_k] \succeq \mathbf{0}$.

Proposition 2.2. *Let C be a symmetric $p \times p$ matrix, let $\mathcal{X} \subset \mathbf{R}^p$ be given by spectratopic representation (2.2),*

$$\mathrm{Opt} = \max_{x\in\mathcal{X}} x^T C x$$

and let

$$\mathrm{Opt}_* = \min_{\Lambda=\{\Lambda_k\}_{k\le K}} \left\{\phi_{\mathcal{T}}(\lambda[\Lambda]) : \Lambda_k \succeq 0,\ k \le K,\ P^T C P \preceq \sum_k \mathcal{R}_k^*[\Lambda_k]\right\}$$
$$\big[\lambda[\Lambda] = [\mathrm{Tr}(\Lambda_1); \dots; \mathrm{Tr}(\Lambda_K)]\big]. \tag{2.10}$$

Then (2.10) is solvable, and

$$\mathrm{Opt} \le \mathrm{Opt}_* \le 2\max\left[\ln(2D), 1\right] \mathrm{Opt}, \tag{2.11}$$

where $D = \sum_k d_k$ is the size of the spectratope $\mathcal{X}$.

To explain where the result of the proposition comes from, let us prove right now its easy part — the first inequality in (2.11); the remaining, essentially less trivial, part of the proof is provided in Section 5.2. Let Λ be a feasible solution to the optimization problem in (2.10), and let $x \in \mathcal{X}$, so that $x = Py$ for some y such that $R_k^2[y] \preceq t_k I_{d_k}$, $k \le K$, for properly selected $t \in \mathcal{T}$. We have

$$\begin{aligned}
x^T C x &= y^T[P^T C P]y \\
&\underbrace{\le}_{(a)} \sum_k y^T \mathcal{R}_k^*[\Lambda_k]\, y = \sum_k \mathrm{Tr}\left(\mathcal{R}_k^*[\Lambda_k] y y^T\right) \underbrace{=}_{(b)} \sum_k \mathrm{Tr}\left(\Lambda_k \mathcal{R}_k[y y^T]\right) \\
&\underbrace{=}_{(c)} \sum_k \mathrm{Tr}\left(\Lambda_k R_k^2[y]\right) \underbrace{\le}_{(d)} \sum_k \mathrm{Tr}\left(\Lambda_k t_k I_{d_k}\right) = \sum_k t_k \,\mathrm{Tr}(\Lambda_k) = \lambda^T[\Lambda]\, t \\
&\underbrace{\le}_{(e)} \phi_{\mathcal{T}}(\lambda[\Lambda]),
\end{aligned}$$

where (a) is due to the fact that Λ is feasible for the optimization problem in (2.10), (b) is by (2.8), (c) is by (2.4), (d) is due to $\Lambda_K \succeq 0$ and $R_k^2[y] \preceq t_k I_{d_k}$, and (e) is by the definition of $\phi_{\mathcal{T}}$. The bottom line is that the value of the objective of the optimization problem in (2.10) at every feasible solution to this problem upper-bounds Opt, implying the first inequality in (2.11). Note that the derivation we have carried out is nothing but a minor modification of the standard semidefinite relaxation scheme.

Remark 2.3. Proposition 2.2 has some history. When $\mathcal{X}$ is an intersection of centered at the origin ellipsoids/elliptic cylinders, it was established in [22]; matrix analogy of the latter result can be traced back to [21], see also [26]. The case when $\mathcal{X}$ is a general-type ellitope (2.3) was considered in [13], with tightness guarantee slightly better than in (2.11), namely,

$$\mathrm{Opt} \le \mathrm{Opt}_* \le 4\ln(5K)\,\mathrm{Opt}\,.$$

Note that in the case where $\mathcal{X}$ is an ellitope (2.3), Proposition 2.2 results in a worse than $O(1)\ln(K)$ "nonoptimality factor" $O(1)\ln(\sum_{k=1}^K \mathrm{Rank}(S_k))$. We remark that passing from ellitopes to spectratopes requires replacing elementary bounds on deviation probabilities used in [13,22] by a more powerful tool — matrix concentration inequalities, see [18,29] and references therein.

3. Near-optimal linear estimation under random noise

3.1. Situation and goal. Given $\nu \times n$ matrix B, consider the problem of estimating linear image Bx of unknown deterministic signal x known to belong to a given set $\mathcal{X} \subset \mathbf{R}^n$ via noisy observation

$$\omega = Ax + \xi, \tag{3.1}$$

where A is a given $m \times n$ matrix A and ξ is random observation noise. In some signal processing applications, the distribution of noise is fixed and is part of the data of the estimation problem. In order to cover some interesting applications (cf. Section 3.4), we allow for "ambiguous" noise distributions; all we know in advance is that this distribution belongs to a family $\mathcal{P}$ of Borel probability distributions on $\mathbf{R}^m$ associated, in the sense of (1.2), with a given convex compact subset $\mathcal{Q}$ of the interior of the cone $\mathbf{S}^m_+$ of positive semidefinite $m \times m$ matrices. Actual distribution of noise in (3.1) is somehow selected from $\mathcal{P}$ by nature (and may, e.g. depend on x).

In the sequel, for a Borel probability distribution P on $\mathbf{R}^m$ we write $P \lessdot \mathcal{Q}$ to express the fact that $\mathrm{Cov}[P]$ is $\succeq$-dominated by a matrix from $\mathcal{Q}$:

$$\{P \lessdot \mathcal{Q}\} \Leftrightarrow \big\{\exists Q \in \mathcal{Q} : \mathrm{Cov}[P] \preceq Q\big\}.$$

From now on we assume that *all matrices from* $\mathcal{Q}$ *are positive definite.*

Given $\mathcal{Q}$ and a norm $\|\cdot\|$ on $\mathbf{R}^\nu$, we quantify the *risk* of a candidate estimate — a Borel function $\widehat{x}(\cdot)\colon \mathbf{R}^m \to \mathbf{R}^\nu$ — by its $(\mathcal{Q}, \|\cdot\|)$-risk on $\mathcal{X}$ defined as

$$\mathrm{Risk}_{\mathcal{Q},\|\cdot\|}[\widehat{x}|\mathcal{X}] = \sup_{x\in\mathcal{X}, P \lessdot \mathcal{Q}} \mathbf{E}_{\xi\sim P}\big\{\|\widehat{x}(Ax+\xi) - Bx\|\big\}, \tag{3.2}$$

where $\|\cdot\|$ is some norm on $\mathbf{R}^\nu$. Our focus is on *linear estimates* — estimates of the form

$$\widehat{x}_H(\omega) = H^T\omega$$

given by $m \times \nu$ matrices H, and our current goal is to demonstrate that under some restrictions on the signal domain $\mathcal{X}$, a "good" linear estimate yielded by an optimal solution to an efficiently solvable convex optimization problem is near-optimal in terms of its risk among *all* estimates, linear and nonlinear alike.

We assume here that set $\mathcal{X}$ is a spectratope (cf. (2.2)). Ideally, to compute a "good" linear estimate one would look for H which minimizes the risk $\mathrm{Risk}_{\mathcal{Q},\|\cdot\|}[\widehat{x}_H|\mathcal{X}]$ in H. This risk is generally difficult to compute even when $\mathcal{X}$ is a spectratope, and with our approach, we minimize in H an efficiently computable upper bound on the risk rather than the risk itself. In order this bound to be tight — good enough to allow to build a near-optimal linear estimate, we have to impose further restrictions, specifically, we make from now on the following

Assumption A. *The unit ball $\mathcal{B}_*$ of the norm $\|\cdot\|_*$* conjugate *to the norm $\|\cdot\|$ in the definition* (3.2) *of the estimation risk* is a spectratope:

$$\begin{aligned}\mathcal{B}_* &= \{z \in \mathbf{R}^\nu : \exists y \in \mathcal{Y} : z = My\},\\ \mathcal{Y} &:= \{y \in \mathbf{R}^q : \exists r \in \mathcal{R} : S_\ell^2[y] \preceq r_\ell I_{f_\ell},\ 1 \le \ell \le L\},\end{aligned} \tag{3.3}$$

where the right hand side data are as required in a spectratopic representation.

Examples of norms $\|\cdot\|$ satisfying Assumption A include $\|\cdot\|_q$-norms on $\mathbf{R}^\nu$, $1 \le q \le 2$ (conjugates of the norms $\|\cdot\|_p$ with $1/p + 1/q = 1$, see Example 1 in Section 2.1). Another example is *nuclear norm* $\|\cdot\|_{\mathrm{Sh},1}$ on the space $\mathbf{R}^\nu = \mathbf{R}^{p\times q}$ of $p \times q$ matrices, $\|V\|_{\mathrm{Sh},1} = \sum \sigma_i(V)$ — the sum of singular values of a matrix V. The conjugate of the nuclear norm is the spectral norm $\|\cdot\|_{\mathrm{Sh},\infty}$ on $\mathbf{R}^\nu = \mathbf{R}^{p\times q}$, and the unit ball of the latter norm is a basic spectratope (cf. Example 2 in Section 2.1):

$$\{Z \in \mathbf{R}^{p\times q} : \|Z\|_{\mathrm{Sh},\infty} \le 1\} = \{Z : \exists r \in \mathcal{R} = [0,1] : S^2[Z] \preceq t I_{p+q}\},\quad S[Z] = \left[\begin{array}{c|c} & Z^T \\ \hline Z & \end{array}\right].$$

It is immediately seen that the case when $\mathcal{X}$ is a spectratope (2.2) can be reduced to the one where $\mathcal{X}$ is a *basic* spectratope — to this end it suffices to replace matrices A and B with AP and BP, respectively, and to treat y rather than $x = Py$ as the signal underlying observation (3.1), see (2.2). We assume that this reduction has been carried out in advance, so that from now on our signal set will be

$$\mathcal{X} = \{x \in \mathbf{R}^n : \exists t \in \mathcal{T} : R_k^2[x] \preceq t_k I_{d_k},\ 1 \le k \le K\}.$$

3.2. Building linear estimate. Observe that the $(\mathcal{Q}, \|\cdot\|)$-risk of the linear estimate $\widehat{x}_H(\omega) = H^T\omega$, $H \in \mathbf{R}^{m\times\nu}$, can be upper-bounded as follows:

$$\begin{aligned}\mathrm{Risk}_{\mathcal{Q},\|\cdot\|}[\widehat{x}_H(\cdot)|\mathcal{X}] &= \sup_{x\in\mathcal{X}, P\preccurlyeq\mathcal{Q}} \mathbf{E}_{\xi\sim P}\{\|H^T(Ax+\xi) - Bx\|\}\\ &\le \sup_{x\in\mathcal{X}} \|H^TAx - Bx\| + \sup_{P\preccurlyeq\mathcal{Q}} \mathbf{E}_{\xi\sim P}\{\|H^T\xi\|\}\\ &\le \Phi_{\mathcal{X}}(H) + \Psi_{\mathcal{Q}}(H),\end{aligned} \tag{3.4}$$

where

$$\Phi_{\mathcal{X}}(H) = \max_{x} \big\{\|(B - H^T A)x\| : x \in \mathcal{X}\big\}, \quad \Psi_{\mathcal{Q}}(H) = \sup_{P \preccurlyeq \mathcal{Q}} \mathbf{E}_{\xi\sim P}\big\{\|H^T \xi\|\big\}.$$

While $\Phi_{\mathcal{X}}(H)$ and $\Psi_{\mathcal{Q}}(H)$ are convex functions of H, these functions can be difficult to compute.[7] In such a case, matrix H of a "good" linear estimate $\widehat{x}_H$ which is also efficiently computable can be chosen as a minimizer of the sum of efficiently computable convex upper bounds on $\Phi_{\mathcal{X}}$ and $\Psi_{\mathcal{Q}}$.

3.2.1. Upper-bounding $\Phi_{\mathcal{X}}(\cdot)$. With Assumption A in force, let us consider the direct product spectratope

$$\begin{aligned} \mathcal{Z} := \mathcal{X} \times \mathcal{Y} = \big\{[x;y] \in \mathbf{R}^n \times \mathbf{R}^q : \exists s = [t;r] \in \mathcal{T} \times \mathcal{R} :\\ R_k^2[x] \preceq t_k I_{d_k},\ 1 \le k \le K,\ S_\ell^2[y] \preceq r_\ell I_{f_\ell},\ 1 \le \ell \le L\big\}\\ = \big\{w = [x;y] \in \mathbf{R}^n \times \mathbf{R}^q : \exists s = [t;r] \in \mathcal{S} = \mathcal{T} \times \mathcal{R} :\\ U_i^2[w] \preceq s_i I_{g_i},\ 1 \le i \le I = K + L\big\} \end{aligned}$$

with $U_i[\cdot]$ readily given by $R_k[\cdot]$ and $S_\ell[\cdot]$. Given a $\nu \times n$ matrix V and setting

$$W[V] = \frac{1}{2}\left[\begin{array}{c|c} & V^T M \\ \hline M^T V & \end{array}\right]$$

it clearly holds

$$\max_{x\in\mathcal{X}} \|Vx\| = \max_{x\in\mathcal{X},\, z\in\mathcal{B}_*} z^T V x = \max_{x\in\mathcal{X},\, y\in\mathcal{Y}} y^T M^T V x = \max_{w\in\mathcal{Z}} w^T W[V] w.$$

Applying Proposition 2.2, we arrive at the following result (cf. Proposition 4.1):

Corollary 3.1. *In the just defined situation, the efficiently computable convex function*

$$\begin{aligned} \overline{\Phi}_{\mathcal{X}}(H) = \min_{\Lambda,\Upsilon} \Big\{ \phi_{\mathcal{T}}\big(\lambda[\Lambda]\big) + \phi_{\mathcal{R}}\big(\lambda[\Upsilon]\big) : \Lambda = \big\{\Lambda_k \in \mathbf{S}_+^{d_k}\big\}_{k\le K},\\ \Upsilon = \big\{\Upsilon_\ell \in \mathbf{S}_+^{f_\ell}\big\}_{\ell\le L},\ \left[\begin{array}{c|c} \sum_k \mathcal{R}_k^*[\Lambda_k] & \frac{1}{2}\big(B - H^T A\big)^T M \\ \hline \frac{1}{2} M^T\big(B - H^T A\big) & \sum_\ell \mathcal{S}_\ell^*[\Upsilon_\ell] \end{array}\right] \succeq 0 \Big\} \end{aligned} \quad (3.5)$$

is a tight upper bound on $\Phi_{\mathcal{X}}(\cdot)$, namely,

$$\forall H \in \mathbf{R}^{m\times\nu} : \Phi_{\mathcal{X}}(H) \le \overline{\Phi}_{\mathcal{X}}(H) \le 2\max\big[\ln(2D), 1\big]\,\Phi_{\mathcal{X}}(H),\qquad D = \sum_k d_k + \sum_\ell f_\ell.$$

[7] For instance, computing $\Psi_{\mathcal{X}}(H)$ reduces to maximizing the convex function $\|(B - H^T A)x\|$ over $x \in \mathcal{X}$, which is computationally intractable even when $\mathcal{X}$ is as simple as the unit box, and $\|\cdot\|$ is the Euclidean norm.

Recall, that here

$$\begin{aligned}\big[\mathcal{R}_k^*[\Lambda_k]\big]_{ij} &= \frac{1}{2}\operatorname{Tr}\big(\Lambda_k\big[R_k^{ki}R_k^{kj}+R_k^{kj}R_k^{ki}\big]\big), \quad \textit{where } R_k[x]=\sum_i x_i R^{ki},\\ \big[\mathcal{S}_\ell^*[\Upsilon_\ell]\big]_{ij} &= \frac{1}{2}\operatorname{Tr}\big(\Upsilon_\ell\big[S_\ell^{\ell i}S_\ell^{\ell j}+S_\ell^{\ell j}S_\ell^{\ell i}\big]\big), \quad \textit{where } S_\ell[y]=\sum_i y_i S^{\ell i},\end{aligned} \tag{3.6}$$

are the mappings (2.7) *associated with* R_k *and* S_ℓ,

$$\phi_{\mathcal{T}}(\lambda)=\max_{t\in\mathcal{T}}\lambda^T t, \quad \phi_{\mathcal{R}}(\lambda)=\max_{r\in\mathcal{R}}\lambda^T r,$$

and

$$\lambda\big[\{\Xi_1,\dots,\Xi_N\}\big]=\big[\operatorname{Tr}(\Xi_1);\dots;\operatorname{Tr}(\Xi_N)\big].$$

3.2.2. Upper-bounding $\Psi_{\mathcal{Q}}(\cdot)$. Our next observation is as follows (for proof, see Section 5.4):

Lemma 3.2. *Let* Y *be a* $m\times\nu$ *matrix,* $Q\in\mathbf{S}^m_+$, *and* P *be a probability distribution on* $\mathbf{R}^m$ *with* $\operatorname{Cov}[P]\preceq Q$. *Let, further,* $\|\cdot\|$ *be a norm on* $\mathbf{R}^\nu$ *with the unit ball* $\mathcal{B}_*$ *of the conjugate norm* $\|\cdot\|_*$ *given by* (3.3). *Finally, let* $\Upsilon=\{\Upsilon_\ell\in\mathbf{S}_+^{f_\ell}\}_{\ell\le L}$ *and a matrix* $\Theta\in\mathbf{S}^m$ *satisfy the constraint*

$$\left[\begin{array}{c|c}\Theta & \frac{1}{2}YM\\ \hline \frac{1}{2}M^TY^T & \sum_\ell \mathcal{S}_\ell^*[\Upsilon_\ell]\end{array}\right]\succeq 0 \tag{3.7}$$

(for notation, see (3.3) and (3.6)*). Then*

$$\mathbf{E}_{\xi\sim P}\big\{\|Y^T\xi\|\big\}\le \operatorname{Tr}(Q\Theta)+\phi_{\mathcal{R}}\big(\lambda[\Upsilon]\big). \tag{3.8}$$

We have the following immediate consequence of Lemma 3.2.

Corollary 3.3. *Let*

$$\Gamma(\Theta)=\max_{Q\in\mathcal{Q}}\operatorname{Tr}(Q\Theta) \tag{3.9}$$

and

$$\overline{\Psi}_{\mathcal{Q}}(H)=\min_{\{\Upsilon_\ell\}_{\ell\le L},\,\Theta\in\mathbf{S}^m}\left\{\Gamma(\Theta)+\phi_{\mathcal{R}}\big(\lambda[\Upsilon]\big):\ \Upsilon_\ell\succeq 0\ \forall\ell,\ \left[\begin{array}{c|c}\Theta & \frac{1}{2}HM\\ \hline \frac{1}{2}M^TH^T & \sum_\ell \mathcal{S}_\ell^*[\Upsilon_\ell]\end{array}\right]\succeq 0\right\}. \tag{3.10}$$

Then $\overline{\Psi}_{\mathcal{Q}}(\cdot):\mathbf{R}^{m\times\nu}\to\mathbf{R}$ *is efficiently computable convex upper bound on* $\Psi_{\mathcal{Q}}(\cdot)$.

Indeed, given Lemma 3.2, the only non-evident part of the corollary is that $\overline{\Psi}_{\mathcal{Q}}(\cdot)$ is a well-defined real-valued function, which is readily given by Lemma 5.1, see Section 5.1.

Remark 3.4. When $\Upsilon = \{\Upsilon_\ell\}_{\ell\leq L}$, Θ is a feasible solution to the right hand side problem in (3.10) and $s > 0$, the pair $\Upsilon' = \{s\Upsilon_\ell\}_{\ell\leq L}$, $\Theta' = s^{-1}\Theta$ also is a feasible solution; since $\phi_{\mathcal{R}}(\cdot)$ and $\Gamma(\cdot)$ are positively homogeneous of degree 1, we conclude that $\overline{\Psi}_{\mathcal{Q}}$ is in fact the infimum of the function

$$2\sqrt{\Gamma(\Theta)\phi_{\mathcal{R}}(\lambda[\Upsilon])} = \inf_{s>0}\left[s^{-1}\Gamma(\Theta) + s\phi_{\mathcal{R}}(\lambda[\Upsilon])\right]$$

over Υ, Θ satisfying the constraints of the problem (3.10).

