Near-Optimal Linear Recovery from Indirect Observations

joint work with

A. Nemirovski, Georgia Tech

http://www2.isye.gatech.edu/~ nemirovs/StatOpt_LN.pdf

Les Houches, April, 2017

Situation: "In the nature" there exists a signal x known to belong to a given convex compact set $\mathcal{X} \subset \mathbb{R}^n$. We observe corrupted by noise affine image of the signal:

$$\omega = Ax + \sigma\xi \in \Omega = \mathbb{R}^m$$

- A: given $m \times n$ sensing matrix
- ξ: random observation noise

• **Our goal** is to recover the image Bx of x under a given affine mapping B: $\mathbb{R}^n \to \mathbb{R}^\nu$.

• **Risk** of a candidate estimate $\widehat{x}(\cdot) : \Omega \to \mathbb{R}^\nu$ is defined as

$$\text{Risk}[\widehat{x}|\mathcal{X}] = \sup_{x\in\mathcal{X}} \sqrt{\mathbf{E}_\xi \left\{\|Bx - \widehat{x}(Ax + \sigma\xi)\|_2^2\right\}}$$

$\Rightarrow$ Risk^2 is the worst-case, over $x \in \mathcal{X}$, expected $\|\cdot\|_2^2$ recovery error.

Agenda: Under appropriate assumptions on $\mathcal{X}$, we are to show that

- *One can build, in a computationally efficient fashion, the (nearly) best, in terms of risk, estimate from the family of linear estimates*

$$\widehat{x}(\omega) = \widehat{x}_H(\omega) = H^T\omega \qquad [H \in \mathbb{R}^{m\times\nu}]$$

- *The resulting linear estimate is nearly optimal among all estimates, linear and nonlinear alike.*

Linear estimation of signal in Gaussian noise

- ...
- Kuks & Olman, 1971, 1972
- Rao 1972, 1973, Pilz, 1981, 1986, ..., Drygas, 1996, Christopeit & Helmes, 1996, Arnold & Stahlecker, 2000, ...
- Pinsker 1980, Efromovich & Pinsker, 1981, 1982, Efromovich & Pinsker 1996, Golubev, Levit & Tsybakov, 1996, ..., Efromovich, 2008, ...
- Donoho, Liu, McGibbon, 1990
- ...

Risk of linear estimation

Assuming that ξ *is zero mean with unit covariance matrix,* we can easily compute the risk of *a linear estimate* $\widehat{x}_H(\omega) = H^T\omega$

$$\begin{array}{rcl}
\text{Risk}^2[\widehat{x}_H|\mathcal{X}] &=& \max\limits_{x\in\mathcal{X}} \mathbf{E}_\xi\left\{\|[B-H^TA]x-\sigma H^T\xi\|_2^2\right\}\\
&=& \max\limits_{x\in\mathcal{X}} \left\{\|[B-H^TA]x\|_2^2+\sigma^2\mathbf{E}_\xi\{\text{Tr}(H^T\xi\xi^TH\}\right\}\\
&=& \sigma^2\text{Tr}(H^TH)+\max\limits_{x\in\mathcal{X}}\text{Tr}(xx^T[B^T-A^TH][B-H^TA]).
\end{array}$$

Note: *building the minimum risk linear estimate reduces to solving convex minimization problem*

$$\min_H \left[\phi(H) := \max_{x\in\mathcal{X}} \text{Tr}(xx^T[B^T-A^TH][B-H^TA]) + \sigma^2\text{Tr}(H^TH)\right]. \qquad (*)$$

Convex function ϕ *is given implicitly and can be difficult to compute, making* $(*)$ *difficult as well.*

Fact: essentially, the only cases when $(*)$ is known to be easy are those when

- $\mathcal{X}$ is given as a convex hull of finite set of moderate cardinality
- $\mathcal{X}$ is an ellipsoid: for $W \in \mathbf{S}^n$ and $S \succ 0$

$$\max_{x^TSx\leq 1} \mathrm{Tr}(xx^TW) = \lambda_{\max}\left(S^{-1/2}WS^{-1/2}\right).$$

where $\lambda_{\max}(\cdot)$ is the maximal eigenvalue.

When $\mathcal{X}$ is a "box," computing ϕ is NP-hard...

• When ϕ is difficult to compute, we can to replace ϕ in the design problem $(*)$ with an efficiently computable convex upper bound $\varphi(H)$.