In addition, for every feasible solution $\Upsilon = \{\Upsilon_\ell\}_{\ell\leq L}$, Θ to the problem (3.10) with $\mathcal{M}[\Upsilon] := \sum_\ell \mathcal{S}_\ell^*[\Upsilon_\ell] \succ 0$, the pair $\Upsilon, \widehat{\Theta} = \frac{1}{4}HM\mathcal{M}^{-1}[\Upsilon]M^TH^T$ is feasible for the problem as well and $0 \preceq \widehat{\Theta} \preceq \Theta$ (Schur Complement Lemma), so that $\Gamma(\widehat{\Theta}) \leq \Gamma(\Theta)$. As a result,

$$\overline{\Psi}_{\mathcal{Q}}(H) = \inf_{\Upsilon}\Big\{\tfrac{1}{4}\Gamma\big(HM\mathcal{M}^{-1}[\Upsilon]M^TH^T\big) + \phi_{\mathcal{R}}(\lambda[\Upsilon]) : \\ \Upsilon = \big\{\Upsilon_\ell \in \mathbf{S}_+^{f_\ell}\big\}_{\ell\leq L},\ \mathcal{M}[\Upsilon] \succ 0\Big\}. \tag{3.11}$$

Illustration. Consider the case when $\|u\| = \|u\|_p$ with $p \in [1,2]$, and let us apply the just described scheme for upper-bounding $\Psi_{\mathcal{Q}}(\cdot)$, assuming $\mathcal{Q} = \{V \in \mathbf{S}_+^m : V \preceq Q\}$ for some given $Q \succ 0$, so that $\Gamma(\Theta) = \mathrm{Tr}(Q\Theta)$, $\Theta \succeq 0$. The unit ball of the norm conjugate to $\|\cdot\|$, that is, the norm $\|\cdot\|_q$, $q = p/(p-1) \in [2,\infty]$, is the basic spectratope (in fact, ellitope)

$$\mathcal{B}_* = \big\{y \in \mathbf{R}^\mu : \exists r \in \mathcal{R} := \big\{\mathbf{R}_+^\nu : \|r\|_{q/2} \leq 1\big\} : S_\ell^2[y] \leq r_\ell,\ 1 \leq \ell \leq L = \nu\big\}, \\ S_\ell[y] = y_\ell.$$

As a result, Υ's from Remark 3.4 are collections of ν positive semidefinite 1×1-matrices, and we can identify them with ν-dimensional nonnegative vectors υ, resulting in $\lambda[\Upsilon] = \upsilon$ and $\mathcal{M}[\Upsilon] = \mathrm{Diag}\{\upsilon\}$. Besides this, for *nonnegative* υ we clearly have $\phi_{\mathcal{R}}(\upsilon) = \|\upsilon\|_{p/(2-p)}$. The optimization problem in (3.11) now reads

$$\overline{\Psi}_{\mathcal{Q}}(H) = \inf_{\upsilon\in\mathbf{R}^\nu}\big\{\tfrac{1}{4}\mathrm{Tr}\big(Q^{1/2}H\,\mathrm{Diag}^{-1}\{\upsilon\}H^TQ^{1/2}\big) + \|\upsilon\|_{p/(2-p)} : \upsilon > 0\big\}.$$

After setting $a_\ell = \|\mathrm{Col}_\ell[Q^{1/2}H]\|_2$, (3.11) becomes

$$\overline{\Psi}_{\mathcal{Q}}(H) = \inf_{\upsilon>0}\left\{\frac{1}{4}\sum_i \frac{a_i^2}{\upsilon_i} + \|\upsilon\|_{p/(2-p)}\right\},$$

resulting in $\overline{\Psi}_Q(H) = \|[a_1;\dots;a_\nu]\|_p$. Recalling what are a_i's, we end up with

$$\Psi_{\mathcal{Q}}(H) \leq \overline{\Psi}_{\mathcal{Q}}(H) := \left\|\left[\|\operatorname{Col}_1[Q^{1/2}H]\|_2;\dots;\|\operatorname{Col}_\nu[Q^{1/2}H]\|_2\right]\right\|_p. \quad (3.12)$$

Note that the bound (3.12) can be easily improved when $\xi \sim \mathcal{N}(0,Q)$. Indeed, in this case $\eta = H^T\xi$ is normal with components $\eta_i \sim \mathcal{N}(0,a_i^2)$, $a_i = \|\operatorname{Col}_\ell[Q^{1/2}H]\|_2$, and therefore

$$\Psi_{\mathcal{Q}}(H) = \mathbf{E}\{\|\eta\|_p\} \leq \left[\mathbf{E}_\eta\{\|\eta\|_p^p\}\right]^{1/p} = \frac{\sqrt{2}\,\Gamma(p+1)/2}{\pi^{1/2p}}\left[\sum_i a_i^p\right]^{1/p}$$
$$= \frac{\sqrt{2}\,\Gamma(p+1)/2}{\pi^{1/2p}}\|[a_1;\dots;a_\nu]\|_p =: \widetilde{\Psi}_{\mathcal{Q}}(H)\left[\leq \|[a_1;\dots;a_\nu]\|_p\right].$$

For instance, when $p=1$ the bound $\widetilde{\Psi}_{\mathcal{Q}}(H)$ becomes exact and equals

$$\sqrt{\frac{2}{\pi}}\sum_i a_i = \sqrt{\frac{2}{\pi}}\Psi_{\mathcal{Q}}(H).$$

3.2.3. Putting things together: building presumably good linear estimate. Corollaries 3.1 and 3.3 imply the following recipe for building a "presumably good" linear estimate:

Proposition 3.5. *In the situation of Section 3.1 and under Assumption A, consider the convex optimization problem (for notation, see* (3.6) *and* (3.9)*)*

$$\begin{array}{l}\mathrm{Opt} = \min\limits_{H,\Lambda,\Upsilon,\Upsilon',\Theta}\Big\{\phi_{\mathcal{T}}(\lambda[\Lambda]) + \phi_{\mathcal{R}}(\lambda[\Upsilon]) + \phi_{\mathcal{R}}(\lambda[\Upsilon']) + \Gamma(\Theta):\\ \quad \Lambda = \{\Lambda_k \succeq 0, k\leq K\},\ \Upsilon = \{\Upsilon_\ell \succeq 0, \ell \leq L\},\ \Upsilon' = \{\Upsilon'_\ell \succeq 0, \ell\leq L\},\\ \quad \left[\begin{array}{c|c}\sum_k \mathcal{R}_k^*[\Lambda_k] & \frac{1}{2}[B^T - A^TH]M\\ \hline \frac{1}{2}M^T[B-H^TA] & \sum_\ell \mathcal{S}_\ell^*[\Upsilon_\ell]\end{array}\right] \succeq 0,\\ \quad \left[\begin{array}{c|c}\Theta & \frac{1}{2}HM\\ \hline \frac{1}{2}M^TH^T & \sum_\ell \mathcal{S}_\ell^*[\Upsilon'_\ell]\end{array}\right] \succeq 0\Big\}. \end{array} \quad (3.13)$$

The problem is solvable, and the H-component H_* of its optimal solution yields linear estimate $\widehat{x}_{H_*}(\omega) = H_*^T\omega$ such that

$$\mathrm{Risk}_{\mathcal{Q},\|\cdot\|}\left[\widehat{x}(\cdot)|\mathcal{X}\right] \leq \mathrm{Opt}. \quad (3.14)$$

The only claim in Proposition 3.5 which is not an immediate consequence of Corollaries 3.1, 3.3 is that problem (3.13) is solvable; this claim is readily given by the fact that the objective clearly is coercive on the feasible set (recall that $\Gamma(\Theta)$ is coercive on $\mathbf{S}_+^m$ due to $\mathcal{Q} \subset \operatorname{int}\mathbf{S}_+^m$ and that $y \mapsto My$ is an onto mapping, since $\mathcal{B}_*$ is full-dimensional).

Remark 3.6. In some applications, observations (1.1) have additional structure, namely, ω is a T-element sample: $\omega = [\bar\omega_1; \dots; \bar\omega_T]$ with components

$$\bar\omega_t = \bar A x + \xi_t, \quad t = 1, \dots, T,$$

and ξ_t are i.i.d. observation noises with zero mean distribution $\bar P$ satisfying $\bar P \lessdot \bar{\mathcal{Q}}$ for some convex compact set $\bar{\mathcal{Q}} \subset \operatorname{int} \mathbf{S}^{\bar m}_+$. In other words, we deal with *repeated observations*, where for $m = T\bar m$,

$$\begin{array}{l} A = [\underbrace{\bar A; \dots; \bar A}_{T}] \in \mathbf{R}^{m\times n} \quad \text{for some } \bar A \in \mathbf{R}^{\bar m \times n}, \\ \mathcal{Q} = \{Q = \mathrm{Diag}\{\underbrace{\bar Q, \dots, \bar Q}_{T}\}, \ \bar Q \in \bar{\mathcal{Q}}\}. \end{array} \tag{3.15}$$

It can be easily verified (see Section 5.5) that in the case of repeated observations the optimization problem (3.13) responsible for the presumably good linear estimate reduces to similar problem *with size independent of* T:

Proposition 3.7. *In the case of repeated observations and under Assumption A, the linear estimate of Bx yielded by an optimal solution to problem (3.13) can be computed as follows. Consider the convex optimization problem*

$$\begin{array}{l} \overline{\mathrm{Opt}} = \min\limits_{\bar H, \Lambda, \Upsilon, \Upsilon', \bar\Theta} \Big\{ \phi_{\mathcal{T}}(\lambda[\Lambda]) + \phi_{\mathcal{R}}(\lambda[\Upsilon]) + \phi_{\mathcal{R}}(\lambda[\Upsilon']) + \frac{1}{T}\bar\Gamma(\bar\Theta) : \\ \quad \Lambda = \{\Lambda_k \succeq 0, k \le K\},\ \Upsilon = \{\Upsilon_\ell \succeq 0, \ell \le L\},\ \Upsilon' = \Upsilon'_\ell \succeq 0, \ell \le L\}, \\ \quad \left[\begin{array}{c|c} \sum_k \mathcal{R}_k^*[\Lambda_k] & \frac{1}{2}[B^T - \bar A^T \bar H] M \\ \hline \frac{1}{2} M^T [B - \bar H^T \bar A] & \sum_\ell \mathcal{S}_\ell^*[\Upsilon_\ell] \end{array}\right] \succeq 0, \\ \quad \left[\begin{array}{c|c} \bar\Theta & \frac{1}{2}\bar H M \\ \hline \frac{1}{2} M^T \bar H^T & \sum_\ell \mathcal{S}_\ell^*[\Upsilon'_\ell] \end{array}\right] \succeq 0 \Big\}, \end{array} \tag{3.16}$$

where

$$\bar\Gamma(\bar\Theta) = \max_{\bar Q \in \bar{\mathcal{Q}}} \mathrm{Tr}(\bar Q \bar\Theta).$$

The problem is solvable, and the estimate in question is yielded by the $\bar H$-component $\bar H_*$ of the optimal solution according to

$$\hat x([\bar\omega_1; \dots; \bar\omega_T]) = \frac{1}{T} \bar H_*^T \sum_{t=1}^T \bar\omega_t,$$

and the upper bound, provided by Proposition 3.5, on the risk $\mathrm{Risk}_{\mathcal{Q},\|\cdot\|}[\hat x(\cdot)|\mathcal{X}]$ *of this estimate is* $\overline{\mathrm{Opt}}$.

3.3. Near-optimality in Gaussian case. The bound (3.14) for the risk of the linear estimate $\widehat{x}_{H_*}(\cdot)$ constructed in (3.13) can be compared to the minimax optimal risk of recovering Bx, $x \in \mathcal{X}$, from observations corrupted by zero mean Gaussian noise with covariance matrix from $\mathcal{Q}$; formally, this minimax optimal risk is defined as

$$\mathrm{RiskOpt}_{\mathcal{Q},\|\cdot\|}[\mathcal{X}] = \sup_{Q\in\mathcal{Q}} \inf_{\widehat{x}(\cdot)} \Big[\sup_{x\in\mathcal{X}} \mathbf{E}_{\xi\sim\mathcal{N}(0,Q)} \big\{ \|Bx - \widehat{x}(Ax+\xi)\| \big\} \Big],$$

where the infimum is taken over all estimates.

Proposition 3.8. *Under the premise and in the notation of Proposition 3.5, let*

$$M_*^2 = \max_W \big\{ \mathbf{E}_{\eta\sim\mathcal{N}(0,I_n)} \|BW^{1/2}\eta\|^2 : \\ W \in \mathcal{W} := \{W \in \mathbf{S}^n_+ : \exists t \in \mathcal{T} : \mathcal{R}_k[W] \preceq t_k I_{d_k},\ 1 \le k \le K\} \big\}; \quad (3.17)$$

we have

$$\mathrm{Risk}_{\mathcal{Q},\|\cdot\|}\left[\widehat{x}_{H_*}|\mathcal{X}\right] \le \mathrm{Opt} \le C\sqrt{\ln(2F)\ln\left(\frac{2DM_*^2}{\mathrm{RiskOpt}^2_{\mathcal{Q},\|\cdot\|}[\mathcal{X}]}\right)}\,\mathrm{RiskOpt}_{\mathcal{Q},\|\cdot\|}[\mathcal{X}], \quad (3.18)$$

where C is a positive absolute constant, and

$$D = \sum_k d_k, \quad F = \sum_\ell f_\ell. \quad (3.19)$$

For the proof, see Section 5.7.

Remark 3.9. The idea of the proof of Proposition 3.8 originates from [24]. Namely, our goal is to upper-bound the optimal value Opt in the optimization problem (3.13), which is an upper bound on the risk of the presumably good linear estimate yielded by Proposition 3.5, in terms of the minimax optimal risk $\mathrm{RiskOpt}_{\mathcal{Q},\|\cdot\|}[\mathcal{X}]$. To this end, we show that:

(1) Opt can be upper-bounded in terms of the optimal *Bayesian risk*

$$\varrho[W,Q] = \inf_{\widehat{x}(\cdot)} \mathbf{E}_{[\eta,\xi]\sim\mathcal{N}(0,W)\times\mathcal{N}(0,Q)} \big\{ \|B\eta - \widehat{x}(A\eta+\xi)\| \big\}$$

associated with observation noise $\xi \sim \mathcal{N}(0,Q)$ and independent of this noise *random Gaussian* signal $\eta \sim \mathcal{N}(0,W)$, with properly selected $Q \in \mathcal{Q}$ and $W \in \mathbf{S}^n_+$;

(2) The above "properly selected W and Q" can be chosen in such a way that the Bayesian risk $\varrho[W,Q]$ can, in turn, be upper-bounded in terms of the minimax optimal risk $\mathrm{RiskOpt}_{\mathcal{Q},\|\cdot\|}[\mathcal{X}]$.

Combining (1) and (2), we arrive at the desired upper-bounding of Opt in terms of $\mathrm{RiskOpt}_{\mathcal{Q},\|\cdot\|}[\mathcal{X}]$.

It should be stressed that implementing the outlined, by itself quite transparent, strategy is by far not straightforward; the main ingredients of the implementation are Lemma 5.3 and the (far from being evident) fact that the upper bound (3.8) on $\mathbf{E}\{\|Y^T\xi\|\}$ in the case of zero mean Gaussian ξ is tight, as stated by the following lemma (see Section 5.6 for the proof).

Lemma 3.10. *Let Y be an $N \times \nu$ matrix, let $\|\cdot\|$ be a norm on $\mathbf{R}^\nu$ such that the unit ball $\mathcal{B}_*$ of the conjugate norm is the spectratope* (3.3)*, and let $\zeta \sim \mathcal{N}(0, Q)$ for some positive semidefinite $N \times N$ matrix Q. Then the best upper bound on $\psi_Q(Y) := \mathbf{E}\{\|Y^T\zeta\|\}$ yielded by Lemma 3.2, that is, the optimal value* Opt$[Q]$ *in the convex optimization problem (cf.* (3.10)*)*

$$\mathrm{Opt}[Q] = \min_{\Theta,\Upsilon}\Big\{\phi_{\mathcal{R}}(\lambda[\Upsilon]) + \mathrm{Tr}(Q\Theta) : \Upsilon = \{\Upsilon_\ell \succeq 0, 1 \le \ell \le L\},\ \Theta \in \mathbf{S}^N,$$
$$\left[\begin{array}{c|c} \Theta & \frac{1}{2}YM \\ \hline \frac{1}{2}M^TY^T & \sum_\ell \mathcal{S}_\ell^*[\Upsilon_\ell] \end{array}\right] \succeq 0\Big\} \tag{3.20}$$

(for notation, see (3.6) *and Lemma 3.2) satisfies the identity*

$$\forall (Q \succeq 0) : \mathrm{Opt}[Q] = \overline{\mathrm{Opt}}[Q] := \min_{G,\Upsilon=\{\Upsilon_\ell,\ell\le L\}}\Big\{\phi_{\mathcal{R}}(\lambda[\Upsilon]) + \mathrm{Tr}(G) : \Upsilon_\ell \succeq 0,$$
$$\left[\begin{array}{c|c} G & \frac{1}{2}Q^{1/2}YM \\ \hline \frac{1}{2}M^TY^TQ^{1/2} & \sum_\ell \mathcal{S}_\ell^*[\Upsilon_\ell] \end{array}\right] \succeq 0\Big\}, \tag{3.21}$$

and is a tight upper bound on $\psi_Q(Y)$, namely,

$$\psi_Q(Y) \le \mathrm{Opt}[Q] \le \frac{8\sqrt{\ln\left(4\sqrt{2}F/(\sqrt{2}-\mathrm{e}^{1/4})\right)}}{\sqrt{2}-\mathrm{e}^{1/4}}\,\psi_Q(Y) \le 62\sqrt{\ln(44F)}\,\psi_Q(Y), \tag{3.22}$$

where $F = \sum_\ell f_\ell$ is the size of the spectratope (3.3)*.*

3.4. Illustration: covariance matrix estimation via indirect observations. Suppose that we observe a sample

$$\{\eta_t = A\xi_t\}_{t\le T}, \tag{3.23}$$

where A is a given $m \times n$ matrix, and $\xi_1, \dots, \xi_T$ are sampled, independently of each other, from zero mean Gaussian distribution with unknown covariance matrix ϑ known to satisfy

$$\gamma\vartheta_* \preceq \vartheta \preceq \vartheta_*, \tag{3.24}$$

where $\gamma \ge 0$ and $\vartheta_* \succ 0$ are given. Our goal is to recover the linear image $\mathcal{B}(\theta)$ of θ, and the norm in which recovery error is measured satisfies Assumption A.

For the covariance estimation problem to fit the framework presented in the previous section we reformulate it as follows.

1. We represent the set $\{\vartheta \in \mathbf{S}^n_+ : \gamma\vartheta_* \preceq \vartheta \preceq \vartheta_*\}$ as the image of the basic spectratope (matrix box)

$$\mathcal{V} = \{v \in \mathbf{S}^n : \|v\|_{\mathrm{Sh},\infty} \leq 1\} \quad \left[\|\cdot\|_{\mathrm{Sh},\infty}\text{: the spectral norm}\right]$$

under affine mapping: we set $\vartheta_0 = ((1+\gamma)/2)\vartheta_*$, $\sigma = (1-\gamma)/2$, and treat the matrix

$$v = \sigma^{-1}\vartheta_*^{-1/2}(\vartheta - \vartheta_0)\vartheta_*^{-1/2} \quad \left[\Leftrightarrow \vartheta = \vartheta_0 + \sigma\vartheta_*^{1/2}v\vartheta_*^{1/2}\right]$$

as the signal underlying our observations. Note that a priori information (3.24) on ϑ reduces to $v \in \mathcal{V}$.

2. We pass from observations η_t to "lifted" observations $\eta_t\eta_t^T \in \mathbf{S}^m$, so that

$$\mathbf{E}\{\eta_t\eta_t^T\} = \mathbf{E}\{A\xi_t\xi_t^TA^T\} = A\vartheta A^T = A\underbrace{\left(\vartheta_0 + \sigma A\vartheta_*^{1/2}v\vartheta_*^{1/2}\right)}_{\vartheta[v]}A^T,$$

and treat as "actual" observations the matrices

$$\omega_t = \eta_t\eta_t^T - A\vartheta_0A^T, \quad 1 \leq t \leq T.$$

We have[8]

$$\omega_t = \mathcal{A}v + \zeta_t \quad \text{with } \mathcal{A}v = \sigma A\vartheta_*^{1/2}v\vartheta_*^{1/2}A^T \text{ and } \zeta_t = \eta_t\eta_t^T - A\vartheta[v]A^T. \quad (3.25)$$

Observe that random matrices $\zeta_1, \ldots, \zeta_T$ are i.i.d. with zero mean and covariance mapping $\mathcal{C}[v]$ (that of the random matrix-valued variable $\zeta = \eta\eta^T - \mathbf{E}\{\eta\eta^T\}$, $\eta \sim \mathcal{N}(0, A\vartheta[v]A^T)$) which satisfies (see Section 5.3 for the derivation)

$$\forall v \in \mathcal{V} : \mathcal{C}[v] \preceq \mathcal{Q}, \quad \langle e, \mathcal{Q}h\rangle = 2\,\mathrm{Tr}(\vartheta_*A^ThA\vartheta_*A^TeA), \quad e, h \in \mathbf{S}^m. \quad (3.26)$$

We have represented the problem of interest in the form described in Section 3.1 and have specified all required data.

Numerical illustration. Here we report on preliminary numerical experiments with the estimation problem stated above. They are restricted to the *diagonal case* where $A \in \mathbf{R}^{n\times n}$ is nonsingular, $\mathcal{B}(\theta) = B\theta B^T$ with $B \in \mathbf{R}^{n\times n}$, $\vartheta_* = I_n$, and $\gamma = 0$; our goal is to recover $\mathcal{B}(\theta) \in \mathbf{S}^n$ in the Frobenius ("Frobenius norm case"), or in

[8]In our current considerations, we need to operate with linear mappings acting from $\mathbf{S}^p$ to $\mathbf{S}^q$. We treat $\mathbf{S}^k$ as Euclidean space equipped with the Frobenius inner product $\langle u, v\rangle = \mathrm{Tr}(uv)$ and denote linear mappings from $\mathbf{S}^p$ into $\mathbf{S}^q$ by capital calligraphic letters, like $\mathcal{A}$, $\mathcal{Q}$, etc. Thus, $\mathcal{A}$ in (3.25) denotes the linear mapping which, on a close inspection, maps matrix $v \in \mathbf{S}^n$ into the matrix $\mathcal{A}v = A[\vartheta[v] - \vartheta[0]]A^T$.

the nuclear norm ("nuclear norm case"). Our first observation is that the case of square nonsingular A reduces immediately to the case of direct observations $A = I_n$; to this end it suffices to treat, as observations, vectors $\xi_t = A^{-1}\eta_t$, see (3.23). It is easily seen that the estimate given by Proposition 3.7 is "intelligent enough" to recognize this possibility. Furthermore, the case of $B \in \mathbf{R}^{n\times n}$ reduces to the case of diagonal B: if $B = UDV^T$ is the singular value decomposition of B, with our ϑ_* and choice of the norm, we lose nothing when replacing B with D, and our design again recognizes this possibility. Therefore, from the start, in our experiment we assume that $A = I_n$ and $\mathcal{B}(\vartheta) = B\vartheta B^T$ with diagonal B, and

$$\vartheta_0 = \tfrac{1}{2}I_n, \quad \sigma = \tfrac{1}{2}, \quad \vartheta[v] = \tfrac{1}{2}I_n + \tfrac{1}{2}v, \quad \mathcal{A}v = \tfrac{1}{2}v.$$

Thus, the estimation problem in question is reduced to that of recovering the matrix

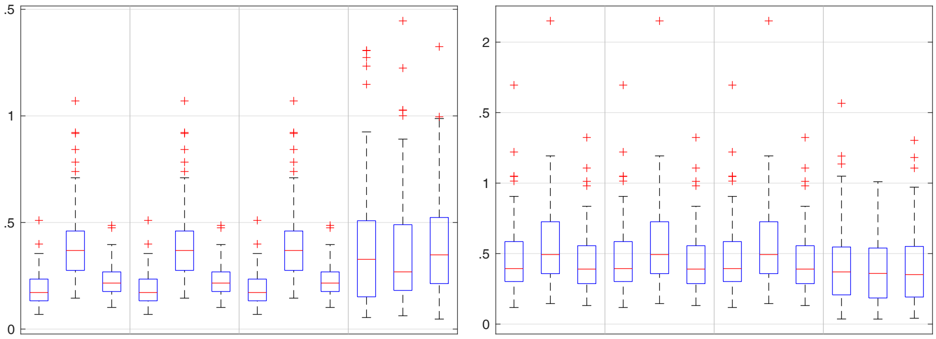
$$B\vartheta[v]B^T = \tfrac{1}{2}B^2 + \tfrac{1}{2}BvB$$

from observations (3.25) stemming from a signal v known to satisfy $v \in \mathcal{V} = \{v \in \mathbf{S}^n : v^2 \preceq I_n\}$. The outlined setup, as compared to the general one, simplifies dramatically optimization problem (3.16) (for details, see Appendix B) and allows to run experiments with n in the range of hundreds.

In our simulations, we use $T \in \{32, 128, 512\}$ and diagonal matrix B with diagonal entries $B_{ii} = i^{-\beta}$, with β running through $\{0, 1, 2, 3\}$. For every combination of T and β from the just outlined ranges, we compute, in the Frobenius and the nuclear norm cases, the linear estimate and (the upper bound on) its risk as given by Proposition 3.7. Next, we run $K = 100$ simulations and record the actual recovery errors as yielded by the linear estimate and the Maximum Likelihood estimate (MLE) in the role of the reference point.[9] In our experiments, the covariance matrices underlying observations were generated as random rotations of diagonal matrices with diagonal entries drawn, independently of each other, from the uniform distribution on [0, 1].

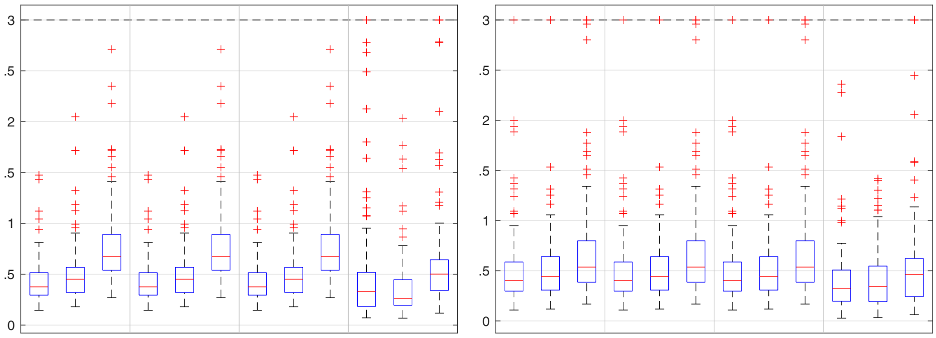
The results of experiments with covariance matrix of size $n = 128$ are presented in Figures 1 and 2. The first figure displays the boxplots of the ratios of actual errors of the linear estimators to the theoretical upper risk bounds for the Frobenius norm case (plot (a)) and nuclear norm case (plot (b)). Each of four boxplot groups corresponds, from left to right, to $\beta = 0$, 1, 2, and 3; three boxplots inside each group correspond to the observation sample lengths $T = 32$, 64, and 128. Boxplots for ratios of errors of linear estimation to those of MLE for each simulation are displayed in Figure 2. The "ordering" of boxplots is the same as in Figure 1; for better readability of the plots

[9]It is immediately seen that with our setup the ML estimate is as follows: given observations η_t, $1 \le t \le T$ (see (3.23) and recall that in our case $A = I_n$), we compute the empirical covariance matrix $\widehat{C} = \frac{1}{T}\sum_{t=1}^T \eta_t\eta_t^T$. The MLE $\widehat{\vartheta}$ of the covariance matrix ϑ of η_t's is obtained by keeping the eigenvectors of $\widehat{C}$ intact and projecting the eigenvalues of $\widehat{C}$ onto $[0, 1]$. The resulting MLE of $\mathcal{B}(\vartheta) = B\vartheta B$ is $B\widehat{\vartheta}B$.



(a) Frobenius norm case. (b) Nuclear norm case.

Figure 1. Ratios of empirical errors of the linear estimation to the upper risk bounds. Boxplots for 100 realisations of randomized experiments.



(a) Frobenius norm case. (b) Nuclear norm case.

Figure 2. Boxplots for ratios of empirical errors of the linear estimation to the error of the MLE; 100 realisations of randomized experiments.

the data is "clipped" at the level 3. We see that no estimate "uniformly dominates" the other one, and that the linear estimate outperforms the MLE when the number T of observations is relatively low.[10]

4. Linear estimation under "uncertain-but-bounded" noise

We present here another application of the result of Proposition 2.2 — construction of a linear estimate of a signal in the case of uncertain but bounded perturbation ξ in the observation (1.1).

[10]The fact that the relative to linear estimation performance of MLE improves as T grows is completely natural — the latter estimate is asymptotically optimal as $T \to \infty$.

4.1. Problem statement. Consider an estimation problem where, given an observation

$$\omega = Ax + \xi$$

of unknown signal x, known to belong to a given signal set $\mathcal{X}$, one wants to recover linear image Bx of x. Here A and B are given $m \times n$ and $\nu \times n$ matrices. Suppose that all we know about ξ is that it belongs to a given compact set $\mathcal{H}$ ("uncertain-but-bounded observation noise"). In the situation in question, given a norm $\|\cdot\|$ on $\mathbf{R}^\nu$, we quantify the accuracy of a candidate estimate $\omega \mapsto \widehat{x}(\omega)$ by its maximal on $\mathcal{X}$ risk

$$\mathrm{Risk}_{\mathcal{H}}\big[\widehat{x}|\mathcal{X}\big] = \sup_{x\in X,\,\xi\in\mathcal{H}} \|Bx - \widehat{x}(Ax+\xi)\|$$

("$\mathcal{H}$-risk").

This is a standard problem of *optimal recovery* (see, e.g. [19,20]). It is well known that when $\mathcal{H}$ and $\mathcal{X}$ are convex compact sets, when specifying $\widehat{x}(\omega)$ as (any) point from $\{x \in X : \omega - Ax \in \mathcal{H}\}$, we get a minimax optimal, within factor 2, estimate, see also [27,28].We are about to show that when $\mathcal{X}$ and $\mathcal{H}$ are spectratopes, and the unit ball of the norm $\|\cdot\|_*$ conjugate to $\|\cdot\|$ is a basic spectratope, an efficiently computable *linear in observation* estimate $\widehat{x}_H = H\omega$ is near-optimal in terms of its $\mathcal{H}$-risk.[11]

Our initial observation is that the situation in question reduces straightforwardly to that where there is no observation noise at all. Indeed, let $\mathcal{Y} = \mathcal{X} \times \mathcal{H}$; then $\mathcal{Y}$ is a spectratope, and we lose nothing when assuming that the signal underlying observation ω is $y = [x;\xi] \in \mathcal{Y}$:

$$\omega = Ax + \xi = \bar{A}y, \quad \bar{A} = [A, I_m],$$

while the entity to be recovered is

$$Bx = \bar{B}y, \quad \bar{B} = [B, 0_{\nu\times m}].$$

With these conventions, the observation noise vanishes, while the $\mathcal{H}$-risk of a candidate estimate $\widehat{x}(\cdot)\colon \mathbf{R}^m \to \mathbf{R}^\nu$ becomes the quantity

$$\mathrm{Risk}_{\|\cdot\|}\big[\widehat{x}|\mathcal{X}\times\mathcal{H}\big] = \sup_{y=[x;\xi]\in\mathcal{X}\times\mathcal{H}} \|\bar{B}y - \widehat{x}(\bar{A}y)\|.$$

To streamline the notation, let us assume that the outlined reduction has already been carried out, so the problem of interest reads: given an observation

$$\omega = Ax \in \mathbf{R}^m,$$

[11]In the case where an "efficient description" of the sets $\mathcal{H}$, $\mathcal{X}$ and the norm $\|\cdot\|$ is available, a minimax optimal, within factor 2, nonlinear estimate can be computed efficiently. On the other hand, its risk is generally hard to compute. Note that the linear estimate we discuss here, which comes with a "reasonably tight" upper bound on its risk, can be of "numerical" interest in the situation where estimates are to be computed repeatedly for different observations sharing common problem data — sets $\mathcal{H}$, $\mathcal{X}$ and the norm $\|\cdot\|$.

estimate the linear image $Bx \in \mathbf{R}^\nu$ of an unknown signal x known to belong to a given spectratope $\mathcal{X}$. The risk of a candidate estimate $\widehat{x}$ is defined as

$$\mathrm{Risk}_{\|\cdot\|}\left[\widehat{x}|\mathcal{X}\right] = \sup_{x\in\mathcal{X}} \|Bx - \widehat{x}(Ax)\|,$$

and the norm $\|\cdot\|$ is such that the unit ball $\mathcal{B}_*$ of the norm $\|\cdot\|_*$ *conjugate* to $\|\cdot\|$ is a *basic spectratope:*

$$\mathcal{B}_* := \left\{u \in \mathbf{R}^\nu : \|u\|_* \le 1\right\} = \left\{u \in \mathbf{R}^\nu : \exists r \in \mathcal{R} : S_\ell^2[u] \preceq r_\ell I_{f_\ell},\ 1 \le \ell \le L\right\},$$

where the right hand side data are as required in a spectratopic representation. By the same reasoning as in Section 3.1, we lose nothing when assuming from now on that *the signal set is a basic spectratope:*

$$\mathcal{X} = \left\{x \in \mathbf{R}^n : \exists t \in \mathcal{T} : R_k^2[x] \preceq t_k I_{d_k},\ 1 \le k \le K\right\}.$$

4.2. Near-optimality of linear estimation. Let $\widehat{x}_H(\omega) = H^T\omega$ be a linear estimate. We have

$$\begin{aligned}\mathrm{Risk}_{\|\cdot\|}\left[\widehat{x}_H|\mathcal{X}\right] &= \max_{x\in\mathcal{X}} \left\|\left(B - H^TA\right)x\right\| \\ &= \max_{[u;x]\in\mathcal{B}_*\times\mathcal{X}} [u;x]^T \left[\begin{array}{c|c} & \frac{1}{2}\left(B - H^TA\right) \\ \hline \frac{1}{2}\left(B - H^TA\right)^T & \end{array}\right][u;x].\end{aligned}$$

Applying Proposition 2.2, we arrive at item (i) of the following proposition (cf. Corollary 3.1):

Proposition 4.1. *In the situation of this section, consider the convex optimization problem*

$$\begin{aligned}\overline{\mathrm{Opt}} = \min_{H,\Upsilon=\{\Upsilon_\ell\},\Lambda=\{\Lambda_k\}} \Bigg\{&\phi_{\mathcal{R}}\left(\lambda[\Upsilon]\right) + \phi_{\mathcal{T}}\left(\lambda[\Lambda]\right) : \Upsilon_\ell \succeq 0,\ \Lambda_k \succeq 0,\ \forall(\ell,k), \\ &\left[\begin{array}{c|c} \sum_\ell \mathcal{S}_\ell^*[\Upsilon_\ell] & \frac{1}{2}\left(B - H^TA\right) \\ \hline \frac{1}{2}\left(B - H^TA\right)^T & \sum_k \mathcal{R}_k^*[\Lambda_k] \end{array}\right] \succeq 0\Bigg\}, \end{aligned} \tag{4.1}$$

where $\mathcal{R}_k^*[\cdot]$ and $\mathcal{S}_\ell^*[\cdot]$ are induced by $R_k[\cdot]$, resp., $S_k[\cdot]$, as explained in Section 2.1.

(i) *The problem is solvable, and the risk of the linear estimate $\widehat{x}_{H_*}(\cdot)$ yielded by the H-component of an optimal solution to (4.1) does not exceed $\overline{\mathrm{Opt}}$.*

(ii) *The linear estimate $\widehat{x}_{H_*}$ is near-optimal in terms of its $\mathcal{H}$-risk:*

$$\mathrm{Risk}_{\|\cdot\|}\left[\widehat{x}_{H_*}|\mathcal{X}\right] \le \overline{\mathrm{Opt}} \le 2\ln(2D)\,\mathrm{Risk}_{\mathrm{opt}}[\mathcal{X}], \quad D = \sum_k d_k + \sum_\ell f_\ell, \tag{4.2}$$

where $\mathrm{Risk}_{\rm opt}[\mathcal{X}]$ *is the minimax optimal risk:*

$$\mathrm{Risk}_{\rm opt}[\mathcal{X}] = \inf_{\widehat{x}} \mathrm{Risk}_{\|\cdot\|}\left[\widehat{x}|\mathcal{X}\right],$$

where inf *is taken w.r.t. all possible estimates.*

For a proof, see Section 5.8.

4.3. Numerical illustration. The construction[12] from the proof of Proposition 4.1 item (ii) can be used to lower bound numerically the minimax risk $\mathrm{Risk}_{\|\cdot\|}[\widehat{x}_{H_*}|\mathcal{X}]$, and we can compare the resulting lower bound on the "true" minimax risk with the upper bound (4.1) on the risk of the linear estimate yielded by our approach, thus quantifying numerically its conservatism.

We have conducted two experiments of the outlined type. In both experiments the signal set $\mathcal{X}$ is the box

$$\mathcal{X} = \left\{x \in \mathbf{R}^n : j|x_j| \le 1,\ 1 \le j \le n\right\} \qquad \left[K = n,\ R_k = k^2 e_k e_k^T,\ k = 1, \dots, K,\ T = [0,1]^K\right],$$

B is the $n \times n$ identity matrix, and $\frac{n}{2} \times n$ sensing matrix A is a randomly rotated matrix with singular values λ_j, $1 \le j \le n$, forming a geometric progression, with $\lambda_1 = 1$ and $\lambda_{n/2} = 0.01$. In the first experiment the "dual-norm spectratope" $\mathcal{B}_*$ is a random parallelotope

$$\mathcal{B}_*^{(P)} = \left\{u \in \mathbf{R}^n : |\eta_i^T x| \le 1,\ 1 \le i \le n\right\} \quad \left[L = n,\ \mathcal{R} = [0,1]^L\right].$$

In the second experiment $\mathcal{B}_*$ is a random "matrix box"

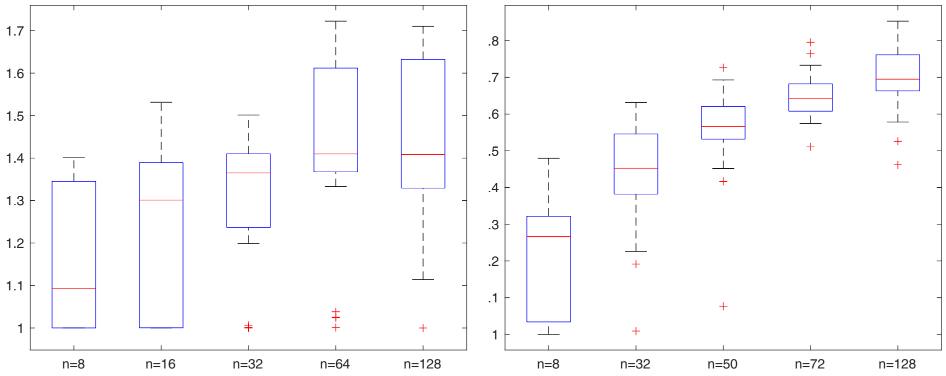
$$\mathcal{B}_*^{(M)} = \left\{u \in \mathbf{R}^n : \left\|\sum_{i=1}^n S^i x_i\right\|_{\mathrm{Sh},\infty} \le 1\right\} \quad \left[L = 1,\ \mathcal{R} = [0,1]\right],$$

where $S^i \in \mathbf{R}^{d\times d}$ are random symmetric matrices ($n = d^2/2$ in the reported experiments). With a natural implementations of the outlined bounding scheme we arrive at simulation results presented in Figure 3. Observe that in all experiments (100 random problems for each problem dimension) the suboptimality factor does not exceed 1.9, while its theoretical estimation as in (4.2) varies in the interval [9.7, 22.2].

[12] In short, the idea of the construction is as follows. We first note that the maximal norm $\|Bx\|$ for x in the intersection of $\mathcal{X}$ and of the kernel of A, i.e. the optimal value of the problem

$$\max_x\{\|Bx\| : x \in \mathcal{X},\ Ax = 0\} = \max_{x,u}\{u^T Bx,\ u \in \mathcal{B}_*,\ x \in \mathcal{X},\ Ax = 0\}, \qquad (*)$$

lower bounds the minimax risk. Then we use semidefinite relaxation to compute a feasible solution $[\overline{u}; \overline{x}]$ to ($*$) and use the value $\overline{u}^T B\overline{x}$ to lower bound $\mathrm{Risk}_{\rm opt}[\mathcal{X}]$. The reader is referred to the proofs of Propositions 2.2 and 4.1 for details.



(a) Random parallelotope $\mathcal{B}_*^{(P)}$. (b) Random "matrix box" $\mathcal{B}_*^{(M)}$.

Figure 3. Suboptimality factors as functions of the problem dimension. Boxplots for 100 realisations of randomized experiments.

Remark 4.2. Note that Propositions 3.5 and 3.8 also apply in the following "mixed" observation scheme:

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

where, as above, A is a given $m \times n$ matrix, x us unknown deterministic signal known to belong to a given signal set $\mathcal{X}$, ξ is a random noise with distribution known to belong to a family $\mathcal{P}$ of Borel probability distributions on $\mathbf{R}^m$ satisfying (1.2) for a given convex compact set $\mathcal{Q} \subset \operatorname{int} \mathbf{S}^m_+$, and η is "uncertain-but-bounded" perturbation known to belong to a given set $\mathcal{H}$. As before, our goal is to recover $Bx \in \mathbf{R}^\nu$ via observation ω. Given a norm $\|\cdot\|$ on $\mathbf{R}^\nu$, we can quantify the performance of a candidate estimate $\omega \mapsto \widehat{x}(\omega)\colon \mathbf{R}^m \to \mathbf{R}^\nu$ by its risk

$$\mathrm{Risk}_{\mathcal{Q},\mathcal{H},\|\cdot\|}\left[\widehat{x}|\mathcal{X}\right] = \sup_{x\in\mathcal{X}, P\preccurlyeq\mathcal{Q}, \eta\in\mathcal{H}} \mathbf{E}_{\xi\sim P}\left\{\|Bx - \widehat{x}(Ax + \xi + \eta)\|\right\}.$$

Observe that the estimation problem associated with this "mixed" observation scheme straightforwardly reduces to similar problem for random observation scheme, by the same trick we have used in Section 4.1 to eliminate the observation noise. Indeed, let us treat $x^+ := [x; \eta] \in \mathcal{X}^+ := \mathcal{X} \times \mathcal{H}$ and $\mathcal{X}^+$ as the new signal/signal set underlying our observation, and denote $\bar{A}x^+ = Ax + \eta$, $\bar{B}x^+ = Bx$, where $\bar{A} = [A, I_m]$ and $\bar{B} = [B, 0_{\nu\times m}]$. With these conventions, the "mixed" observation scheme becomes

$$\omega = \bar{A}x^+ + \xi,$$

and for every candidate estimate $\widehat{x}(\cdot)$ it clearly holds

$$\mathrm{Risk}_{\mathcal{Q},\mathcal{H},\|\cdot\|}\left[\widehat{x}|\mathcal{X}\right] = \mathrm{Risk}_{\mathcal{Q},\|\cdot\|}\left[\widehat{x}|\mathcal{X}^+\right].$$

In other words, we are now in the situation of Section 3.1; assuming that $\mathcal{X}$ and $\mathcal{H}$ are spectratopes, so is $\mathcal{X}^+$, meaning that all results of Section 3 on constructing linear estimates and their near-optimality are applicable in our present setup.

Note that within its scope (uncertain-but-bounded perturbation), Proposition 4.1 provides a stronger near-optimality characterisation of linear estimates than Proposition 3.8. Indeed, the "suboptimality factor" in Proposition 4.1 depends (logarithmically) solely on the sizes of the participating spectratopes, while in Proposition 3.8 this factor is affected also by the actual minimax risk and deteriorates, albeit only logarithmically, as the minimax risk goes to 0.

5. Proofs

5.1. Technical lemma. In the sequel, we repeatedly use the following technical fact:

Lemma 5.1. *Given basic spectratope (2.1), a positive definite $n \times n$ matrix Q and setting $\Lambda_k = \mathcal{R}_k[Q]$, we get a collection of positive semidefinite matrices such that $\sum_k \mathcal{R}_k^*[\Lambda_k]$ is positive definite. As a corollary, whenever M_k, $k \le K$, are positive definite matrices, the matrix $\sum_k \mathcal{R}_k^*[M_k]$ is positive definite. In addition, the set*

$$\mathcal{W} = \left\{Q \in \mathbf{S}^n : Q \succeq 0, \exists t \in \mathcal{T} : \mathcal{R}_k[Q] \preceq t_k I_{d_k}, k \le K\right\}$$

is nonempty convex compact set containing a neighbourhood of the origin.

Proof. Let us prove the first claim, Assuming the opposite, we can find a nonzero vector y such that $\sum_k y^T \mathcal{R}_k^*[\Lambda_k] y \le 0$, whence

$$0 \ge \sum_k y^T \mathcal{R}_k^*[\Lambda_k] y = \sum_k \mathrm{Tr}\left(\mathcal{R}_k^*[\Lambda_k][yy^T]\right) = \sum_k \mathrm{Tr}\left(\Lambda_k \mathcal{R}_k[yy^T]\right)$$

(we have used (2.8) and (2.4)). Since $\Lambda_k = \mathcal{R}_k[Q] \succeq 0$ due to $Q \succeq 0$, see (2.5), it follows that $\mathrm{Tr}(\Lambda_k \mathcal{R}_k[yy^T]) = 0$ for all k. Now, the linear mapping $\mathcal{R}_k[\cdot]$ is $\succeq$-monotone, and Q is positive definite, implying that $Q \succeq r_k yy^T$ for some $r_k > 0$, whence $\Lambda_k \succeq r_k \mathcal{R}_k[yy^T]$. Therefore, $\mathrm{Tr}(\Lambda_k \mathcal{R}_k[yy^T]) = 0$ implies that $\mathrm{Tr}(\mathcal{R}_k^2[yy^T]) = 0$, that is, $\mathcal{R}_k[yy^T] = R_k^2[y] = 0$. Since $R_k[\cdot]$ takes values in $\mathbf{S}^{d_k}$, we get $R_k[y] = 0$ for all k, which is impossible due to $y \ne 0$ and property (S_3), see Section 2.1.

The second claim is an immediate consequence of the first one. Indeed, when M_k are positive definite, we can find $\gamma > 0$ such that $\Lambda_k \preceq \gamma M_k$ for all $k \le K$; invoking (2.9), we conclude that $\mathcal{R}_k^*[\Lambda_k] \preceq \gamma \mathcal{R}_k^*[M_k]$, whence $\sum_k \mathcal{R}_k^*[M_k]$ is positive definite along with $\sum_k \mathcal{R}_k^*[\Lambda_k]$.