• We are about to consider a family of sets $\mathcal{X}$ – *ellitopes* – for which reasonably tight bounds φ of desired type are available.

An ellitope is a set $\mathcal{X} \subset \mathbb{R}^n$ given as

$$\mathcal{X} = \{x \in \mathbb{R}^n : \exists y \in \mathbb{R}^N, t \in \mathcal{T} : x = Py, \ y^T S_k y \le t_k, \ 1 \le k \le K\}$$

where

- P is a given $n \times N$ matrix (we can assume that $P = I_n$),
- $S_k \succeq 0$ are positive semidefinite matrices with $\sum_k S_k \succ 0$
- $\mathcal{T}$ is a convex compact subset of K-dimensional nonnegative orthant $\mathbb{R}^K_+$ such that
  - $\mathcal{T}$ contains some positive vectors
  - $\mathcal{T}$ is *monotone*: if $0 \le t' \le t$ and $t \in \mathcal{T}$, then $t' \in \mathcal{T}$ as well.

Note: *every ellitope is a symmetric w.r.t. the origin convex compact set.*

Examples

[A.] A centered at the origin ellipsoid $(K = 1,\ \mathcal{T} = [0; 1])$

[B.] (Bounded) intersection of K ellispoids/elliptic cylinders centered at the origin $(\mathcal{T} = \{t \in \mathbb{R}^K : 0 \leq t_k \leq 1,\ k \leq N\})$

[C.] Box $\{x \in \mathbb{R}^n : -1 \leq x_i \leq 1\}$ $(\mathcal{T} = \{t \in \mathbb{R}^n : 0 \leq t_k \leq 1,\ k \leq K = n\},\ x^T S_k x = x_k^2)$

[D.] $\mathcal{X} = \{x \in \mathbb{R}^n : \|x\|_p \leq 1\}$ with $p \geq 2$ $(\mathcal{T} = \{t \in \mathbb{R}^n_+ : \|t\|_{p/2} \leq 1\},\ x^T S_k x = x_k^2,\ k \leq K = n)$

Ellitopes admit fully algorithmic calculus: if $\mathcal{X}_i$, $1 \leq i \leq I$, are ellitopes, so are

- linear images of $\mathcal{X}_i$
- inverse linear images of $\mathcal{X}_i$ under linear embeddings
- $\bigcap_i \mathcal{X}_i$
- $\mathcal{X}_1 \times ... \times \mathcal{X}_I$
- $\mathcal{X}_1 + ... + \mathcal{X}_I$
- ...

Observation

Let

$$\mathcal{X} = \{x \in \mathbb{R}^n : \exists t \in \mathcal{T} : x^T S_k x \le t_k , 1 \le k \le K\}$$

be an ellitope. Given a quadratic form $x^T W x$, $W \in \mathbf{S}^n$, one has

$$\max_{x \in \mathcal{X}} x^T W x = \max_{x \in \mathcal{X}} \mathrm{Tr}(x x^T W) \le \max_{Q \in \mathcal{Q}} \mathrm{Tr}(QW),$$

where

$$\mathcal{Q} := \{Q \in \mathbf{S}^n : Q \succeq 0, \exists t \in \mathcal{T}: \mathrm{Tr}(QS_k) \le t_k, k \le K\}.$$

We conclude that

$$\phi(H) \le \varphi(H) := \sigma^2 \mathrm{Tr}(H^T H) + \max_{Q \in \mathcal{Q}} \mathrm{Tr}\big(Q(A^T H - B^T)(H^T A - B)\big),$$

and

$$\mathrm{Risk}^2[\widehat{x}_H | \mathcal{X}] \le \min_H \varphi(H).$$

This attracts our attention to the optimization problem

$$\mathrm{Opt}^P = \min_H \left\{\varphi(H) = \max_{Q\in\mathcal{Q}} \left[\underbrace{\sigma^2\mathrm{Tr}(H^TH) + \mathrm{Tr}\left(Q(A^TH - B^T)(H^TA - B)\right)}_{\Phi(H,Q)}\right]\right\}. \quad (P)$$

Note that (P) is the primal problem

$$\min_H \left[\max_{Q\in\mathcal{Q}} \Phi(H,Q)\right]$$

associated with the convex-concave saddle point function $\Phi(H,Q)$. The dual problem associated with $\Phi(H,Q)$ is

$$\max_{Q\in\mathcal{Q}} \left[\min_H \Phi(H,Q)\right],$$

that is, the problem

$$\mathrm{Opt}^D = \max_{Q\in\mathcal{Q}} \left\{\psi(Q) := \min_H \left[\sigma^2\mathrm{Tr}(H^TH) + \mathrm{Tr}\left(Q(A^TH - B^T)(H^TA - B)\right)\right]\right\}. \quad (D)$$

By the Sion-Kakutani theorem, (P) and (D) are solvable with equal optimal values: $\mathrm{Opt}^D = \mathrm{Opt}^P = \mathrm{Opt}$.

Note that the minimizer of $\Phi(\cdot, Q)$ can be easily computed:

$$H(Q) = (\sigma^2 I_m + AQA^T)^{-1} AQB^T,$$

so that

$$\psi(Q) = \mathrm{Tr}\big(B[Q - QA^T(\sigma^2 I_m + AQA^T)^{-1} AQ]B^T\big),$$

and the dual problem reads

$$\begin{array}{rl}\mathrm{Opt}^D = \max\limits_{Q,t} \Big\{ \mathrm{Tr}\big(B[Q - QA^T(\sigma^2 I_m + AQA^T)^{-1} AQ]B^T\big), & \\ Q \succeq 0,\, t \in \mathcal{T},\, \mathrm{Tr}(QS_k) \leq t_k,\, k \leq K \Big\} & \end{array} \qquad (D)$$

In fact, both (P) and (D) can be cast as Semidefinite Optimization problems. In particular, (P) can be rewritten as

$$\mathrm{Opt} = \min_{H,\lambda} \left\{ \sigma^2 \mathrm{Tr}(H^T H) + \phi_{\mathcal{T}}(\lambda) : \left[\begin{array}{cc} \sum_k \lambda_k S_k & B^T - A^T H \\ B - H^T A & I_\nu \end{array} \right] \succeq 0, \lambda \geq 0 \right\} \qquad (P)$$

where $\phi_{\mathcal{T}} : \mathbb{R}^K \to \mathbb{R}$ is the support function of $\mathcal{T}$:

$$\phi_{\mathcal{T}}(\lambda) = \max_{t \in \mathcal{T}} \lambda^T t.$$

Note that (P) is efficiently solvable whenever $\mathcal{T}$ is computationally tractable.

Bottom line: *Given matrices* $A \in \mathbb{R}^{m\times n}$, $B \in \mathbb{R}^{k\times n}$ *and an ellitope*

$$\mathcal{X} = \{x \in \mathbb{R}^n : \exists t \in \mathcal{T} : \; x^T S_k x \le t_k \,, 1 \le k \le K\} \qquad (*)$$

consider the convex optimization problems

$$\mathrm{Opt}^P = \min_H \varphi(H) \quad \textit{and} \quad \mathrm{Opt}^D = \max_{Q \in \mathcal{Q}} \psi(Q),$$

where $\mathcal{Q} := \{Q \in \mathbf{S}^n : \; Q \succeq 0, \exists t \in : \mathrm{Tr}(QS_k) \le t_k, \, k \le K\}$.

- *The optimal values of two problems coincide,* $\mathrm{Opt}^P = \mathrm{Opt}^D = \mathrm{Opt}$.
- *When noise* ξ *satisfies* $\mathbf{E}\{\xi\} = 0$, *and* $\mathbf{E}\{\xi\xi^T\} = I_m$, *the risk of the linear estimate* $\widehat{x}_{H_*}(\cdot)$ *induced by the optimal solution* H_* *to the problem (this solution clearly exists provided that* $\sigma > 0$*) satisfies the risk bound*

$$\mathrm{Risk}[\widehat{x}_{H_*}|\mathcal{X}] \le \sqrt{\mathrm{Opt}}.$$

- We are to compare the bound $\sqrt{\mathrm{Opt}}$ for the risk of $\widehat{x}_{H_*}$ to the minimax risk

$$\mathrm{Risk}_{\mathrm{Opt}}[\mathcal{X}] = \inf_{\widehat{x}(\cdot)} \mathrm{Risk}[\widehat{x}|\mathcal{X}].$$

Bayesian risks

• *Minimax risk* $\mathrm{Risk}_{\mathrm{Opt}}[\mathcal{X}]$ is defined as the *worst*, over the signals of interest, performance of $\widehat{x}(\cdot)$

• *Bayesian risk* is the *average performance*, with the average taken over some *prior* probability distribution on the signals.