Finally, the only unevident component in the last claim of the lemma is that $\mathcal{W}$ is bounded. To see that it is the case, let us fix a collection $\{M_k\}$ of positive definite matrices $M_k \in \mathbf{S}^{d_k}$, and let us set $M = \sum_k \mathcal{R}_k^*[M_k]$, so that $M \succ 0$ by already proved part of the lemma. For $Q \in \mathcal{W}$, we have $\mathcal{R}_k[Q] \preceq t_k I_{d_k}$, $k \leq K$, for properly selected $t \in \mathcal{T}$, so that

$$\mathrm{Tr}(QM) = \sum_k \mathrm{Tr}\left(Q\mathcal{R}_k^*[M_k]\right) = \sum_k \mathrm{Tr}\left(\mathcal{R}_k[Q]M_k\right) \leq \sum_k t_k\,\mathrm{Tr}(M_k)$$

(we have used (2.8)), and the concluding quantity does not exceed properly selected $C < \infty$ (since $\mathcal{T}$ is compact). Thus, $\mathcal{W} \subset \{Q : Q \succeq 0, \mathrm{Tr}(QM) \leq C\}$, whence $\mathcal{W}$ is bounded due to $M \succ 0$. □

5.2. Proof of Proposition 2.2.

5.2.1. Preliminaries: matrix concentration. We are about to use the following deep matrix concentration result, see [29, Theorem 4.6.1]:

Theorem 5.2. *Let $Q_i \in \mathbf{S}^n$, $1 \leq i \leq I$, and let ξ_i, $i = 1, \dots, I$, be independent Rademacher (± 1 with probabilities $1/2$) or $\mathcal{N}(0,1)$ random variables. Then for all $s \geq 0$ one has*

$$\mathrm{Prob}\left\{\left\|\sum_{i=1}^{I} \xi_i Q_i\right\| > s\right\} \leq 2n \exp\left\{-\frac{s^2}{2v_Q}\right\},$$

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

We also need the following immediate consequence of the theorem:

Lemma 5.3. *Given spectratope (2.1), let $Q \in \mathbf{S}^n_+$ be such that*

$$\mathcal{R}_k[Q] \preceq \rho t_k I_{d_k}, \quad 1 \leq k \leq K, \tag{5.1}$$

for some $t \in \mathcal{T}$ and some $\rho \in (0,1]$. Then

$$\mathrm{Prob}_{\xi \sim \mathcal{N}(0,Q)}\{\xi \notin \mathcal{X}\} \leq \min\left[2D\mathrm{e}^{-1/2\rho}, 1\right], \quad D := \sum_{k=1}^{K} d_k.$$

Proof. When setting $\xi = Q^{1/2}\eta$, $\eta \sim \mathcal{N}(0, I_n)$, we have

$$R_k[\xi] = R_k\left[Q^{1/2}\eta\right] =: \sum_{i=1}^{n} \eta_i \bar{R}^{ki} = \bar{R}_k[\eta]$$

with

$$\sum_i \left[\bar{R}^{ki}\right]^2 = \mathbf{E}_{\eta \sim \mathcal{N}(0,I_n)}\left\{\bar{R}_k^2[\eta]\right\} = \mathbf{E}_{\xi \sim \mathcal{N}(0,Q)}\left\{R_k^2[\xi]\right\} = \mathcal{R}_k[Q] \preceq \rho t_k I_{d_k}$$

due to (2.6). Hence, by Theorem 5.2 as applied with $Q_i = \bar{R}^{ki}$, $s = \sqrt{t_k}$, we get $v_Q \le \rho t_k$, and therefore

$$\mathrm{Prob}_{\xi\sim\mathcal{N}(0,Q)}\left\{\|R_k[\xi]\|^2 \ge t_k\right\} = \mathrm{Prob}_{\eta\sim\mathcal{N}(0,I_n)}\left\{\|\bar{R}_k[\eta]\|^2 \ge t_k\right\} \le 2d_k \mathrm{e}^{-1/2\rho}.$$

We conclude that

$$\mathrm{Prob}_{\xi\sim\mathcal{N}(0,Q)}\{\xi \notin \mathcal{X}\} \le \mathrm{Prob}_{\xi\sim\mathcal{N}(0,Q)}\left\{\exists k : \|R_k[\xi]\|^2 > t_k\right\} \le 2D\mathrm{e}^{-1/2\rho}. \quad \square$$

5.2.2. Proving Proposition 2.2.

1^o. Under the premise of Proposition 2.2, let us set $\bar{C} = P^T C P$, and consider the conic problem

$$\mathrm{Opt}_\# = \max_{Q,t}\left\{\mathrm{Tr}(\bar{C}Q) : Q \succeq 0,\ \mathcal{R}_k[Q] \preceq t_k I_{d_k},\ \forall k \le K,\ \underbrace{[t;1] \in \mathbf{K}[\mathcal{T}]}_{\Leftrightarrow t\in\mathcal{T}}\right\}. \tag{5.2}$$

Since $\mathcal{T}$ contains positive vectors, this problem is strictly feasible. Besides this, the feasible set of the problem is bounded by Lemma 5.1 and since $\mathcal{T}$ is compact. Thus, problem (5.2) is strictly feasible with bounded feasible set and thus is solvable along with its conic dual, both problems sharing a common optimal value (Conic Duality Theorem, see Appendix C):

$$\begin{aligned}
\mathrm{Opt}_\# = &\min_{\substack{\Lambda=\{\Lambda_k\}_{k\le K},\\ [g;s],\,L}}\Big\{s : \mathrm{Tr}\Big(\Big[\sum_k \mathcal{R}_k^*[\Lambda_k] - L\Big]Q\Big) - \sum_k\big[\mathrm{Tr}(\Lambda_k) + g_k\big]t_k \\
&\qquad = \mathrm{Tr}(\bar{C}Q),\ \forall(Q,t),\ \Lambda_k \succeq 0\ \forall k,\ L \succeq 0,\ s \ge \phi_{\mathcal{T}}(-g)\Big\} \\
&\qquad \big[\text{recall that the cone dual to } \mathbf{K}[\mathcal{T}] \text{ is } \{[g;s] : s \ge \phi_{\mathcal{T}}(-g)\}\big] \\
= &\min_{\Lambda,\,[g;s],\,L}\Big\{s : \sum_k \mathcal{R}_k^*[\Lambda_k] - L = \bar{C},\ g = -\lambda[\Lambda], \\
&\qquad \Lambda_k \succeq 0,\ \forall k,\ L \succeq 0,\ s \ge \phi_{\mathcal{T}}(-g)\Big\} \\
= &\min_{\Lambda}\Big\{\phi_{\mathcal{T}}(\lambda[\Lambda]) : \sum_k \mathcal{R}_k^*[\Lambda_k] \succeq \bar{C},\ \Lambda_k \succeq 0,\ \forall k\Big\} = \mathrm{Opt}_*\,.
\end{aligned}$$

We see that (2.10) is solvable along with conic dual to problem (5.2), and

$$\mathrm{Opt}_\# = \mathrm{Opt}_*\,.$$

2^o. Problem (5.2), as we already know, is solvable; let Q_*, t^* be an optimal solution to the problem. Next, let us set $R_* = Q_*^{1/2}$, $\widehat{C} = R_*\bar{C}R_*$, and let $\widehat{C} = UDU^T$ be the eigenvalue decomposition of $\widehat{C}$, so that the matrix $D = U^T R_*\bar{C}R_*U$ is diagonal, and the trace of this matrix is $\mathrm{Tr}(R_*\bar{C}R_*) = \mathrm{Tr}(\bar{C}Q_*) = \mathrm{Opt}_\# = \mathrm{Opt}_*$. Now let $V = R_*U$, and let $\xi = V\eta$, where $\eta \sim \mathcal{R}$ (i.e. η is n-dimensional random (Rademacher) vector with independent entries taking values ± 1 with probabilities $1/2$). We have

$$\xi^T \bar{C} \xi = \eta^T \big[V^T \bar{C} V\big]\eta = \eta^T \big[U^T R_* \bar{C} R_* U\big]\eta = \eta^T D \eta \equiv \mathrm{Tr}(D) = \mathrm{Opt}_*, \quad (5.3)$$

(recall that D is diagonal) and

$$\mathbf{E}_\xi\{\xi\xi^T\} = \mathbf{E}_\eta\{V\eta\eta^T V^T\} = VV^T = R_* UU^T R_* = R_*^2 = Q_*.$$

From the latter relation,

$$\mathbf{E}_\xi\{R_k^2[\xi]\} = \mathbf{E}_\xi\{\mathcal{R}_k[\xi\xi^T]\} = \mathcal{R}_k\big[\mathbf{E}_\xi\{\xi\xi^T\}\big] = \mathcal{R}_k[Q_*] \preceq t_k^* I_{d_k}, \quad 1 \le k \le K. \quad (5.4)$$

On the other hand, with properly selected symmetric matrices $\bar{R}^{ki}$ we have

$$\bar{R}_k[y] := R_k[Vy] = \sum_i \bar{R}^{ki} y_i$$

identically in $y \in \mathbf{R}^n$, whence

$$\mathbf{E}_\xi\{R_k^2[\xi]\} = \mathbf{E}_\eta\{R_k^2[V\eta]\} = \mathbf{E}_\eta\Big\{\Big[\sum_i \eta_i \bar{R}^{ki}\Big]^2\Big\}$$
$$= \sum_{i,j} \mathbf{E}_\eta\{\eta_i\eta_j\}\bar{R}^{ki}\bar{R}^{kj} = \sum_i [\bar{R}^{ki}]^2.$$

This combines with (5.4) to imply that

$$\sum_i [\bar{R}^{ki}]^2 \preceq t_k^* I_{d_k}, \quad 1 \le k \le K. \quad (5.5)$$

3^o. Let us fix $k \le K$. Applying Theorem 5.2 with $Q_i = \bar{R}^{ki}$ and $s = \sqrt{t_k^*/\rho}$, we derive from (5.5) that

$$\mathrm{Prob}_{\eta\sim\mathcal{R}}\left\{\|\bar{R}_k[\eta]\|^2 > t_k^*/\rho\right\} \le 2d_k \mathrm{e}^{-1/2\rho},$$

and recalling the relation between ξ and η, we arrive at

$$\mathrm{Prob}\left\{\xi : \|R_k[\xi]\|^2 > t_k^*/\rho\right\} \le 2d_k \mathrm{e}^{-1/2\rho}, \quad \forall \rho \in (0,1]. \quad (5.6)$$

Now let us set $\bar{\rho} = 1/2\max[\ln(2D), 1]$, and let $\rho \in (0, \bar{\rho})$. For this ρ, the sum over $k \le K$ of the right hand sides in inequalities (5.6) is < 1, implying that there exists a realization $\bar{\xi}$ of ξ such that

$$\|R_k[\bar{\xi}]\|^2 \le t_k^*/\rho, \quad \forall k,$$

or, equivalently,

$$\bar{x} := \rho^{1/2} P \bar{\xi} \in \mathcal{X},$$

and

$$\mathrm{Opt} \ge \bar{x}^T C \bar{x} = \rho \xi^T \bar{C} \xi = \rho\,\mathrm{Opt}_*$$

(the concluding equality is due to (5.3)). The resulting inequality holds true for every $\rho \in (0, \bar{\rho})$, and we arrive at the right inequality in (2.11). □

5.3. Derivation of relation (3.26) of Section 3.4. Let us $\succeq$-upper-bound the covariance mapping $\mathcal{C}[v]$ of $\zeta = \eta\eta^T - \mathbf{E}\{\eta\eta^T\}$, $\eta \sim \mathcal{N}(0, A\vartheta[v]A^T)$. Observe that $\mathcal{C}[v]$ is a symmetric linear mapping of $\mathbf{S}^m$ into itself given by

$$\langle h, \mathcal{C}[v]h\rangle = \mathbf{E}\big\{\langle h, \zeta\rangle^2\big\} = \mathbf{E}\big\{\langle h, \eta\eta^T\rangle^2\big\} - \langle h, \mathbf{E}\{\eta\eta^T\}\rangle^2, \quad h \in \mathbf{S}^m.$$

Given $v \in \mathcal{V}$, setting $\theta = \vartheta[v]$, so that $0 \preceq \theta \preceq \vartheta_*$, and denoting $\mathcal{H}(h) = \theta^{1/2}A^T hA\theta^{1/2}$, we obtain

$$\begin{aligned}\langle h, \mathcal{C}[v]h\rangle &= \mathbf{E}_{\xi\sim\mathcal{N}(0,\theta)}\big\{\mathrm{Tr}^2\big(hA\xi\xi^T A^T\big)\big\} - \mathrm{Tr}^2\big(h\mathbf{E}_{\xi\sim\mathcal{N}(0,\theta)}\{A\xi\xi^T A^T\}\big)\\ &= \mathbf{E}_{\chi\sim\mathcal{N}(0,I_n)}\big\{\mathrm{Tr}^2\big(hA\theta^{1/2}\chi\chi^T\theta^{1/2}A^T\big)\big\} - \mathrm{Tr}^2\big(hA\theta A^T\big)\\ &= \mathbf{E}_{\chi\sim\mathcal{N}(0,I_n)}\big\{\big(\chi^T\mathcal{H}(h)\chi\big)^2\big\} - \mathrm{Tr}^2\big(\mathcal{H}(h)\big).\end{aligned}$$

We have $\mathcal{H}(h) = U\,\mathrm{Diag}\{\lambda\}U^T$ with orthogonal U, so that for $\overline{\chi} = U^T\chi$ we get

$$\begin{aligned}&\mathbf{E}_{\chi\sim\mathcal{N}(0,I_n)}\big\{\big(\chi^T\mathcal{H}(h)\chi\big)^2\big\} - \mathrm{Tr}^2\big(\mathcal{H}(h)\big)\\ &\qquad= \mathbf{E}_{\overline{\chi}\sim\mathcal{N}(0,I_n)}\big\{\big(\overline{\chi}^T\,\mathrm{Diag}\{\lambda\}\overline{\chi}\big)^2\big\} - \Big(\sum_i \lambda_i\Big)^2\\ &\qquad= \mathbf{E}_{\overline{\chi}\sim\mathcal{N}(0,I_n)}\Big\{\Big(\sum_i \lambda_i\overline{\chi}_i^2\Big)^2\Big\} - \Big(\sum_i \lambda_i\Big)^2\\ &\qquad= \sum_{i\neq j}\lambda_i\lambda_j + 3\sum_i \lambda_i^2 - \Big(\sum_i\lambda_i\Big)^2 = 2\sum_i\lambda_i^2 = 2\,\mathrm{Tr}\big([\mathcal{H}(h)]^2\big).\end{aligned}$$

Thus,

$$\begin{aligned}\langle h, \mathcal{C}[v]h\rangle &= 2\,\mathrm{Tr}\big([\mathcal{H}(h)]^2\big) = 2\,\mathrm{Tr}\big(\theta^{1/2}A^ThA\theta A^ThA\theta^{1/2}\big)\\ &\leq 2\,\mathrm{Tr}\big(\theta^{1/2}A^ThA\vartheta_*A^ThA\theta^{1/2}\big) &&\big[\text{since } 0 \preceq \theta \preceq \vartheta_*\big]\\ &= 2\,\mathrm{Tr}\big(\vartheta_*^{1/2}A^ThA\theta A^ThA\vartheta_*^{1/2}\big)\\ &\leq 2\,\mathrm{Tr}\big(\vartheta_*^{1/2}A^ThA\vartheta_*A^ThA\vartheta_*^{1/2}\big)\\ &= 2\,\mathrm{Tr}\big(\vartheta_*A^ThA\vartheta_*A^ThA\big),\end{aligned}$$

which implies (3.26). □

5.4. Proof of Lemma 3.2. In the case of (3.7), we have

$$
\begin{aligned}
\|Y^T\xi\| &= \max_{z\in\mathcal{B}_*} z^TY^T\xi = \max_{y\in\mathcal{Y}} y^TM^TY^T\xi \\
&\underbrace{\le}_{(3.7)} \max_{y\in\mathcal{Y}}\left[\xi^T\Theta\xi + \sum_\ell y^T\mathcal{S}_\ell^*[\Upsilon_\ell]y\right] \\
&= \max_{y\in\mathcal{Y}}\left[\xi^T\Theta\xi + \sum_\ell \mathrm{Tr}\left(\mathcal{S}_\ell^*[\Upsilon_\ell]yy^T\right)\right] \\
&\underbrace{=}_{(2.4),(2.8)} \max_{y\in\mathcal{Y}}\left[\xi^T\Theta\xi + \sum_\ell \mathrm{Tr}\left(\Upsilon_\ell S_\ell^2[y]\right)\right] \\
&\underbrace{=}_{(3.3)} \xi^T\Theta\xi + \max_{y,r}\left\{\sum_\ell \mathrm{Tr}\left(\Upsilon_\ell S_\ell^2[y]\right) : S_\ell^2[y]\preceq r_\ell I_{f_\ell},\ \ell\le L,\ r\in\mathcal{R}\right\} \\
&\underbrace{\le}_{\Upsilon_\ell\succeq 0} \xi^T\Theta\xi + \max_{r\in\mathcal{R}}\sum_\ell \mathrm{Tr}(\Upsilon_\ell)r_\ell \le \xi^T\Theta\xi + \phi_{\mathcal{R}}(\lambda[\Upsilon]).
\end{aligned}
$$

Taking expectation of both sides of the resulting inequality w.r.t. distribution P of ξ and taking into account that $\mathrm{Tr}(\mathrm{Cov}[P]\Theta)\le\mathrm{Tr}(Q\Theta)$ due to $\Theta\succeq 0$ (by (3.7)) and $\mathrm{Cov}[P]\preceq Q$, we get (3.8). □

5.5. Proof of Proposition 3.7. In the case of (3.15), problem (3.13) reads

$$
\begin{aligned}
\mathrm{Opt} = \min_{\substack{H=[\widetilde{H}_1;\dots;\widetilde{H}_T],\,\Lambda,\,\Upsilon,\,\Upsilon',\\ \Theta=[\theta^{t\tau}]_{1\le t,\tau\le T}}} \Bigg\{ & \phi_{\mathcal{T}}(\lambda[\Lambda]) + \phi_{\mathcal{R}}(\lambda[\Upsilon]) + \phi_{\mathcal{R}}(\lambda[\Upsilon']) + \sum_{t=1}^T \overline{\Gamma}(\theta^{tt}) : \\
& \Lambda = \{\Lambda_k\succeq 0,\ k\le K\},\ \Upsilon=\{\Upsilon_\ell\succeq 0,\ \ell\le L\},\ \Upsilon'=\{\Upsilon'_\ell\succeq 0,\ \ell\le L\}, \\
& \left[\begin{array}{c|c} \sum_k \mathcal{R}_k^*[\Lambda_k] & \frac12[B^T-\overline{A}^T\sum_{t=1}^T\widetilde{H}_t]M \\ \hline \frac12 M^T[B-[\sum_{t=1}^T\widetilde{H}_t^T]\overline{A}] & \sum_\ell \mathcal{S}_\ell^*[\Upsilon_\ell]\end{array}\right]\succeq 0, \\
& \left[\begin{array}{ccc|c} \theta^{1,1} & \cdots & \theta^{1,T} & \frac12\widetilde{H}_1M \\ \vdots & \ddots & \vdots & \vdots \\ \theta^{T,1} & \cdots & \theta^{T,T} & \frac12\widetilde{H}_TM \\ \hline \frac12 M^T\widetilde{H}_1^T & \cdots & \frac12 M^T\widetilde{H}_T^T & \sum_\ell\mathcal{S}_\ell^*[\Upsilon'_\ell]\end{array}\right]\succeq 0\Bigg\},
\end{aligned}
$$

$$
\overline{\Gamma}(\theta) = \max_{\bar{Q}\in\bar{\mathcal{Q}}}\mathrm{Tr}(\bar{Q}\theta), \tag{5.7}
$$

where $\widetilde{H}_t$ are $\overline{m}\times\nu$ matrices, and $\theta^{t\tau}=[\theta^{\tau t}]^T$, $1\le t,\tau\le T$, form a partition of $\Theta\in\mathbf{S}^{\overline{m}T}$ into $\overline{m}\times\overline{m}$ blocks. Problem (5.7) clearly admits a group of symmetries: a permutation σ of $\{1,\dots,T\}$ induces the transformation on the space of decision variables which keeps $\Lambda, \Upsilon, \Upsilon'$ intact and maps $\widetilde{H}_t$ into $\widetilde{H}_{\sigma(t)}$, and $\theta^{t\tau}$ into $\theta^{\sigma(t)\sigma(\tau)}$; this transformation preserves the feasible set and keeps intact the value of the objective. Since the problem is convex and solvable, it admits a "symmetric" optimal solution — one with $\widetilde{H}_t = \widetilde{H}$, and $\theta^{tt} = \theta$, $1 \leq t \leq T$. From the concluding semidefinite constraint in (5.7) it follows that

$$\left[\begin{array}{c|c} \theta & \frac{1}{2}\widetilde{H}M \\ \hline \frac{1}{2}\widetilde{H}^T M^T & \sum_\ell \mathcal{S}_\ell^*[\Upsilon'_\ell] \end{array}\right] \succeq 0,$$

where Υ' stems from the symmetric solution in question. We conclude that $\mathrm{Opt} \geq \overline{\mathrm{Opt}}$, where

$$\begin{array}{l} \overline{\mathrm{Opt}} = \min\limits_{\widetilde{H}\in\mathbf{R}^{\overline{m}\times\nu},\Lambda,\Upsilon,\Upsilon',\theta} \Big\{ \phi_{\mathcal{T}}(\lambda[\Lambda]) + \phi_{\mathcal{R}}(\lambda[\Upsilon]) + \phi_{\mathcal{R}}(\lambda[\Upsilon']) + T\overline{\Gamma}(\theta) : \\ \quad \Lambda = \{\Lambda_k \succeq 0, k \leq K\},\ \Upsilon = \{\Upsilon_\ell \succeq 0, \ell \leq L\},\ \Upsilon' = \{\Upsilon'_\ell \succeq 0, \ell \leq L\}, \\ \quad \left[\begin{array}{c|c} \sum_k \mathcal{R}_k^*[\Lambda_k] & \frac{1}{2}[B^T - T\bar{A}^T\widetilde{H}]M \\ \hline \frac{1}{2}M^T[B - T\widetilde{H}^T\bar{A}] & \sum_\ell \mathcal{S}_\ell^*[\Upsilon_\ell] \end{array}\right] \succeq 0, \\ \quad \left[\begin{array}{c|c} \theta & \frac{1}{2}\widetilde{H}M \\ \hline \frac{1}{2}M^T\widetilde{H}^T & \sum_\ell \mathcal{S}_\ell^*[\Upsilon'_\ell] \end{array}\right] \succeq 0 \Big\}. \end{array} \tag{5.8}$$

It is immediately seen that a feasible solution $\widetilde{H} \in \mathbf{R}^{\overline{m}\times\nu}, \Lambda, \Upsilon, \Upsilon', \theta$ to the optimization problem in (5.8) gives rise to a feasible solution to (5.7) with the same value of the objective, specifically, the solution $H = [\widetilde{H}; \ldots; \widetilde{H}], \Lambda, \Upsilon, \Upsilon', \Theta = [\theta^{t\tau} = \theta]_{t,\tau\leq T}$. We conclude that $\mathrm{Opt} = \overline{\mathrm{Opt}}$, and an optimal solution $\widetilde{H} \in \mathbf{R}^{\overline{m}\times\nu}, \Lambda, \Upsilon, \Upsilon', \theta$ to the optimization problem in (5.8) gives rise to a symmetric optimal solution to (5.7). The associated linear estimate is

$$\widetilde{H}^T \sum_{t=1}^T \omega_t,$$

and its risk is upper-bounded by $\mathrm{Opt} = \overline{\mathrm{Opt}}$. It remains to note that the optimization problem in (3.16) is obtained from the optimization problem in (5.8) by substituting $\widetilde{H} = T^{-1}\bar{H}$ and $\theta = T^{-2}\bar{\Theta}$. □

5.6. Proof of Lemma 3.10.

1^o. Let us verify (3.21). When $Q \succ 0$, passing from variables (Θ, Υ) in problem (3.20) to the variables $(G = Q^{1/2}\Theta Q^{1/2}, \Upsilon)$, the problem becomes exactly the optimization problem in (3.21), implying that $\mathrm{Opt}[Q] = \overline{\mathrm{Opt}}[Q]$ when $Q \succ 0$. As it is easily seen, both sides in this equality are continuous in $Q \succeq 0$, and (3.21) follows.