For the problem of $\|\cdot\|_2$-recovering Bx via noisy observation

$$\omega = Ax + \sigma\xi,\ \xi \sim P$$

this alternative reads as follows:

(!) Given a probability distribution π of signals $x \in \mathbb{R}^n$, find an estimate $\widehat{x}(\cdot)$ which minimizes

$$\mathrm{Risk}^2(\widehat{x}|\pi) := \int_\pi \left\{ \int_{\mathbb{R}^m} \|Bx - \widehat{x}(Ax + \sigma\xi)\|_2^2 P(d\xi) \right\} \pi(dx)$$

– the average, over the distribution π of signals x, of expected $\|\cdot\|_2^2$ estimation error of Bx via observation $Ax + \sigma\xi$.

Let $P_{x,\omega}$ be the induced by π and P_ξ *joint distribution* of $(x, \omega = Ax + \sigma\xi)$ on $\mathbb{R}^n_x \times \mathbb{R}^m_\omega$. $P_{x,\omega}$ gives rise to

- *marginal distribution* P_ω *of* ω,
- *conditional distribution* $P_{x|\omega}$ *of* x *given* ω.

We have

$$\begin{array}{rcl} \mathrm{Risk}^2(\widehat{x}|\pi) & = & \int\limits_{\mathbb{R}^n_x \times \mathbb{R}^m_\omega} \|Bx - \widehat{x}(\omega)\|_2^2 P_{x,\omega}(dx, d\omega) \\ & = & \int_{\mathbb{R}^m} \left\{ \int_{\mathbb{R}^n} \|Bx - \widehat{x}(\omega)\|_2^2 P_{x|\omega}(dx) \right\} P_\omega(d\omega) \end{array}$$

Assuming that the probability distribution π possesses finite second moments, one has

$$\min_{\widehat{x}(\cdot)} \int\limits_{\mathbb{R}^n \times \mathbb{R}^m} \|Bx - \widehat{x}(\omega)\|_2^2 P_{x,\omega}(dx) = \int\limits_{\mathbb{R}^n \times \mathbb{R}^m} \|Bx - \widehat{x}_*(\omega)\|_2^2 P_{x,\omega}(dx),$$

where

$$\widehat{x}_*(\omega) = \int_{\mathbb{R}^n} Bx P_{x|\omega}(dx).$$

Corollary [Gauss-Markov theorem]: *Let $x \in \mathbb{R}^n$ and $\xi \in \mathbb{R}^m$ be independent zero-mean Gaussian random vectors. Assuming $\sigma > 0$ and the covariance matrix of ξ to be positive definite,*

- *conditional, given ω, distribution of x is normal, and the conditional expectation $\widehat{x}_*(\omega)$ is a linear function of ω,*
- *as a result, an optimal solution $\widehat{x}_*(\cdot)$ to the risk minimization problem*

$$\min_{\widehat{x}(\cdot)} \mathbf{E}_{x\sim\pi,\xi} \left\{ \|Bx - \widehat{x}(Ax + \sigma\xi)\|_2^2 \right\}$$

exists and is a linear function of $\omega = Ax + \sigma\xi$.

In particular, when $\xi \sim \mathcal{N}(0, I_m)$ and $x \sim \mathcal{N}(0, Q)$, one has

$$\begin{array}{rcl} \widehat{x}_*(\omega) & = & \left[[\sigma^2 I_m + AQA^T)]^{-1} AQB^T \right] \omega \\ \mathrm{Risk}^2(\widehat{x}_* | \mathcal{N}(0,Q)) & = & \mathrm{Tr}\big(B[Q - QA^T[\sigma^2 I_m + AQA^T]^{-1} AQ]B^T \big) \end{array}$$

Course of actions (Pinsker's program)

• Let $\mathcal{N}(0,Q)$ be a Gaussian prior for the signal x which "sits on $\mathcal{X}$ with high probability." Then by the Gauss-Markov theorem the ("slightly reduced") quantity

$$\psi(Q) = \mathrm{Tr}(B[Q - QA^T[\sigma^2 I_m + AQA^T]^{-1}AQ]B^T)$$

would be a lower bound on $\mathrm{Risk}^2_{\mathrm{Opt}}$.