2^o. Let us set $\zeta = Q^{1/2}\eta$ with $\eta \sim \mathcal{N}(0, I_N)$ and $Z = Q^{1/2}Y$. All we need to complete the proof of Lemma 3.10 is to show that the quantity

$$[\overline{\mathrm{Opt}}[Q] =] \;\; \mathrm{Opt} := \min_{\Theta,\Upsilon=\{\Upsilon_\ell,\ell\le L\}} \left\{ \phi_{\mathcal{R}}(\lambda[\Upsilon]) + \mathrm{Tr}(\Theta) : \Upsilon_\ell \succeq 0, \left[\begin{array}{c|c} \Theta & \frac{1}{2}ZM \\ \hline \frac{1}{2}M^TZ^T & \sum_\ell \mathcal{S}_\ell^*[\Upsilon_\ell] \end{array}\right] \succeq 0 \right\} \tag{5.9}$$

satisfies

$$\begin{array}{l} \psi_I(Z) \le \mathrm{Opt} \le \dfrac{8\sqrt{\ln\left(4\sqrt{2}F/(\sqrt{2}-\mathrm{e}^{1/4})\right)}}{\sqrt{2}-\mathrm{e}^{1/4}} \psi_I(Z), \\ \psi_I(Z) = \mathbf{E}_{\eta\sim\mathcal{N}(0,I_N)}\{\|Z^T\eta\|\}. \end{array} \tag{5.10}$$

3^o. Let us represent Opt as the optimal value of a conic problem. Setting

$$\mathbf{K} = \mathbf{K}[\mathcal{R}] = \mathrm{cl}\left\{[r;s] : s > 0, r/s \in \mathcal{R}\right\},$$

we ensure that

$$\mathcal{R} = \left\{r : [r;1] \in \mathbf{K}\right\}, \quad \mathbf{K}_* = \left\{[g;s] : s \ge \phi_{\mathcal{R}}(-g)\right\},$$

where $\mathbf{K}_*$ is the cone dual to $\mathbf{K}$. Consequently, (5.9) reads

$$\mathrm{Opt} = \min_{\Theta,\Upsilon,\theta} \left\{ \theta + \mathrm{Tr}(\Theta) : \begin{array}{lr} \Upsilon_\ell \succeq 0,\ 1 \le \ell \le L & (a) \\ \left[\begin{array}{c|c} \Theta & \frac{1}{2}ZM \\ \hline \frac{1}{2}M^TZ^T & \sum_\ell \mathcal{S}_\ell^*[\Upsilon_\ell] \end{array}\right] \succeq 0 & (b) \\ \left[-\lambda[\Upsilon];\theta\right] \in \mathbf{K}_* & (c) \end{array} \right\}. \tag{P}$$

4^o. Now let us prove that there exists matrix $W \in \mathbf{S}^q_+$ and $r \in \mathcal{R}$ such that

$$\mathcal{S}_\ell[W] \preceq r_\ell I_{f_\ell}, \quad \ell \le L, \tag{5.11}$$

and

$$\mathrm{Opt} \le \sum_i \sigma_i\left(ZMW^{1/2}\right), \tag{5.12}$$

where $\sigma_1(\cdot) \ge \sigma_2(\cdot) \ge \cdots$ are singular values.

To get the announced result, let us pass from problem (P) to its conic dual. Applying Lemma 5.1 we conclude that (P) is strictly feasible; in addition, (P) clearly is bounded, so that the dual to (P) problem (D) is solvable with optimal value Opt. Let us build (D). Denoting by

$$\Lambda_\ell \succeq 0,\ \ell \le L, \quad \left[\begin{array}{c|c} G & -R \\ \hline -R^T & W \end{array}\right] \succeq 0, \quad [r;\tau] \in \mathbf{K}$$

the Lagrange multipliers for the respective constraints in (P), and aggregating these constraints, the multipliers being the aggregation weights, we arrive at the following aggregated constraint:

$$\mathrm{Tr}(\Theta G)+\mathrm{Tr}\left(W\sum_\ell \mathcal{S}_\ell^*[\Upsilon_\ell]\right)+\sum_\ell \mathrm{Tr}(\Lambda_\ell\Upsilon_\ell)-\sum_\ell r_\ell\,\mathrm{Tr}(\Upsilon_\ell)+\theta\tau \geq \mathrm{Tr}\left(ZMR^T\right).$$

To get the dual problem, we impose on the Lagrange multipliers, in addition to the initial conic constraints like $\Lambda_\ell \succeq 0$, $1 \leq \ell \leq L$, the restriction that the left hand side in the aggregated constraint, identically in Θ, Υ_ℓ and θ, is equal to the objective of (P), that is,

$$G = I,\quad \mathcal{S}_\ell[W] + \Lambda_\ell - r_\ell I_{f_\ell} = 0,\ 1 \leq \ell \leq L,\quad \tau = 1,$$

and maximize, under the resulting restrictions, the right-hand side of the aggregated constraint. After immediate simplifications, we arrive at

$$\mathrm{Opt} = \max_{W,R,r}\left\{\mathrm{Tr}(ZMR^T) : W \succeq R^TR,\ r \in \mathcal{R},\ \mathcal{S}_\ell[W] \preceq r_\ell I_{f_\ell},\ 1 \leq \ell \leq L\right\}.$$

Note that $r \in \mathcal{R}$ is equivalent to $[r; 1] \in \mathbf{K}$, and $W \succeq R^TR$ is the same as

$$\left[\begin{array}{c|c} I & -R \\ \hline -R^T & W \end{array}\right] \succeq 0.$$

Now, to say that $R^TR \preceq W$ is exactly the same as to say that $R = SW^{1/2}$ with the spectral norm $\|S\|_{\mathrm{Sh},\infty}$ of S not exceeding 1, so that

$$\mathrm{Opt} = \max_{W,S,r}\Big\{\underbrace{\mathrm{Tr}\left([ZM[SW^{1/2}]^T\right)}_{=\mathrm{Tr}([ZMW^{1/2}]S^T)} : W \succeq 0,\ \|S\|_{\mathrm{Sh},\infty} \leq 1,\ r \in \mathcal{R},\ \mathcal{S}_\ell[W] \preceq r_\ell I_{f_\ell},\ \ell \leq L\Big\}$$

and we can immediately eliminate the S-variable, using the well-known fact that for every $p \times q$ matrix J, it holds

$$\max_{S\in\mathbf{R}^{p\times q},\|S\|_{\mathrm{Sh},\infty}\leq 1} \mathrm{Tr}\left(JS^T\right) = \|J\|_{\mathrm{Sh},1},$$

where $\|J\|_{\mathrm{Sh},1}$ is the nuclear norm (the sum of singular values) of J. We arrive at

$$\mathrm{Opt} = \max_{W,r}\left\{\|ZMW^{1/2}\|_{\mathrm{Sh},1} : r \in \mathcal{R},\ W \succeq 0,\ \mathcal{S}_\ell[W] \preceq r_\ell I_{d_\ell},\ \ell \leq L\right\}.$$

The resulting problem clearly is solvable, and its optimal solution W ensures the target relations (5.11) and (5.12).

5^o. Given W satisfying (5.11) and (5.12), let $UJV = W^{1/2}M^TZ^T$ be the singular value decomposition of $W^{1/2}M^TZ^T$, so that U and V are, respectively, $q\times q$ and $N\times N$ orthogonal matrices, J is $q\times N$ matrix with diagonal $\sigma=[\sigma_1;\dots;\sigma_p]$, $p=\min[q,N]$, and zero off-diagonal entries; the diagonal entries σ_i, $1\leq i\leq p$ are the singular values of $W^{1/2}M^TZ^T$, or, which is the same, of $ZMW^{1/2}$. Therefore, we have

$$\sum_i \sigma_i \geq \mathrm{Opt}\,. \tag{5.13}$$

Now consider the following construction. Let $\eta\sim\mathcal{N}(0,I_N)$; we denote by υ the vector comprised of the first p entries in $V\eta$; note that $\upsilon\sim\mathcal{N}(0,I_p)$, since V is orthogonal. We then augment, if necessary, υ by $q-p$ independent of each other and of η $\mathcal{N}(0,1)$ random variables to obtain a q-dimensional normal vector $\upsilon'\sim\mathcal{N}(0,I_q)$, and set $\chi=U\upsilon'$; because U is orthogonal we also have $\chi\sim\mathcal{N}(0,I_q)$. Observe that

$$\chi^TW^{1/2}M^TZ^T\eta=\chi^TUJV\eta=[\upsilon']^TJ\upsilon=\sum_{i=1}^p\sigma_i\upsilon_i^2. \tag{5.14}$$

To continue we need the following simple observations.

(1) *One has*

$$\alpha:=\mathrm{Prob}\left\{\sum_{i=1}^p\sigma_i\upsilon_i^2<\frac{1}{2}\sum_{i=1}^p\sigma_i\right\}\leq\frac{\mathrm{e}^{1/4}}{\sqrt{2}}\ [<1]. \tag{5.15}$$

The claim is evident when $\sigma:=\sum_i\sigma_i=0$. Now let $\sigma>0$, and let us apply the Cramer bounding scheme. Namely, given $\gamma>0$, consider the random variable

$$\omega=\exp\left\{\tfrac{1}{2}\gamma\sum_i\sigma_i-\gamma\sum_i\sigma_i\upsilon_i^2\right\}.$$

Note that $\omega>0$ a.s., and is >1 when $\sum_{i=1}^p\sigma_i\upsilon_i^2<\frac{1}{2}\sum_{i=1}^p\sigma_i$, so that $\alpha\leq\mathbf{E}\{\omega\}$, or, equivalently, thanks to $\upsilon\sim\mathcal{N}(0,I_p)$,

$$\begin{aligned}\ln(\alpha)\leq\ln\left(\mathbf{E}\{\omega\}\right)&=\frac{1}{2}\gamma\sum_i\sigma_i+\sum_i\ln\left(\mathbf{E}\left\{\exp\{-\gamma\sigma_i\upsilon_i^2\}\right\}\right)\\&\leq\frac{1}{2}\Big[\gamma\sigma-\sum_i\ln(1+2\gamma\sigma_i)\Big].\end{aligned}$$

Function $-\sum_i\ln(1+2\gamma\sigma_i)$ is convex in $[\sigma_1;\dots;\sigma_p]\geq0$, therefore, its maximum over the simplex $\{\sigma_i\geq0,\ i\leq p,\ \sum_i\sigma_i=\sigma\}$ is attained at a vertex, and we get

$$\ln(\alpha)\leq\tfrac{1}{2}\big[\gamma\sigma-\ln(1+2\gamma\sigma)\big].$$

Minimizing the right hand side in $\gamma>0$, we arrive at (5.15).

(2) *Whenever* $\varkappa \geq 1$, *one has*

$$\mathrm{Prob}\left\{\|MW^{1/2}\chi\|_* > \varkappa\right\} \leq 2F\exp\{-\varkappa^2/2\}, \tag{5.16}$$

with F *given by* (3.19).

Indeed, setting $\rho = 1/\varkappa^2 \leq 1$ and $\omega = \sqrt{\rho}W^{1/2}\chi$, we get $\omega \sim \mathcal{N}(0, \rho W)$. Let us apply Lemma 5.3 to $Q = \rho W$, $\mathcal{R}$ in the role of $\mathcal{T}$, L in the role of K, and $\mathcal{S}_\ell[\cdot]$ in the role of $\mathcal{R}_k[\cdot]$. Denoting

$$\mathcal{Y} := \left\{y : \exists r \in \mathcal{R} : S_\ell^2[y] \preceq r_\ell I_{f_\ell},\ \ell \leq L\right\},$$

we have $\mathcal{S}_\ell[Q] = \rho\mathcal{S}_\ell[W] \preceq \rho r_\ell I_{f_\ell}$, $\ell \leq L$, with $r \in \mathcal{R}$ (see (5.11)), so we are under the premise of Lemma 5.3. Applying the lemma, we conclude that

$$\mathrm{Prob}\left\{\chi : \varkappa^{-1}W^{1/2}\chi \notin \mathcal{Y}\right\} \leq 2F\exp\{-1/(2\rho)\} = 2F\exp\{-\varkappa^2/2\}.$$

Recalling that $\mathcal{B}_* = M\mathcal{Y}$, we see that $\mathrm{Prob}\{\chi : \varkappa^{-1}MW^{1/2}\chi \notin \mathcal{B}_*\}$ is indeed upper-bounded by the right hand side of (5.16), and (5.16) follows.

(3) *For* $\varkappa \geq 1$, *let*

$$E_\varkappa = \left\{(\chi, \eta) : \|MW^{1/2}\chi\|_* \leq \varkappa,\ \sum_i \sigma_i \upsilon_i^2 \geq \frac{1}{2}\sum_i \sigma_i\right\}.$$

Then one has

$$\mathrm{Prob}\{E_\varkappa\} \geq \beta(\varkappa) := 1 - \frac{\mathrm{e}^{1/4}}{\sqrt{2}} - 2F\exp\{-\varkappa^2/2\}. \tag{5.17}$$

Indeed, relation (5.17) follows from (5.15) and (5.16) due to the union bound.

When $(\chi, \eta) \in E_\varkappa$, we have

$$\begin{aligned}\varkappa\|Z^T\eta\| &\geq \|MW^{1/2}\chi\|_*\|Z^T\eta\|\\ &\geq \chi^T W^{1/2}M^T Z^T\eta\\ &= \sum_i \sigma_i\upsilon_i^2 \geq \frac{1}{2}\sum_i \sigma_i \geq \tfrac{1}{2}\,\mathrm{Opt},\end{aligned}$$

(we have used (5.14) and (5.13)), so that whenever $(\chi, \eta) \in E_\varkappa$ one has $\|Z^T\eta\| \geq \frac{1}{2\varkappa}\,\mathrm{Opt}$. Hence, finally,

$$\begin{aligned}2\mathbf{E}_{\eta\sim\mathcal{N}(0,I_N)}\left\{\|Z^T\eta\|\right\} &\geq \mathrm{Prob}\left\{(\chi, \eta) \in E_\varkappa\right\}\varkappa^{-1}\,\mathrm{Opt}\\ &\geq \left[1 - \frac{\mathrm{e}^{1/4}}{\sqrt{2}} - 2F\exp\{-\varkappa^2/2\}\right]\varkappa^{-1}\,\mathrm{Opt},\end{aligned}$$

and we arrive at (5.10) when specifying $\varkappa$ as

$$\varkappa = \sqrt{2\ln\left(\frac{4\sqrt{2}F}{\sqrt{2} - \mathrm{e}^{1/4}}\right)}. \qquad \square$$

5.7. Proof of Proposition 3.8.

1^o. Let

$$\begin{aligned}\Phi\big(H,\Lambda,\Upsilon,\Upsilon',\Theta;Q\big)\\ &= \phi_{\mathcal{T}}\big(\lambda[\Lambda]\big)+\phi_{\mathcal{R}}\big(\lambda[\Upsilon]\big)+\phi_{\mathcal{R}}\big(\lambda[\Upsilon']\big)+\mathrm{Tr}(Q\Theta)\colon \mathcal{M}\times\mathcal{Q}\to\mathbf{R},\end{aligned}$$

where

$$\begin{aligned}\mathcal{M}=\Big\{&(H,\Lambda,\Upsilon,\Upsilon',\Theta):\Lambda=\{\Lambda_k\succeq 0,\,k\le K\},\\ &\Upsilon=\{\Upsilon_\ell\succeq 0,\,\ell\le L\},\ \Upsilon'=\{\Upsilon'_\ell\succeq 0,\,\ell\le L\},\\ &\left[\begin{array}{c|c}\sum_k \mathcal{R}_k^*[\Lambda_k] & \frac12\big[B^T-A^TH\big]M\\ \hline \frac12 M^T\big[B-H^TA\big] & \sum_\ell \mathcal{S}_\ell^*[\Upsilon_\ell]\end{array}\right]\succeq 0,\\ &\left[\begin{array}{c|c}\Theta & \frac12 HM\\ \hline \frac12 M^TH^T & \sum_\ell \mathcal{S}_\ell^*[\Upsilon'_\ell]\end{array}\right]\succeq 0\Big\}.\end{aligned}$$

Looking at (3.13), we conclude immediately that the optimal value Opt in (3.13) is nothing but

$$\mathrm{Opt}=\min_{(H,\Lambda,\Upsilon,\Upsilon',\Theta)\in\mathcal{M}}\Big[\overline{\Phi}(H,\Lambda,\Upsilon,\Upsilon',\Theta):=\max_{Q\in\mathcal{Q}}\Phi(H,\Lambda,\Upsilon,\Upsilon',\Theta;Q)\Big]. \tag{5.18}$$

Note that the sets $\mathcal{M}$ and $\mathcal{Q}$ are closed and convex, $\mathcal{Q}$ is compact, and Φ is a continuous convex-concave function on $\mathcal{M}\times\mathcal{Q}$. In view of these observations, the fact that $\mathcal{Q}\subset\mathrm{int}\,\mathbf{S}^m_+$ combines with the Sion–Kakutani theorem to imply that Φ possesses saddle point $(H_*,\Lambda_*,\Upsilon_*,\Upsilon'_*,\Theta_*;Q_*)$ (min in $(H,\Lambda,\Upsilon,\Upsilon',\Theta)$, max in Q) on $\mathcal{M}\times\mathcal{Q}$, whence Opt is the saddle point value of Φ by (5.18). We conclude that for properly selected $Q_*\in\mathcal{Q}$ it holds

$$\begin{aligned}\mathrm{Opt}&=\min_{(H,\Lambda,\Upsilon,\Upsilon',\Theta)\in\mathcal{M}}\Phi(H,\Lambda,\Upsilon,\Upsilon',\Theta;Q_*)\\ &=\min_{H,\Lambda,\Upsilon,\Upsilon',\Theta}\Big\{\phi_{\mathcal{T}}\big(\lambda[\Lambda]\big)+\phi_{\mathcal{R}}\big(\lambda[\Upsilon]\big)+\phi_{\mathcal{R}}\big(\lambda[\Upsilon']\big)+\mathrm{Tr}(Q_*\Theta):\\ &\qquad\Lambda=\{\Lambda_k\succeq 0,\,k\le K\},\ \Upsilon=\{\Upsilon_\ell\succeq 0,\,\ell\le L\},\ \Upsilon'=\{\Upsilon'_\ell\succeq 0,\,\ell\le L\},\\ &\qquad\left[\begin{array}{c|c}\sum_k \mathcal{R}_k^*[\Lambda_k] & \frac12\big[B^T-A^TH\big]M\\ \hline \frac12 M^T\big[B-H^TA\big] & \sum_\ell \mathcal{S}_\ell^*[\Upsilon_\ell]\end{array}\right]\succeq 0,\\ &\qquad\left[\begin{array}{c|c}\Theta & \frac12 HM\\ \hline \frac12 M^TH^T & \sum_\ell \mathcal{S}_\ell^*[\Upsilon'_\ell]\end{array}\right]\succeq 0\Big\}\end{aligned}$$

$$
\begin{aligned}
&= \min_{H,\Lambda,\Upsilon,\Upsilon',G}\Big\{\phi_{\mathcal{T}}\big(\lambda[\Lambda]\big)+\phi_{\mathcal{R}}\big(\lambda[\Upsilon]\big)+\phi_{\mathcal{R}}\big(\lambda[\Upsilon']\big)+\mathrm{Tr}(G):\\
&\qquad \Lambda=\{\Lambda_k\succeq 0,\,k\le K\},\ \Upsilon=\{\Upsilon_\ell\succeq 0,\,\ell\le L\},\ \Upsilon'=\{\Upsilon'_\ell\succeq 0,\,\ell\le L\},\\
&\qquad \left[\begin{array}{c|c}\sum_k\mathcal{R}_k^*[\Lambda_k] & \frac{1}{2}[B^T-A^TH]M\\ \hline \frac{1}{2}M^T[B-H^TA] & \sum_\ell \mathcal{S}_\ell^*[\Upsilon_\ell]\end{array}\right]\succeq 0,\\
&\qquad \left[\begin{array}{c|c}G & \frac{1}{2}Q_*^{1/2}HM\\ \hline \frac{1}{2}M^TH^TQ_*^{1/2} & \sum_\ell \mathcal{S}_\ell^*[\Upsilon'_\ell]\end{array}\right]\succeq 0\Big\}\\
&= \min_{H,\Lambda,\Upsilon}\Big\{\phi_{\mathcal{T}}\big(\lambda[\Lambda]\big)+\phi_{\mathcal{R}}\big(\lambda[\Upsilon]\big)+\overline{\Psi}(H):\\
&\qquad \Lambda=\{\Lambda_k\succeq 0,\,k\le K\},\ \Upsilon=\{\Upsilon_\ell\succeq 0,\,\ell\le L\},\\
&\qquad \left[\begin{array}{c|c}\sum_k\mathcal{R}_k^*[\Lambda_k] & \frac{1}{2}[B^T-A^TH]M\\ \hline \frac{1}{2}M^T[B-H^TA] & \sum_\ell \mathcal{S}_\ell^*[\Upsilon_\ell]\end{array}\right]\succeq 0\Big\},\\
&\overline{\Psi}(H):=\min_{G,\Upsilon'}\Big\{\phi_{\mathcal{R}}\big(\lambda[\Upsilon']\big)+\mathrm{Tr}(G):\Upsilon'=\{\Upsilon'_\ell\succeq 0,\,\ell\le L\},\\
&\qquad \left[\begin{array}{c|c}G & \frac{1}{2}Q_*^{1/2}HM\\ \hline \frac{1}{2}M^TH^TQ_*^{1/2} & \sum_\ell \mathcal{S}_\ell^*[\Upsilon'_\ell]\end{array}\right]\succeq 0\Big\},
\end{aligned}
\tag{5.19}
$$

where Opt is given by (3.13), and the equalities are due to (3.20) and (3.21).

2^o. From now on we assume that the observation noise ξ in observation (3.1) is $\xi\sim\mathcal{N}(0,Q_*)$. Besides this, we assume that $B\neq 0$, since otherwise the conclusion of Proposition 3.8 is evident.

3^o. Let W be a positive semidefinite $n\times n$ matrix, let $\eta\sim\mathcal{N}(0,W)$ be random signal, and let $\xi\sim\mathcal{N}(0,Q_*)$ be independent of η; vectors (η,ξ) induce random vector

$$
\omega=A\eta+\xi\sim\mathcal{N}\big(0,AWA^T+Q_*\big).
$$

Now, consider the problem where given ω we are interested to recover $B\eta$, and the Bayesian risk of a candidate estimate $\widehat{x}(\cdot)$ is quantified by $\mathbf{E}_{\eta,\xi}\{\|B\eta-\widehat{x}(A\eta+\xi)\|\}$. Let us set

$$
\varrho[W]=\inf_{\widehat{x}(\cdot)}\mathbf{E}_{\eta,\xi}\big\{\|B\eta-\widehat{x}(A\eta+\xi)\|\big\}. \tag{5.20}
$$

Our first observation is that $\varrho[W]$ is "nearly attainable" with a linear estimate. Since $[\omega;B\eta]$ is zero mean Gaussian, the conditional expectation $\mathbf{E}_{|\omega}\{B\eta\}$ of $B\eta$ given ω is linear in ω: $\mathbf{E}_{|\omega}\{B\eta\}=\bar{H}^T\omega$ for some $\bar{H}$ depending on W only. Given an estimate $\widehat{x}(\cdot)$, its Bayesian risk satisfies

$$
\begin{aligned}
\varrho &:= \mathbf{E}_{\eta,\omega}\big\{\|B\eta-\widehat{x}(\omega)\|\big\}=\mathbf{E}_\omega\big\{\mathbf{E}_{|\omega}\big\{\|B\eta-\widehat{x}(\omega)\|\big\}\big\}\\
&\geq \mathbf{E}_\omega\big\{\mathbf{E}_{|\omega}\big\{\|B\eta-\mathbf{E}_{|\omega}\{B\eta\}\|\big\}\big\}=\mathbf{E}_\omega\big\{\|B\eta-\bar{H}^T\omega\|\big\},
\end{aligned}
\tag{5.21}
$$

where the last inequality is due to the Anderson lemma [1]. Now (5.21) combines with independence of ξ, η and the Jensen inequality to imply that

$$\begin{aligned}\varrho &\geq \mathbf{E}_\eta\{\mathbf{E}_\xi\{\|B\eta - \bar{H}^T(A\eta+\xi)\|\}\}\\ &\geq \mathbf{E}_\eta\{\|\mathbf{E}_\xi\{B\eta - \bar{H}^T(A\eta+\xi)\}\|\} = \mathbf{E}_\eta\{\|(B-\bar{H}^TA)\eta\|\},\end{aligned}$$

that is,

$$\mathbf{E}_\eta\{\|(B-\bar{H}^TA)\eta\|\} \leq \varrho. \tag{5.22}$$

By "symmetric" reasoning,

$$\mathbf{E}_\xi\{\|\bar{H}^T\xi\|\} \leq \varrho. \tag{5.23}$$

In relations (5.22) and (5.23), $\bar{H}$ depends solely on W, and ϱ can be made arbitrarily close to $\varrho[W]$, thus we arrive at the following

Lemma 5.4. *Let W be a positive semidefinite $n \times n$ matrix. Then the risk $\varrho[W]$ defined by (5.20) satisfies the inequality*

$$\varrho[W] \geq \tfrac{1}{2}\inf_{H\in\mathbf{R}^{m\times\nu}}\left[\mathbf{E}_{\eta\sim\mathcal{N}(0,W)}\{\|[B-H^TA]\eta\|\} + \mathbf{E}_{\xi\sim\mathcal{N}(0,Q_*)}\{\|H^T\xi\|\}\right]. \tag{5.24}$$

4^o. Lemma 5.4 combines with Lemma 3.10 to imply the following result:

Lemma 5.5. *Let W be a positive semidefinite $n \times n$ matrix. Then the risk $\varrho[W]$ defined by (5.20) satisfies the inequality*

$$\begin{aligned}\varrho[W] \geq (2\kappa[F])^{-1}\min_{\Upsilon=\{\Upsilon_\ell,\ell\leq L\},G,H}\Big\{&\mathrm{Tr}(WG) + \phi_{\mathcal{R}}(\lambda[\Upsilon]) + \overline{\Psi}(H):\\ &\Upsilon_\ell \succeq 0,\ \forall \ell, \left[\begin{array}{c|c} G & \frac{1}{2}[B^T - A^TH]M \\ \hline \frac{1}{2}M^T[B-H^TA] & \sum_\ell \mathcal{S}_\ell^*[\Upsilon_\ell]\end{array}\right] \succeq 0\Big\},\end{aligned} \tag{5.25}$$

where $\overline{\Psi}(H)$ is given by (5.19) and

$$\kappa[F] = \frac{8}{\sqrt{2}-\mathrm{e}^{1/4}}\sqrt{\ln\left(\frac{4\sqrt{2}F}{\sqrt{2}-\mathrm{e}^{1/4}}\right)}.$$

Proof. Let H be $m\times\nu$ matrix. Applying Lemma 3.10 to $N = m$, $Y = H$, $Q = Q_*$, we get

$$\mathbf{E}_{\xi\sim\mathcal{N}(0,Q_*)}\{\|H^T\xi\|\} \geq \kappa^{-1}[F]\,\overline{\Psi}(H). \tag{5.26}$$

Applying Lemma 3.10 to $N = n$, $Y = (B - H^TA)^T$, $Q = W$, we get

$$\begin{aligned}\kappa[F]\mathbf{E}_{\eta\sim\mathcal{N}(0,W)}\{\|[B-H^TA]\eta\|\} \geq \min_{\Upsilon=\{\Upsilon_\ell\succeq 0,\,\ell\leq L\},G}\Big\{&\phi_{\mathcal{R}}(\lambda[\Upsilon]) + \mathrm{Tr}(WG):\\ &\left[\begin{array}{c|c} G & \frac{1}{2}[B^T - A^TH]M \\ \hline \frac{1}{2}M^T[B-H^TA] & \sum_\ell \mathcal{S}_\ell^*[\Upsilon_\ell]\end{array}\right] \succeq 0\Big\}.\end{aligned}$$

The resulting inequality combines with (5.24) and (5.26) to imply (5.25). □

5^o. For $0 < \varkappa \le 1$, let us set

$$\begin{array}{ll}(a) & \mathcal{W}_\varkappa = \left\{W \in \mathbf{S}^n_+ : \exists t \in \mathcal{T} : \mathcal{R}_k[W] \preceq \varkappa t_k I_{d_k},\ 1 \le k \le K\right\},\\ (b) & \mathcal{Z} = \left\{\big(\Upsilon = \{\Upsilon_\ell,\ \ell \le L\}, G, H\big) : \Upsilon_\ell \succeq 0,\ \forall \ell,\right.\\ & \left.\left[\begin{array}{c|c} G & \frac{1}{2}\left[B^T - A^T H\right]M \\ \hline \frac{1}{2}M^T\left[B - H^T A\right] & \sum_\ell \mathcal{S}^*_\ell[\Upsilon_\ell]\end{array}\right] \succeq 0\right\}.\end{array} \tag{5.27}$$

Note that $\mathcal{W}_\varkappa$ is a nonempty convex compact (by Lemma 5.1) set such that $\mathcal{W}_\varkappa = \varkappa \mathcal{W}_1$, and $\mathcal{Z}$ is a nonempty closed convex set. Consider the parametric saddle point problem

$$\mathrm{Opt}(\varkappa) = \max_{W \in \mathcal{W}_\varkappa} \min_{(\Upsilon, G, H) \in \mathcal{Z}} \left[E(W; \Upsilon, G, H) := \mathrm{Tr}(WG) + \phi_{\mathcal{R}}\big(\lambda[\Upsilon]\big) + \overline{\Psi}(H)\right]. \tag{5.28}$$

This problem is convex-concave; utilizing the fact that $\mathcal{W}_\varkappa$ is compact and contains positive definite matrices, it is immediately seen that the Sion–Kakutani theorem ensures the existence of a saddle point whenever $\varkappa \in (0, 1]$. We claim that

$$0 < \varkappa \le 1 \Rightarrow \mathrm{Opt}(\varkappa) \ge \sqrt{\varkappa}\,\mathrm{Opt}(1). \tag{5.29}$$

Indeed, $\mathcal{Z}$ is invariant w.r.t. scalings

$$\big(\Upsilon = \{\Upsilon_\ell, \ell \le L\}, G, H\big) \mapsto \big(\theta\Upsilon := \{\theta\Upsilon_\ell, \ell \le L\}, \theta^{-1}G, H\big), \quad [\theta > 0].$$

When taking into account that $\phi_{\mathcal{R}}(\lambda[\theta\Upsilon]) = \theta\phi_{\mathcal{R}}(\lambda[\Upsilon])$, we get

$$\begin{aligned}\underline{E}(W) :=& \min_{(\Upsilon,G,H)\in\mathcal{Z}} E(W; \Upsilon, G, H)\\ =& \min_{(\Upsilon,G,H)\in\mathcal{Z}} \inf_{\theta>0} E(W; \theta\Upsilon, \theta^{-1}G, H)\\ =& \min_{(\Upsilon,G,H)\in\mathcal{Z}} \left[2\sqrt{\mathrm{Tr}(WG)\phi_{\mathcal{R}}\big(\lambda[\Upsilon]\big)} + \overline{\Psi}(H)\right].\end{aligned}$$

Because $\overline{\Psi}$ is nonnegative we conclude that whenever $W \succeq 0$ and $\varkappa \in (0, 1]$, one has

$$\underline{E}(\varkappa W) \ge \sqrt{\varkappa}\,\underline{E}(W),$$

which combines with $\mathcal{W}_\varkappa = \varkappa \mathcal{W}_1$ to imply that

$$\mathrm{Opt}(\varkappa) = \max_{W\in\mathcal{W}_\varkappa} \underline{E}(W) = \max_{W\in\mathcal{W}_1} \underline{E}(\varkappa W) \ge \sqrt{\varkappa} \max_{W\in\mathcal{W}_1} \underline{E}(W) = \sqrt{\varkappa}\,\mathrm{Opt}(1),$$

and (5.29) follows.

6^o. We claim that

$$\mathrm{Opt}(1) = \mathrm{Opt}, \tag{5.30}$$

where Opt is given by (3.13) (and, as we have seen, by (5.19) as well). Note that (5.30) combines with (5.29) to imply that

$$0 < \varkappa \le 1 \Rightarrow \mathrm{Opt}(\varkappa) \ge \sqrt{\varkappa}\,\mathrm{Opt}\,. \tag{5.31}$$

Verification of (5.30) is given by the following computation. By the Sion–Kakutani theorem,

$$\begin{aligned}
\mathrm{Opt}(1) &= \max_{W\in\mathcal{W}_1}\ \min_{(\Upsilon,G,H)\in\mathcal{Z}} \left\{\mathrm{Tr}(WG) + \phi_{\mathcal{R}}(\lambda[\Upsilon]) + \overline{\Psi}(H)\right\}\\
&= \min_{(\Upsilon,G,H)\in\mathcal{Z}}\ \max_{W\in\mathcal{W}_1} \left\{\mathrm{Tr}(WG) + \phi_{\mathcal{R}}(\lambda[\Upsilon]) + \overline{\Psi}(H)\right\}\\
&= \min_{(\Upsilon,G,H)\in\mathcal{Z}} \Big\{\overline{\Psi}(H) + \phi_{\mathcal{R}}(\lambda[\Upsilon])\\
&\qquad + \max_{W}\left\{\mathrm{Tr}(GW) : W\succeq 0,\ \exists t\in\mathcal{T} : \mathcal{R}_k[W]\preceq t_k I_{d_k},\ k\le K\right\}\Big\}\\
&= \min_{(\Upsilon,G,H)\in\mathcal{Z}} \Big\{\overline{\Psi}(H) + \phi_{\mathcal{R}}(\lambda[\Upsilon])\\
&\qquad + \max_{W,t}\left\{\mathrm{Tr}(GW) : W\succeq 0,\ [t;1]\in\mathbf{K}[\mathcal{T}],\ \mathcal{R}_k[W]\preceq t_k I_{d_k},\ k\le K\right\}\Big\}.
\end{aligned}$$

Now, using Conic Duality combined with the fact that $(\mathbf{K}[\mathcal{T}])_* = \{[g;s] : s\ge\phi_{\mathcal{T}}(-g)\}$ we obtain

$$\begin{aligned}
&\max_{W,t}\left\{\mathrm{Tr}(GW) : W\succeq 0, [t;1]\in\mathbf{K}[\mathcal{T}],\ \mathcal{R}_k[W]\preceq t_k I_{d_k},\ k\le K\right\}\\
&\quad= \min_{Z,[g;s],\,\Lambda=\{\Lambda_k\}} \Big\{s : \big\{Z\succeq 0, [g;s]\in\big(\mathbf{K}[\mathcal{T}]\big)_*,\ \Lambda_k\succeq 0,\ k\le K,\\
&\qquad -\mathrm{Tr}(ZW) - g^T t + \sum_k \mathrm{Tr}\left(\mathcal{R}_k^*[\Lambda_k]W\right) - \sum_k t_k\,\mathrm{Tr}(\Lambda_k) = G,\\
&\qquad \forall\big(W\in\mathbf{S}^n,\ t\in\mathbf{R}^K\big)\big\}\Big\}\\
&\quad= \min_{Z,[g;s],\Lambda=\{\Lambda_k\}} \Big\{s : \big\{Z\succeq 0,\ s\ge\phi_{\mathcal{T}}(-g),\ \Lambda_k\succeq 0,\ k\le K,\\
&\qquad G = \sum_k \mathcal{R}_k^*[\Lambda_k] - Z,\ g = -\lambda[\Lambda]\big\}\Big\}\\
&\quad= \min_{\Lambda}\Big\{\phi_{\mathcal{T}}\big(\lambda[\Lambda]\big) : \Lambda = \{\Lambda_k\succeq 0,\ k\le K\},\ G\preceq \sum_k \mathcal{R}_k^*[\Lambda_k]\Big\},
\end{aligned}$$

and we arrive at

$$\begin{array}{rl}\mathrm{Opt}(1) = & \min\limits_{\Upsilon,G,H,\Lambda}\Big\{\overline{\Psi}(H)+\phi_{\mathcal{R}}(\lambda[\Upsilon])+\phi_{\mathcal{T}}(\lambda[\Lambda]):\\ & \Upsilon=\{\Upsilon_\ell\succeq 0,\,\ell\leq L\},\ \Lambda=\{\Lambda_k\succeq 0,\,k\leq K\}, G\preceq\sum_k \mathcal{R}_k^*[\Lambda_k],\\ & \left[\begin{array}{c|c} G & \frac{1}{2}[B^T-A^TH]M\\ \hline \frac{1}{2}M^T[B-H^TA] & \sum_\ell \mathcal{S}_\ell^*[\Upsilon_\ell]\end{array}\right]\succeq 0\Big\}\\ = & \min\limits_{\Upsilon,H,\Lambda}\Big\{\overline{\Psi}(H)+\phi_{\mathcal{R}}(\lambda[\Upsilon])+\phi_{\mathcal{T}}(\lambda[\Lambda]):\\ & \Upsilon=\{\Upsilon_\ell\succeq 0,\,\ell\leq L\},\Lambda=\{\Lambda_k\succeq 0,\,k\leq K\},\\ & \left[\begin{array}{c|c} \sum_k \mathcal{R}_k^*[\Lambda_k] & \frac{1}{2}[B^T-A^TH]M\\ \hline \frac{1}{2}M^T[B-H^TA] & \sum_\ell \mathcal{S}_\ell^*[\Upsilon_\ell]\end{array}\right]\succeq 0\Big\}\\ = & \mathrm{Opt}\ \ [\text{see } (5.19)].\end{array}$$

Now we can complete the proof.

7^o. Let us set

$$\varrho_* = \inf_{\widehat{x}(\cdot)}\mathrm{Risk}\big[\widehat{x}|\mathcal{X}\big],\quad \mathrm{Risk}\big[\widehat{x}|\mathcal{X}\big] = \sup_{x\in\mathcal{X}}\mathbf{E}_{\xi\sim\mathcal{N}(0,Q_*)}\big\{\|Bx-\widehat{x}(Ax+\xi)\|\big\}, \tag{5.32}$$

where inf is taken over all estimates. It is immediately seen that $\varrho_*>0$ due to $Q_*\succ 0$ (recall that $Q_*\in\mathcal{Q}$ and that $\mathcal{Q}\subset\mathrm{int}\,\mathbf{S}^m_+$) combined with $B\neq 0$ and $0\in\mathrm{int}\,\mathcal{X}$. Consequently, there is an estimate $\widetilde{x}(\cdot)$ such that $\mathrm{Risk}[\widetilde{x}|\mathcal{X}]\leq\frac{3}{2}\varrho_*$. Further, when $x\in\mathcal{X}\backslash\{0\}$, we have $W:=xx^T\in\mathcal{W}$, see (3.17) and (2.4), and $W^{1/2}=W/\|x\|_2$. Hence for M_* as defined in (3.17) we have

$$M_*^2\geq\mathbf{E}_{\eta\sim\mathcal{N}(0,I_n)}\big\{\|BW^{1/2}\eta\|^2\big\}=\|x\|_2^{-2}\|Bx\|^2\mathbf{E}_{\eta\sim\mathcal{N}(0,I_n)}\big\{(x^T\eta)^2\big\}=\|Bx\|^2,$$

and we arrive at

$$x\in\mathcal{X}\Rightarrow\|Bx\|\leq M_*. \tag{5.33}$$

Now let us convert the estimate $\widetilde{x}$ into the estimate $\widehat{x}$ defined as follows: $\widehat{x}(\omega)$ is the $\|\cdot\|$-closest to $\widetilde{x}(\omega)$ point of the set $\mathcal{B}_{M_*}=\{u:\|u\|\leq M_*\}$. We have $Bx\in\mathcal{B}_{M_*}$ by (5.33), and because, by construction, $\widehat{x}$ is the closest to $\widetilde{x}$ point of $\mathcal{B}_{M_*}$, we get

$$x\in\mathcal{X}\Rightarrow\|Bx-\widehat{x}(\omega)\|\leq\|Bx-\widetilde{x}(\omega)\|+\|\widetilde{x}(\omega)-\widehat{x}(\omega)\|\leq 2\|Bx-\widetilde{x}(\omega)\|.$$

We conclude that $\|\widehat{x}(\omega)\|\leq M_*\ \forall\omega$, and

$$\mathrm{Risk}\big[\widehat{x}|\mathcal{X}\big]\leq 2\,\mathrm{Risk}\big[\widetilde{x}|\mathcal{X}\big]\leq 3\varrho_*. \tag{5.34}$$

8^o. For $\varkappa \in (0,1]$, let $W_\varkappa$ be the W-component of a saddle point solution to the saddle point problem (5.28). Then, by (5.31),

$$\begin{array}{l}\sqrt{\varkappa}\,\mathrm{Opt} \leq \mathrm{Opt}(\varkappa) = \min\limits_{(\Upsilon,G,H)\in\mathcal{Z}} \left\{\mathrm{Tr}(W_\varkappa G) + \phi_{\mathcal{R}}\big(\lambda[\Upsilon]\big) + \overline{\Psi}(H)\right\}\\
\quad = \min\limits_{(\Upsilon,G,H)} \left\{\mathrm{Tr}(W_\varkappa G) + \phi_{\mathcal{R}}\big(\lambda[\Upsilon]\big) + \overline{\Psi}(H) : \Upsilon_\ell \succeq 0\,\forall \ell,\right.\\
\qquad\qquad \left.\left[\begin{array}{c|c} G & \frac{1}{2}[B^T - A^T H]M \\ \hline \frac{1}{2}M^T[B - H^T A] & \sum_\ell \mathcal{S}_\ell^*[\Upsilon_\ell]\end{array}\right] \succeq 0\right\}\\
\quad \leq 2\kappa[F]\varrho[W_\varkappa]\end{array} \tag{5.35}$$

(we have used (5.27.b) and (5.25); recall that $\varrho[\cdot]$ is given by (5.20)). On the other hand, when applying Lemma 5.3 to $Q = W_\varkappa$ and $\rho = \varkappa$, we obtain, in view of relations $0 < \varkappa \leq 1$, $W_\varkappa \in \mathcal{W}_\varkappa$,

$$\delta(\varkappa) := \mathrm{Prob}_{\eta\sim\mathcal{N}(0,I_n)}\left\{W_\varkappa^{1/2}\eta \not\in \mathcal{X}\right\} \leq 2D\exp\left\{-(2\varkappa)^{-1}\right\}, \tag{5.36}$$

with D given by (3.19). Setting

$$\mathcal{E}_\varkappa = \left\{\zeta : W_\varkappa^{1/2}\zeta \in \mathcal{X}\right\},\quad \mathcal{E}_\varkappa^c = \mathbf{R}^n\backslash\mathcal{E}_\varkappa,\quad \Sigma = \mathrm{Diag}\{I_n, Q_*\},$$

we have by definition of the risk $\varrho[W_\varkappa]$

$$\begin{array}{rl}\varrho[W_\varkappa] \leq & \mathbf{E}_{(\eta,\xi)\sim\mathcal{N}(0,\Sigma)}\left\{\|BW_\varkappa^{1/2}\eta - \widehat{x}\big(AW_\varkappa^{1/2}\eta + \xi\big)\|\right\}\\
= & \mathbf{E}_{\eta\sim\mathcal{N}(0,I_n)}\left\{\mathbf{E}_{\xi\sim\mathcal{N}(0,Q_*)}\left\{\|BW_\varkappa^{1/2}\eta - \widehat{x}(AW_\varkappa^{1/2}\eta + \xi)\|\right\}\right\}\\
= & \mathbf{E}_{\eta\sim\mathcal{N}(0,I_n)}\left\{\mathbf{E}_{\xi\sim\mathcal{N}(0,Q_*)}\left\{\|BW_\varkappa^{1/2}\eta - \widehat{x}\big(AW_\varkappa^{1/2}\eta + \xi\big)\|\right\}1\{\eta\in\mathcal{E}_\varkappa\}\right\}\\
& + \mathbf{E}_{\eta\sim\mathcal{N}(0,I_n)}\left\{\mathbf{E}_{\xi\sim\mathcal{N}(0,Q_*)}\left\{\|BW_\varkappa^{1/2}\eta - \widehat{x}(AW_\varkappa^{1/2}\eta + \xi)\|\right\}1\{\eta\in\mathcal{E}_\varkappa^c\}\right\}\\
\leq & \mathrm{Risk}\left[\widehat{x}|\mathcal{X}\right] + \mathbf{E}_{\eta\sim\mathcal{N}(0,I_n)}\left\{\big(\|BW_\varkappa^{1/2}\eta\| + M_*\big)1\{\eta\in\mathcal{E}_\varkappa^c\}\right\}\\
& \hfill \left[\text{since } \|\widehat{x}(\cdot)\| \leq M_*\right]\\
\leq & 3\varrho_* + M_*\delta(\varkappa) + \mathbf{E}_{\eta\sim\mathcal{N}(0,I_n)}\left\{\|BW_\varkappa^{1/2}\eta\|1\{\eta\in\mathcal{E}_\varkappa^c\}\right\}\\
& \hfill \left[\text{we have used (5.34)}\right].\end{array}$$

We conclude that

$$\begin{array}{rl}\varrho[W_\varkappa] \leq & 3\varrho_* + M_*\delta(\varkappa) + \left[\mathbf{E}_{\eta\sim\mathcal{N}(0,I_n)}\left\{\|BW_\varkappa^{1/2}\eta\|^2\right\}\right]^{1/2}\left[\mathrm{Prob}_{\eta\sim\mathcal{N}(0,I_n)}\{\eta\in\mathcal{E}_\varkappa^c\}\right]^{1/2}\\
\leq & 3\varrho_* + M_*\left[\delta(\varkappa) + \sqrt{\delta(\varkappa)}\right] \quad \left[\text{by (3.17); note that } W_\varkappa \in \mathcal{W} \text{ due to } \varkappa \leq 1\right]\\
\leq & 3\varrho_* + 2M_*\sqrt{\delta(\varkappa)} \quad \left[\text{since } \delta(\varkappa) \leq 1\right]\\
\leq & 3\varrho_* + 2M_*\sqrt{2D}\exp\left\{-(4\varkappa)^{-1}\right\} \quad \left[\text{we have used (5.36)}\right].\end{array}$$

The bottom line here is that

$$0 < \varkappa \leq 1 \Rightarrow \varrho[W_\varkappa] \leq 3\varrho_* + 2M_*\sqrt{2D}\exp\left\{-\frac{1}{4\varkappa}\right\}. \tag{5.37}$$

Observe that $\varrho_* \leq M_*$, since due to (5.33), for the trivial — identically zero — estimate $\bar{x}(\cdot)$ of Bx one has $\mathrm{Risk}[\bar{x}|\mathcal{X}] \leq M_*$. It follows that setting

$$\bar{\varkappa} = \frac{1}{4\ln\left(2M_*\sqrt{2D}/\varrho_*\right)}$$

we ensure that $\bar{\varkappa} \in (0,1]$, whence, by (5.37),

$$\varrho[W_{\bar{\varkappa}}] \leq 4\varrho_*.$$

This combines with (5.35) to imply that

$$\sqrt{\bar{\varkappa}}\,\mathrm{Opt} \leq 2\kappa[F]\varrho[W_{\bar{\varkappa}}] \leq 8\kappa[F]\varrho_*.$$

Hence, finally,

$$\mathrm{Opt} \leq \frac{8\kappa[F]}{\sqrt{\bar{\varkappa}}}\varrho_* \leq \frac{64\sqrt{2}}{\sqrt{2}-\mathrm{e}^{1/4}}\sqrt{\ln\left(\frac{4\sqrt{2}F}{\sqrt{2}-\mathrm{e}^{1/4}}\right)\ln\left(\frac{8M_*^2D}{\varrho_*^2}\right)}\varrho_*.$$

Noting that by definition of ϱ_* and $\mathrm{RiskOpt}_{\mathcal{Q},\|\cdot\|}[\mathcal{X}]$ we have

$$\varrho_* \leq \mathrm{RiskOpt}_{\mathcal{Q},\|\cdot\|}[\mathcal{X}] \leq M_*$$

(the concluding $\leq$ is due to $\|Bx\| \leq M_*$ for $x \in \mathcal{X}$), we arrive at (3.18). □

5.8. Proof of Proposition 4.1.

1^o. Item (i) is a direct consequence of Proposition 2.2, modulo the claim that problem (4.1) is solvable, and we start with justifying this claim. Let $F = \mathrm{Im}\,A$. Clearly, feasibility of a candidate solution (H, Λ, Υ) to the problem depends solely on the restriction of the linear mapping $z \mapsto H^T z$ onto F, so that adding to the constraints of the problem the requirement that the restriction of this linear mapping on the orthogonal complement of F in $\mathbf{R}^m$ is identically zero, we get an equivalent problem. It is immediately seen that in the resulting problem, the feasible solutions with the value of the objective $\leq a$ for every $a \in \mathbf{R}$ form a compact set, so that the latter problem (and thus, the original one) indeed is solvable.

Let us prove the near-optimality result of (ii).

2^o. Observe that setting

$$\varrho = \max_x\left\{\|Bx\| : x \in \mathcal{X},\ Ax = 0\right\}, \tag{5.38}$$

we ensure that

$$\mathrm{Risk}_{\mathrm{opt}}[\mathcal{X}] \geq \varrho. \tag{5.39}$$

Indeed, let $\bar{x}$ be an optimal solution to the (clearly solvable) optimization problem in (5.38). Then observation $\omega = 0$ can be obtained from both the signals $x = \bar{x}$ and $x = -\bar{x}$, and therefore the risk of any (deterministic) recovery routine is at least $\|B\bar{x}\| = \varrho$, as claimed.

3^o. It may happen that $\mathrm{Ker}\, A = \{0\}$. In this case the situation is trivial: specifying $A^\dagger$ as a partial inverse to A: $A^\dagger A = I_n$ and setting $H^T = BA^\dagger$ (so that $B - H^T A = 0$), $\Upsilon_\ell = 0_{f_\ell \times f_\ell}$, $\ell \leq L$, $\Lambda_k = 0_{d_k \times d_k}$, $k \leq K$, we get a feasible solution to the optimization problem in (4.1) with zero value of the objective, implying that $\mathrm{Opt}_\# = 0$; consequently, the linear estimate induced by an optimal solution to the problem is with zero risk, and the conclusion of Proposition 4.1 is clearly true. With this in mind, we assume from now on that $\mathrm{Ker}\, A \neq \{0\}$. Denoting $\kappa = \dim \mathrm{Ker}\, A$, we can build an $n \times \kappa$ matrix E of rank κ such that $\mathrm{Ker}\, A$ is the image space of E.

4^o. Setting

$$
\begin{aligned}
\mathcal{Z} &:= \left\{z \in \mathbf{R}^\kappa : Ez \in \mathcal{X}\right\} \\
&= \left\{z \in \mathbf{R}^\kappa : \exists (t \in \mathcal{T}) : \bar{R}_k^2[z] \preceq t_k I_{d_k}, k \leq K\right\}, \quad \bar{R}_k[z] = R_k[Ez], \\
C &= \left[\begin{array}{c|c} & \frac{1}{2}BE \\ \hline \frac{1}{2}E^T B^T & \end{array}\right],
\end{aligned}
$$

note that when z runs trough the spectratope $\mathcal{Z}$, Ez runs exactly through the entire set $\{x \in \mathcal{X} : Ax = 0\}$. With this in mind, invoking Proposition 2.2, we arrive at

$$
\begin{aligned}
\varrho = & \max_{g:\|g\|_* \leq 1} \max_{z \in \mathcal{Z}} g^T BEz = \max_{[u;z] \in \mathcal{B}_* \times \mathcal{Z}} [u;z]^T C [u;z] \\
\leq & \ \mathrm{Opt} := \min_{\substack{\Upsilon = \{\Upsilon_\ell : \ell \leq L\}, \\ \Lambda = \{\Lambda_k, k \leq K\}}} \Big\{ \phi_{\mathcal{R}}(\lambda[\Upsilon]) + \phi_{\mathcal{T}}(\lambda[\Lambda]) : \Upsilon_\ell \succeq 0, \ \Lambda_k \succeq 0, \ \forall (\ell, k), \\
& \left[\begin{array}{c|c} \sum_\ell S_\ell^*[\Upsilon_\ell] & \frac{1}{2}BE \\ \hline \frac{1}{2}E^T B^T & E^T\left[\sum_k \mathcal{R}_k^*[\Lambda_K]\right]E \end{array}\right] \succeq 0 \Big\}
\end{aligned}
\tag{5.40}
$$

(we have used the straightforward identity $\bar{\mathcal{R}}_k^*[\Lambda_k] = E^T \mathcal{R}_k^*[\Lambda_k]E$). By the same Proposition 2.2, the optimization problem in (5.40) specifying Opt is solvable, and

$$
\varrho \leq \mathrm{Opt} \leq 2\ln(2D)\varrho, \quad D = \sum_k d_k + \sum_\ell f_\ell. \tag{5.41}
$$

5^o. Let $\bar{\Upsilon} = \{\bar{\Upsilon}_\ell\}$, $\bar{\Lambda} = \{\bar{\Lambda}_k\}$ be an optimal solution to the optimization problem specifying Opt, see (5.40), and let

$$
\boldsymbol{\Upsilon} = \sum_\ell \mathcal{S}_\ell^*[\bar{\Upsilon}_\ell], \quad \boldsymbol{\Lambda} = \sum_k \mathcal{R}_k^*[\bar{\Lambda}_k],
$$

so that

$$
\mathrm{Opt} = \phi_{\mathcal{R}}(\lambda[\bar{\Upsilon}]) + \phi_{\mathcal{T}}(\lambda[\bar{\Lambda}]) \quad \text{and} \quad \left[\begin{array}{c|c} \boldsymbol{\Upsilon} & \frac{1}{2}BE \\ \hline \frac{1}{2}E^T B^T & E^T \boldsymbol{\Lambda} E \end{array}\right] \succeq 0. \tag{5.42}
$$

We claim that for properly selected $m \times \nu$ matrix H it holds

$$\left[\begin{array}{c|c} \boldsymbol{\Upsilon} & \frac{1}{2}(B - H^T A) \\ \hline \frac{1}{2}(B - H^T A)^T & \boldsymbol{\Lambda} \end{array}\right] \succeq 0. \tag{5.43}$$

This claim implies the conclusion of Proposition 4.1: by the claim, we have $\overline{\mathrm{Opt}} \leq \mathrm{Opt}$, which combines with (5.41) and (5.39) to imply (4.2).

In order to justify the claim, assume that it fails to be true, and let us lead this assumption to a contradiction. To this end, consider the semidefinite program

$$\tau_* = \min_{\tau, H} \left\{ \tau : \left[\begin{array}{c|c} \boldsymbol{\Upsilon} & \frac{1}{2}(B - H^T A) \\ \hline \frac{1}{2}(B - H^T A)^T & \boldsymbol{\Lambda} \end{array}\right] + \tau I_{\nu+n} \succeq 0 \right\}. \tag{5.44}$$

The problem clearly is strictly feasible, and the value of the objective at every feasible solution is positive. In addition, the problem is solvable (by exactly the same argument as in item 1^o of the proof).

4^o.b. As we have seen, (5.44) is a strictly feasible solvable problem with positive optimal value τ_*, so that the problem dual to (5.44) is solvable with positive optimal value. Let us build the dual problem. Denoting by

$$\left[\begin{array}{c|c} U & V \\ \hline V^T & W \end{array}\right] \succeq 0$$

the Lagrange multipliers for the semidefinite constraint in (5.44) and taking inner product of the left hand side of the constraint with the multiplier, we get the aggregated constraint

$$\mathrm{Tr}(U\boldsymbol{\Upsilon}) + \mathrm{Tr}(W\boldsymbol{\Lambda}) + \tau\big[\mathrm{Tr}(U) + \mathrm{Tr}(W)\big] + \mathrm{Tr}\left((B - H^T A)V^T\right) \geq 0.$$

The equality constraints of the dual problem should make the homogeneous in τ, H part of the left hand side in the aggregated constraint identically equal to τ, which amounts to

$$\mathrm{Tr}(U) + \mathrm{Tr}(W) = 1, \quad VA^T = 0, \tag{5.45}$$

so the aggregated constraint reads

$$\tau \geq -\big[\mathrm{Tr}(U\boldsymbol{\Upsilon}) + \mathrm{Tr}(W\boldsymbol{\Lambda}) + \mathrm{Tr}(BV^T)\big].$$

The dual problem is to maximize the right hand side of the latter constraint over Lagrange multiplier

$$\left[\begin{array}{c|c} U & V \\ \hline V^T & W \end{array}\right] \succeq 0$$

satisfying (5.45), and its optimal value is $\tau_* > 0$, that is, there exists

$$\left[\begin{array}{c|c} \bar{U} & \bar{V} \\ \hline \bar{V}^T & \overline{W} \end{array}\right] \succeq 0$$

such that $A\bar{V}^T = 0$ and

$$\mathrm{Tr}(\bar{U}\Upsilon) + \mathrm{Tr}(\overline{W}\Lambda) + \mathrm{Tr}(B\bar{V}^T) < 0. \tag{5.46}$$

Adding to $\bar{U}$ a small positive multiple of the unit matrix, we can assume, in addition, that $\bar{U} \succ 0$. Now, the relation $A\bar{V}^T = 0$ combines with the definition of E to imply that $\bar{V}^T = EF$ for properly selected matrix F, so that

$$\left[\begin{array}{c|c} \bar{U} & F^T E^T \\ \hline EF & \overline{W} \end{array}\right] \succeq 0.$$

Hence, by the Schur Complement Lemma,

$$\overline{W} \succeq EF\bar{U}^{-1}F^T E^T,$$

and (5.46) combines with $\Lambda \succeq 0$ to imply that

$$\begin{aligned} 0 &> \mathrm{Tr}(\bar{U}\Upsilon) + \mathrm{Tr}(\overline{W}\Lambda) + \mathrm{Tr}(B\bar{V}^T) \\ &= \mathrm{Tr}(\bar{U}\Upsilon) + \mathrm{Tr}(\overline{W}\Lambda) + \mathrm{Tr}(BEF) \\ &\geq \mathrm{Tr}(\bar{U}\Upsilon) + \mathrm{Tr}(EF\bar{U}^{-1}F^T E^T\Lambda) + \mathrm{Tr}(BEF) \\ &= \mathrm{Tr}\left(\left[\begin{array}{c|c} \Upsilon & \frac{1}{2}BE \\ \hline \frac{1}{2}E^T B^T & E^T\Lambda E \end{array}\right]\left[\begin{array}{c|c} \bar{U} & F^T \\ \hline F & F\bar{U}^{-1}F^T \end{array}\right]\right). \end{aligned}$$

Both matrix factors in the concluding the chain $\mathrm{Tr}(\cdot)$ are positive semidefinite (the first one due to (5.42), and the second — by the Schur Complement Lemma); consequently, the concluding quantity in the chain is nonnegative, which is impossible. We have arrived at a desired contradiction. □

A. Calculus of spectratopes

The principal rules of the calculus of spectratopes are as follows:

Finite intersections. If

$$\mathcal{X}_\ell = \big\{x \in \mathbf{R}^\nu : \exists (y^\ell \in \mathbf{R}^{n_\ell}, t^\ell \in \mathcal{T}_\ell) : \\ x = P_\ell y^\ell,\ R_{k\ell}^2[y^\ell] \preceq t_k^\ell I_{d_{k\ell}},\ k \leq K_\ell\big\}, \quad 1 \leq \ell \leq L,$$

are spectratopes, so is $\mathcal{X} = \bigcap_{\ell \le L} \mathcal{X}_\ell$. Indeed, let

$$E = \{[y = [y^1; \dots; y^L] \in \mathbf{R}^{n_1} \times \dots \times \mathbf{R}^{N_L} : P_1 y^1 = P_2 y^2 = \dots = P_L y^L\}.$$

When $E = \{0\}$, we have $\mathcal{X} = \{0\}$, so that $\mathcal{X}$ is a spectratope; when $E \neq \{0\}$, we have

$$\begin{aligned}\mathcal{X} = \big\{x \in \mathbf{R}^\nu : \exists \big(y = [y^1; \dots; y^L] \in E,\ t = [t^1; \dots; t^L] \in \mathcal{T} := \mathcal{T}_1 \times \dots \times \mathcal{T}_L\big) : \\ x = Py := P_1 y^1,\ R^2_{k\ell}[y^\ell] \preceq t^\ell_k I_{d_{k\ell}},\ 1 \le k \le K_\ell,\ 1 \le \ell \le L\big\};\end{aligned}$$

identifying E and appropriate $\mathbf{R}^n$, we arrive at a valid spectratopic representation of $\mathcal{X}$.

Direct product. If

$$\begin{aligned}\mathcal{X}_\ell = \big\{x^\ell \in \mathbf{R}^{\nu_\ell} : \exists (y^\ell \in \mathbf{R}^{n_\ell},\ t^\ell \in \mathcal{T}_\ell) : \\ x^\ell = P_\ell y^\ell,\ R^2_{k\ell}[y^\ell] \preceq t^\ell_k I_{d_{k\ell}},\ k \le K_\ell\big\}, \quad 1 \le \ell \le L,\end{aligned}$$

are spectratopes, so is $\mathcal{X} = \mathcal{X}_1 \times \dots \times \mathcal{X}_L$:

$$\begin{aligned}\mathcal{X}_1 \times \dots \times \mathcal{X}_L = \big\{x = [x^1; \dots; x^L] : \\ \exists \big(y = [y^1; \dots; y^L], t = [t^1; \dots; t_L] \in \mathcal{T} = \mathcal{T}_1 \times \dots \times \mathcal{T}_\ell\big) : \\ x = Py := [P_1 y^1; \dots; P_L y^L], R^2_{k\ell}[y^\ell] \preceq t^\ell_k I_{d_{k\ell}}, 1 \le k \le K_\ell, 1 \le \ell \le L\big\};\end{aligned}$$

Linear image. If

$$\mathcal{X} = \big\{x \in \mathbf{R}^\nu : \exists (y \in \mathbf{R}^n,\ t \in \mathcal{T}) : x = Py,\ R^2_k[y] \preceq t_k I_{d_k},\ k \le K\big\}$$

is a spectratope and S is a $\mu \times \nu$ matrix, the set $S\mathcal{X} = \{z = Sx : x \in \mathcal{X}\}$ is a spectratope:

$$S\mathcal{X} = \big\{z \in \mathbf{R}^\mu : \exists (y \in \mathbf{R}^n,\ t \in \mathcal{T}) : z = SPy,\ R^2_k[y] \preceq t_k I_{d_k},\ k \le K\big\}.$$

Inverse linear image under embedding. If

$$\mathcal{X} = \big\{x \in \mathbf{R}^\nu : \exists (y \in \mathbf{R}^m,\ t \in \mathcal{T}) : x = Py,\ R^2_k[y] \preceq t_k I_{d_k},\ k \le K\big\}$$

is a spectratope, and S is a $\nu \times \mu$ matrix with trivial kernel, the set $S^{-1}\mathcal{X} = \{z : Sz \in \mathcal{X}\}$ is a spectratope. Indeed, setting $E = \{y \in \mathbf{R}^m : Py \in \operatorname{Im} S\}$, we get a linear subspace of $\mathbf{R}^n$; if $E = \{0\}$, $S^{-1}\mathcal{X} = \{0\}$ is a spectratope, otherwise we have

$$S^{-1}\mathcal{X} = \big\{z \in \mathbf{R}^\mu : \exists (y \in E,\ t \in \mathcal{T}) : z = Qy,\ R^2_k[y] \preceq t_k I_{d_k},\ k \le K\big\},$$

where the linear mapping $y \mapsto Qy : E \to \mathbf{R}^\mu$ is uniquely defined by the relation $Py = SQy$. When identifying E with appropriate $\mathbf{R}^n$, we get a valid spectratopic representation of $S^{-1}\mathcal{X}$.

Arithmetic sum. If $\mathcal{X}_\ell$, $\ell \leq L$, are spectratopes in $\mathbf{R}^\nu$, so is the arithmetic sum $\mathcal{X} = \mathcal{X}_1 + \cdots + \mathcal{X}_L$ of $\mathcal{X}_\ell$. Indeed, $\mathcal{X}$ is the image of $\mathcal{X}_1 \times \cdots \times \mathcal{X}_L$ under the linear mapping $[x^1; \ldots; x^L] \mapsto x^1 + \cdots + x^L$, and taking direct products and linear images preserve spectratopes.

B. Processing covariance estimation problem in the diagonal case

We start with setting some additional notation to be used when operating with Euclidean space $\mathbf{S}^n$.

– We denote $\overline{n} = n(n+1)/2 = \dim \mathbf{S}^n$, $\mathcal{I} = \{(i,j) : 1 \leq i \leq j \leq n\}$, and for $(i,j) \in \mathcal{I}$ we set

$$e^{ij} = \begin{cases} e_i e_i^T, & i = j, \\ \frac{1}{\sqrt{2}}\left[e_i e_j^T + e_j e_i^T\right], & i < j, \end{cases}$$

where e_i are the standard basic orths in $\mathbf{R}^n$. Note that $\{e^{ij} : (i,j) \in \mathcal{I}\}$ is the standard orthonormal basis in $\mathbf{S}^n$. Given $v \in \mathbf{S}^n$, we denote by $x(v)$ the vector of coordinates of v in this basis:

$$x_{ij}(v) = \mathrm{Tr}(ve^{ij}) = \begin{cases} v_{ii}, & i = j, \\ \sqrt{2} v_{ij}, & i < j, \end{cases} \quad (i,j) \in \mathcal{I}.$$

Similarly, for $x \in \mathbf{R}^{\overline{n}}$, we index the entries in x by pairs ij, $(i,j) \in \mathcal{I}$, and set $v(x) = \sum_{(i,j)\in\mathcal{I}_p} x_{ij} e^{ij}$, so that $v \mapsto x(v)$ and $x \mapsto v(x)$ are inverse to each other linear isometries identifying the Euclidean spaces $\mathbf{S}^n$ and $\mathbf{R}^{\overline{n}}$ (recall that the inner products on these spaces are, respectively, the Frobenius and the standard one).

– Recall that $\mathcal{V}$ is the matrix box

$$\left\{v \in \mathbf{S}^n : v^2 \preceq I_n\right\} = \left\{v \in \mathbf{S}^n : \exists t \in \mathcal{T} := [0,1] : v^2 \preceq t I_n\right\}.$$

We denote by $\mathcal{X}$ the image of $\mathcal{V}$ under the mapping x:

$$\mathcal{X} = \left\{x \in \mathbf{R}^{\overline{n}} : \exists t \in [0,1] : v^2[x] \preceq t I_n\right\}.$$

Note that $\mathcal{X}$ is a basic spectratope.

Now we can assume that the signal underlying our observations is $x \in \mathcal{X}$, and the observations themselves are

$$w_t = x(\omega_t) = \underbrace{x\left(\tfrac{1}{2} v(x)\right)}_{=\frac{1}{2}x} + z_t, \quad z_t = x(\zeta_t).$$

Note that $z_t \in \mathbf{R}^{\bar{n}}$, $1 \le t \le T$, are zero mean i.i.d. random vectors with covariance matrix $D[x]$ satisfying, in view of (3.26), the relation

$$x \in \mathcal{X} \;\Rightarrow\; D[x] \preceq 2I_{\bar{n}}$$

(recall that we are in the case of $A = I_n$, $\vartheta_* = I_n$). Our goal is to estimate $B\vartheta[v]B^T = \frac{1}{2}BB^T + \frac{1}{2}BvB^T$, or, what is the same, to recover

$$\bar{B}x := \frac{1}{2}x\big(Bv(x)B^T\big).$$

Recall that we are in the situation where the norm in which the recovery error is measured is either the Frobenius, or the nuclear norm on $\mathbf{S}^n$; we "transfer" this norm from $\mathbf{S}^n$ to $\mathbf{R}^{\bar{n}}$. In the situation in question, Proposition 3.7 supplies the linear estimate

$$\widehat{x}\big(w^{(T)}\big) = \frac{1}{T}H_*^T\sum_{t=1}^{T} w_t$$

of $\bar{B}x$ with H_* stemming from the optimal solution to the convex optimization problem presented in Proposition 3.7. It is immediately seen that under the circumstances, this optimization problem reads:

in the Frobenius norm case:

$$\begin{aligned}\overline{\mathrm{Opt}} = \min_{\bar{H},\Lambda,\upsilon,\upsilon',\Theta}\Big\{&\mathrm{Tr}(\Lambda) + \upsilon + \upsilon' + \frac{2}{T}\mathrm{Tr}(\Theta) : \\ &\bar{H} \in \mathbf{R}^{\bar{n}\times\bar{n}},\ \Lambda \in \mathbf{S}^n_+,\ \upsilon \in \mathbf{R}_+,\ \upsilon' \in \mathbf{R}_+,\ \Theta \in \mathbf{S}^{\bar{n}}, \\ &\left[\begin{array}{c|c}\mathcal{R}^*[\Lambda] & \frac{1}{2}[\bar{B}^T - \frac{1}{2}\bar{H}] \\ \hline \frac{1}{2}[\bar{B} - \frac{1}{2}\bar{H}^T] & \upsilon I_{\bar{n}}\end{array}\right] \succeq 0, \\ &\left[\begin{array}{c|c}\Theta & \frac{1}{2}\bar{H} \\ \hline \frac{1}{2}\bar{H}^T & \upsilon' I_{\bar{n}}\end{array}\right] \succeq 0\Big\};\end{aligned} \tag{B.1}$$

in the nuclear norm case:

$$\begin{aligned}\overline{\mathrm{Opt}} = \min_{\bar{H},\Lambda,\Upsilon,\Upsilon',\Theta}\Big\{&\mathrm{Tr}(\Lambda) + \mathrm{Tr}(\Upsilon) + \mathrm{Tr}(\Upsilon') + \frac{2}{T}\mathrm{Tr}(\Theta) : \\ &\bar{H} \in \mathbf{R}^{\bar{n}\times\bar{n}}, \Lambda \in \mathbf{S}^n_+, \Upsilon \in \mathbf{S}^n_+, \Upsilon' \in \mathbf{S}^n_+, \Theta \in \mathbf{S}^{\bar{n}}, \\ &\left[\begin{array}{c|c}\mathcal{R}^*[\Lambda] & \frac{1}{2}[\bar{B}^T - \frac{1}{2}\bar{H}] \\ \hline \frac{1}{2}[\bar{B} - \frac{1}{2}\bar{H}^T] & \mathcal{R}^*[\Upsilon]\end{array}\right] \succeq 0, \\ &\left[\begin{array}{c|c}\Theta & \frac{1}{2}\bar{H} \\ \hline \frac{1}{2}\bar{H}^T & \mathcal{R}^*[\Upsilon']\end{array}\right] \succeq 0\Big\},\end{aligned} \tag{B.2}$$

where

$$\mathcal{R}^*[S] = \left[\mathrm{Tr}\left(e^{ij} S e^{k\ell}\right)\right]_{\substack{(i,j)\in\mathcal{I},\\ (k,\ell)\in\mathcal{I}}} \in \mathbf{S}^{\overline{n}}, \quad S \in \mathbf{S}^n.$$

So far, we did not use the fact that B is diagonal; this is what we intend to utilize now. Specifically, it is immediately seen that with diagonal B,

(1) The $\overline{n} \times \overline{n}$ matrix $\overline{B}$ is diagonal, with diagonal entries $\overline{B}_{ij,ij} = \frac{1}{2} B_{ii} B_{jj}$.

(2) Let $\mathcal{E}$ be the multiplicative group comprised of $n \times n$ diagonal matrices with diagonal entries ± 1. Every matrix $E \in \mathcal{E}$ induces diagonal $\overline{n} \times \overline{n}$ matrix F_E with diagonal entries ± 1 such that

$$\mathcal{R}^*[ESE] = F_E \mathcal{R}^*[S] F_E, \quad \forall \big(E \in \mathcal{E},\, S \in \mathbf{S}^n\big).$$

Indeed, when $E \in \mathcal{E}$ and $S \in \mathbf{S}^n$, we have

$$\begin{aligned}
\left[\mathcal{R}^*[ESE]\right]_{ij,k\ell} &= \mathrm{Tr}\left(e^{ij}[ESE]e^{k\ell}\right) \\
&= \mathrm{Tr}\left(S[Ee^{k\ell}E][Ee^{ij}E]\right) \\
&= \mathrm{Tr}\left(S[E_{ii}E_{jj}E_{kk}E_{\ell\ell}e^{k\ell}e^{ij}]\right) \\
&= [F_E]_{ij,ij}[F_E]_{k\ell,k\ell}\,\mathrm{Tr}\left(e^{ij} S e^{k\ell}\right) = [F_E \mathcal{R}^*[S] F_E]_{ij,k\ell},
\end{aligned}$$

where F_E is the diagonal $\overline{n} \times \overline{n}$ matrix with diagonal entries $[F_E]_{pq,pq} = E_{pp}E_{qq}$, $(p,q) \in \mathcal{I}$.

(3) When $S \in \mathbf{S}^n$ is diagonal, $\mathcal{R}^*[S]$ is diagonal as well.

Indeed, when S is diagonal and $(i,j) \in \mathcal{I}$, $(k,\ell) \in \mathcal{I}$ with $(i,j) \neq (k,\ell)$, the supports of matrices e^{ij} and $(Se^{k\ell})^T$ (the sets of cells where the entries of the respective matrices are nonzero) do not intersect, whence

$$\left[\mathcal{R}^*[S]\right]_{ij,k\ell} = \mathrm{Tr}\left(e^{ij} S e^{k\ell}\right) = \sum_{p,q} [e^{ij}]_{pq}\left[\left(Se^{k\ell}\right)^T\right]_{pq} = 0.$$

Observe that by (1)–(3) problems (B.1) and (B.2) have optimal solutions with diagonal matrix components. To see this, let us start with the nuclear norm case. Let $\overline{H}_*, \Lambda_*, \Upsilon_*, \Upsilon'_*, \Theta_*$ be an optimal solution to the problem. We claim that when $E \in \mathcal{E}$, the collection $F_E \overline{H}_* F_E, E\Lambda_* E, E\Upsilon_* E, E\Upsilon'_* E, F_E \Theta_* F_E$ is an optimal solution as well. Indeed, the values of the objective at the original and the transformed solution clearly are the same, so that all we need in order to justify our claim is to check that the transformed solution is feasible, which boils down to verifying that it satisfies the LMI constraints of the problem. We have:

$$
\begin{aligned}
&\left[\begin{array}{c|c}\mathcal{R}^*[E\Lambda_*E] & \frac{1}{2}\left[\bar{B}^T-\frac{1}{2}F_E\bar{H}_*F_E\right]\\ \hline \frac{1}{2}\left[\bar{B}-\frac{1}{2}[F_E\bar{H}_*F_E]^T\right] & \mathcal{R}^*[E\Upsilon_*E]\end{array}\right]\\
&\quad=\left[\begin{array}{c|c}F_E\mathcal{R}^*[\Lambda]F_E & \frac{1}{2}F_E\left[\bar{B}^T-\frac{1}{2}\bar{H}_*\right]F_E\\ \hline \frac{1}{2}F_E\left[\bar{B}-\frac{1}{2}\bar{H}_*^T\right]F_E & F_E\mathcal{R}^*[\Upsilon_*]F_E\end{array}\right]\\
&\quad=\mathrm{Diag}\{F_E,F_E\}\left[\begin{array}{c|c}\mathcal{R}^*[\Lambda] & \frac{1}{2}\left[\bar{B}^T-\frac{1}{2}\bar{H}_*\right]\\ \hline \frac{1}{2}\left[\bar{B}-\frac{1}{2}\bar{H}_*^T\right] & \mathcal{R}^*[\Upsilon_*]\end{array}\right]\mathrm{Diag}\{F_E,F_E\}\succeq 0,
\end{aligned}
$$

where the first equality is due to (3) combined with diagonality of $\bar{B}$ stated in (1). Similarly,

$$
\begin{aligned}
&\left[\begin{array}{c|c}F_E\Theta_*F_E & \frac{1}{2}F_E\bar{H}_*F_E\\ \hline \frac{1}{2}F_E\bar{H}_*^TF_E & \mathcal{R}^*[E\Upsilon'_*E]\end{array}\right]\\
&\qquad\qquad=\left[\begin{array}{c|c}F_E\Theta_*F_E & \frac{1}{2}F_E\bar{H}_*F_E\\ \hline \frac{1}{2}F_E\bar{H}_*^TF_E & F_E\mathcal{R}^*[\Upsilon'_*]F_E\end{array}\right]\\
&\qquad\qquad=\mathrm{Diag}\{F_E,F_E\}\left[\begin{array}{c|c}\Theta_* & \frac{1}{2}\bar{H}_*\\ \hline \frac{1}{2}\bar{H}_*^T & \mathcal{R}^*[\Upsilon'_*]\end{array}\right]\mathrm{Diag}\{F_E,F_E\}\succeq 0.
\end{aligned}
$$

Thus, the transformed solution indeed is feasible, as claimed. Since the problem is convex, the average, over $E\in\mathcal{E}$, of the above transformations of an optimal solution again is an optimal solution, let it be denoted $\bar{H}_\#,\Lambda_\#,\Upsilon_\#,\Upsilon'_\#,\Theta_\#$. By construction, $\Lambda_\#,\Upsilon_\#,\Upsilon'_\#$ are diagonal, whence by (3), $\mathcal{R}^*[\Lambda_\#],\mathcal{R}^*[\Upsilon_\#],\mathcal{R}^*[\Upsilon'_\#]$ are diagonal as well. This combines with diagonality of $\bar{B}$ to imply, similarly to the above, that if L is a diagonal $\bar{n}\times\bar{n}$ matrix with diagonal entries ± 1, the collection $L\bar{H}_\#L,\Lambda_\#,\Upsilon_\#,\Upsilon'_\#,L\Theta_\#L$ is an optimal solution to (B.2). Averaging these optimal solutions over L's, we conclude that the problem has an optimal solution comprised of diagonal matrices, as claimed. The same reasoning, with evident simplifications, works in the case of Frobenius norm.

We see that when solving (B.1) and (B.2), we lose nothing when restricting ourselves to candidate solutions with diagonal matrix components, which, by (1) and (3), automatically ensures the diagonality of blocks in the LMI constraints of the problem. As a result, we, first, reduce dramatically the design dimension of the problem, and, second, can now replace "large-scale" LMI constraints (which now state that some 2×2 block matrices with diagonal $\bar{n}\times\bar{n}$ blocks should be $\succeq 0$) with a bunch of small — just 2×2 — LMI's, thus making the problems easily solvable by the existing software, e.g. `CVX` [10], provided n is in the range of hundreds.

C. Conic duality

A conic problem is an optimization problem of the form

$$\mathrm{Opt}(P) = \max_x \left\{c^T x : A_i x - b_i \in \mathbf{K}_i,\ i = 1, \dots, m,\ Px = p\right\}, \tag{P}$$

where $\mathbf{K}_i$ are regular (i.e. closed, convex, pointed and with a nonempty interior) cones in Euclidean spaces E_i. Conic dual of (P) is "responsible" for upper-bounding the optimal value in (P) and is built as follows: selecting somehow *Lagrange multipliers* λ_i for the conic constraints $A_i x - b_i \in \mathbf{K}_i$ in the cones dual to $\mathbf{K}_i$:

$$\lambda_i \in \mathbf{K}_i^* := \left\{\lambda : \langle \lambda, y\rangle \geq 0,\ \forall y \in \mathbf{K}_i\right\},$$

and a Lagrange multiplier $\mu \in \mathbf{R}^{\dim p}$ for the equality constraints, every feasible solution x to (P) satisfies the linear inequalities $\langle \lambda_i, A_i x\rangle \geq \langle \lambda_i, b_i\rangle$, $i \leq m$, same as the inequality $\mu^T P x \geq \mu^T p$, and thus satisfies the aggregated inequality

$$\sum_i \langle \lambda_i, A_i x\rangle + \mu^T P x \geq \sum_i \langle \lambda_i, b_i\rangle + \mu^T p.$$

If the left hand side of this inequality is, *identically in* x, equal to $-c^T x$ (or, which is the same, $-c = \sum_i A_i^* \lambda_i + P^T \mu$, where A_i^* is the conjugate of A_i), the inequality produces an upper bound $-\langle \lambda_i, b_i\rangle - p^T \mu$ on $\mathrm{Opt}(P)$. The dual problem

$$\begin{aligned}\mathrm{Opt}(D) = \min_{\lambda_1,\dots,\lambda_m,\mu} \Big\{ &-\sum_i \langle \lambda_i, b_i\rangle - p^T \mu : \\ &\lambda_i \in \mathbf{K}_i^*,\ i \leq m,\ \sum_i A_i^* \lambda_i + P^T \mu = -c \Big\}\end{aligned} \tag{D}$$

is the problem of minimizing this upper bound. Note that (D) is a conic problem along with (P) — it is a problem of optimizing a linear objective under a bunch of linear equality constraints and conic inclusions of the form "affine function of the decision vector should belong to a given regular cone." Conic problem, like (P), is called *strictly feasible*, if it admits a feasible solution x for which all conic inclusions are satisfied strictly: $A_i x - b_i \in \mathrm{int}\, K_i$ for all i. Conic Duality Theorem (see, e.g. [2]) states that when one of the problems (P), (D) is bounded[13] and strictly feasible, then the other problem in the pair is solvable, and $\mathrm{Opt}(P) = \mathrm{Opt}(D)$.

References

[1] T. W. Anderson, The integral of a symmetric unimodal function over a symmetric convex set and some probability inequalities, *Proc. Amer. Math. Soc.*, **6** (1955), no. 2, 170–176. Zbl 0066.37402 MR 69229

[13] For a maximization (minimization) problem, boundedness means that the objective is bounded from above (resp., from below) on the feasible set.

[2] A. Ben-Tal and A. Nemirovski, *Lectures on modern convex optimization. Analysis, algorithms, and engineering applications*, MPS/SIAM Series on Optimization, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2001. Zbl 0986.90032 MR 1857264

[3] A. Buchholz, Operator Khintchine inequality in non-commutative probability, *Math. Ann.*, **319** (2001), no. 1, 1–16. Zbl 0991.46035 MR 1812816

[4] D. L. Donoho, Statistical estimation and optimal recovery, *Ann. Statist.*, **22** (1994), no. 1, 238–270. Zbl 0805.62014 MR 1272082

[5] D. L. Donoho, De-noising by soft-thresholding, *IEEE Trans. Inform. Theory*, **41** (1995), no. 3, 613–627. Zbl 0820.62002 MR 1331258

[6] D. L. Donoho and I. M. Johnstone, Minimax estimation via wavelet shrinkage, *Ann. Statist.*, **26** (1998), no. 3, 879–921. Zbl 0935.62041 MR 1635414

[7] D. L. Donoho, R. C. Liu, and B. MacGibbon, Minimax risk over hyperrectangles, and implications, *Ann. Statist.*, (1990), 1416–1437. Zbl 0705.62018 MR 1062717

[8] S. Efromovich, *Nonparametric curve estimation. Methods, theory, and applications*, Springer Series in Statistics, Springer-Verlag, New York, 1999. Zbl 0935.62039 MR 1705298

[9] S. Efromovich and M. Pinsker, Sharp-optimal and adaptive estimation for heteroscedastic nonparametric regression, *Statist. Sinica*, (1996), 925–942. Zbl 0857.62037 MR 1422411

[10] M. Grant and S. Boyd, *The* `CVX` *Users' Guide. Release 2.1*, 2014. Available at: `http://web.cvxr.com/cvx/doc/CVX.pdf`

[11] I. A. Ibragimov and R. Z. Khasminskii, *Statistical estimation. Asymptotic theory*, translated from the Russian by Samuel Kotz, Applications of Mathematics, 16, Springer-Verlag, Berlin, New York, 1981. Zbl 0467.62026 MR 620321

[12] I. A. Ibragimov and R. Z. Khas'minskii, Estimation of linear functionals in gaussian noise, *Theory Probab. Appl.*, **32** (1988), no. 1, 30–39.

[13] A. Juditsky and A. Nemirovski, Near-optimality of linear recovery in gaussian observation scheme under $\|\cdot\|_2^2$-loss, to appear in *Ann. Statist.* arXiv:1602.01355

[14] A. B. Juditsky and A. S. Nemirovski, Nonparametric estimation by convex programming, *Ann. Statist.*, **37** (2009), no. 5A, 2278–2300. Zbl 1173.62024 MR 2543692

[15] J. A. Kuks and W. Olman, Linear minimax estimation of regression coefficients. II (Russian), *Izv. Akad. Nauk Ehst. SSR, Fiz. Mat.*, **20** (1971), 480–482. Zbl 0282.62054 MR 319298

[16] J. A. Kuks and W. Olman, Linear minimax estimation of regression coefficients. I (Russian), *Izv. Akad. Nauk Ehst. SSR, Fiz. Mat.*, **21** (1972), 66–72. Zbl 0227.62019 MR 319297

[17] F. Lust-Piquard, Inégalités de Khintchine dans C_p $(1 < p < \infty)$, *CR Acad. Sci. Paris*, **303** (1986), 289–292. Zbl 0592.47038 MR 859804

[18] L. Mackey, M. I. Jordan, R. Y. Chen, B. Farrell, and J. A. Tropp, Matrix concentration inequalities via the method of exchangeable pairs, *Ann. Probab.*, **42** (2014), no. 3, 906–945. Zbl 1294.60008 MR 3189061

[19] C. Micchelli and T. Rivlin, Lectures on optimal recovery, in *Numerical analysis (Lancaster, 1984)*, 21–93, Lecture Notes in Mathematics, 1129, Springer, 1985. Zbl 0698.41024 MR 799030

[20] C. A. Micchelli and T. J. Rivlin, A survey of optimal recovery, in *Optimal estimation in approximation theory (Freudenstadt, 1976)*, 1–54, Proc. Internat. Sympos., Plenum, New York, 1977. Zbl 0386.93045 MR 617931

[21] A. Nemirovski, Sums of random symmetric matrices and quadratic optimization under orthogonality constraints, *Math. Program. Ser. B*, **109** (2007), no. 2-3, 283–317. Zbl 1156.90012 MR 2295145

[22] A. Nemirovski, C. Roos, and T. Terlaky, On maximization of a quadratic form over intersection of ellipsoids with common center, *Math. Program. Ser. A*, **86** (1999), no. 3, 463–473. Zbl 0944.90056 MR 1733748

[23] A. Nemirovskii, Nonparametric estimation of smooth regression functions, *J. Comput. Sysr. Sci.*, **23** (1986), no. 6, 1–11. MR 844292

[24] M. Pinsker, Optimal filtration of square-integrable signals in gaussian noise, *Prob. Info. Transmission*, **16** (1980), no. 2, 120–133. MR 624591

[25] G. Pisier, Non-commutative vector valued L_p-spaces and completely p-summing maps, *Astérisque*, **247** (1998), vi+131pp. Zbl 0937.46056 MR 1648908

[26] A. M.-C. So, Moment inequalities for sums of random matrices and their applications in optimization, *Math. Program. Ser. A*, **130** (2011), no. 1, 125–151. Zbl 1231.60007 MR 2853163

[27] J. F. Traub, G. W. Wasilkowski, and H. Woźniakowski, *Information, uncertainty, complexity*, Addison-Wesley Publishing Company, Advanced Book Program/World Science Division, 1983. Zbl 0522.68041 MR 680041

[28] J. F. Traub, G. W. Wasilkowski, and H. Woźniakowski, *Information-based complexity*, Academic Press, Boston, San Diego, and New York, 1988. Zbl 0654.94004 MR 958691

[29] J. A. Tropp, An introduction to matrix concentration inequalities, *Found. Trends Mach. Learn.*, **8** (2015), no. 1-2, 1–230. Zbl 06468266

[30] A. B. Tsybakov, *Introduction to nonparametric estimation*, revised and extended from the 2004 French original, translated by Vladimir Zaiats, Springer Series in Statistics. Springer, New York, 2009. Zbl 1176.62032 MR 2724359

[31] L. Wasserman, *All of nonparametric statistics*, Springer Texts in Statistics, Springer, New York, 2006. Zbl 1099.62029 MR 2172729

Received 24 May, 2017; revised 04 February, 2018

A. Juditsky, Laboratoire Jean Kuntzmann, Université Grenoble-Alpes,
700 Avenue Centrale, 38401 Domaine Universitaire de Saint-Martin-d'Hères, France

E-mail: anatoli.juditsky@univ-grenoble-alpes.fr

A. Nemirovski, H. Milton Stewart School of Industrial and Systems Engineering,
Georgia Institute of Technology, Atlanta, GA 30332, USA

E-mail: nemirovs@isye.gatech.edu