• Note that $\mathbf{E}_{\eta\sim\mathcal{N}(0,Q)}\{\eta^T S\eta\} = \mathrm{Tr}(SQ)$. Thus, *selecting* $Q \succeq 0$ *according to*

$$\exists t \in \mathcal{T} : \mathrm{Tr}(QS_k) \leq t_k, k \leq K$$

we ensure that $\eta \sim \mathcal{N}(0,Q)$ *sits in* $\mathcal{X}$ *"on average."*

Imposing on $Q \succeq 0$ *restriction*

$$\exists t \in \mathcal{T} : \mathrm{Tr}(QS_k) \leq \rho t_k, k \leq K, \qquad [\rho > 0]$$

we enforce $\eta \sim \mathcal{N}(0,Q)$ *to take values in* $\mathcal{X}$ *with probability controlled by* ρ *and approaching 1 as* $\rho \to +0$.

• The above considerations give rise to parametric optimization problem

$$\mathrm{Opt}_*(\rho) = \max_{Q\succeq 0} \{\psi(Q) : \exists t \in \mathcal{T} : \mathrm{Tr}(QS_k) \le \rho t_k,\ 1 \le k \le K\} \qquad (P_\rho)$$

We *may expect* that for small ρ a "slightly corrected" $\mathrm{Opt}_*(\rho)$ is a lower bound on $\mathrm{Risk}^2_{\mathrm{Opt}}$.

• As we have just seen, $\mathrm{Opt}_*(1) = \mathrm{Opt}$ (!). Since the optimal value of the (concave) optimization problem (P_ρ) is a a concave function of ρ, we have

$$\mathrm{Opt}_*(\rho) \ge \rho\mathrm{Opt},\ \ 0 < \rho < 1.$$

Now, all we need is a simple result as follows:

Lemma *Let S and Q be positive semidefinite $n\times n$ matrices with $\rho := \mathrm{Tr}(SQ) \le 1$, and let $\eta \sim \mathcal{N}(0,Q)$. Then*

$$\mathrm{Prob}\left\{\eta^T S\eta > 1\right\} \le \mathrm{e}^{-\frac{1-\rho+\rho\ln(\rho)}{2\rho}}$$

We arrive at the following

Theorem. *Let us associate with ellitope* $\mathcal{X} = \{x \in \mathbb{R}^n : \exists t \in \mathcal{T} : x^T S_k x \leq t_k,\ k \leq K\}$ *the convex compact set*

$$\mathcal{Q} = \{Q \in \mathbf{S}^n : Q \succeq 0, \exists t \in \mathcal{T} : \mathrm{Tr}(QS_k) \leq t_k, k \leq K\},$$

and the quantity

$$M_* = \max_{Q \in \mathcal{Q}} \sqrt{\mathrm{Tr}(BQB^T)}.$$

Then the linear estimate $\widehat{x}_{H_*}(\omega) = H_*^T \omega$ *of* Bx, $x \in \mathcal{X}$, *via observation* $\omega = Ax + \sigma\xi$, $\xi \sim \mathcal{N}(0, I_m)$, *given by the optimal solution* H_* *to the convex optimization problem*

$$\mathrm{Opt} = \min_{H,\lambda} \left\{ \phi_{\mathcal{T}}(\lambda) + \sigma^2 \mathrm{Tr}(HH^T) : \begin{array}{l} \lambda \geq 0 \\ \left[\begin{array}{c|c} \sum_k \lambda_k S_k & B^T - A^T H \\ \hline B - H^T A & I_k \end{array}\right] \succeq 0 \end{array} \right\}$$

satisfies the risk bound

$$\mathrm{Risk}[\widehat{x}_{H_*}|\mathcal{X}] \leq \sqrt{\mathrm{Opt}} \leq \sqrt{6 \ln\left(\frac{8M_*^2 K}{\mathrm{Risk}^2_{\mathrm{Opt}}[\mathcal{X}]}\right)} \mathrm{Risk}_{\mathrm{Opt}}[\mathcal{X}].$$

Numerical illustration

In these experiments

- B is $n \times n$ identity matrix,
- $n \times n$ sensing matrix A is a randomly rotated matrix with singular values λ_j, $1 \leq j \leq n$, forming a geometric progression, with $\lambda_1 = 1$ and $\lambda_n = 0.01$.

- In the first experiment the signal set $\mathcal{X}_1$ is an ellipsoid:

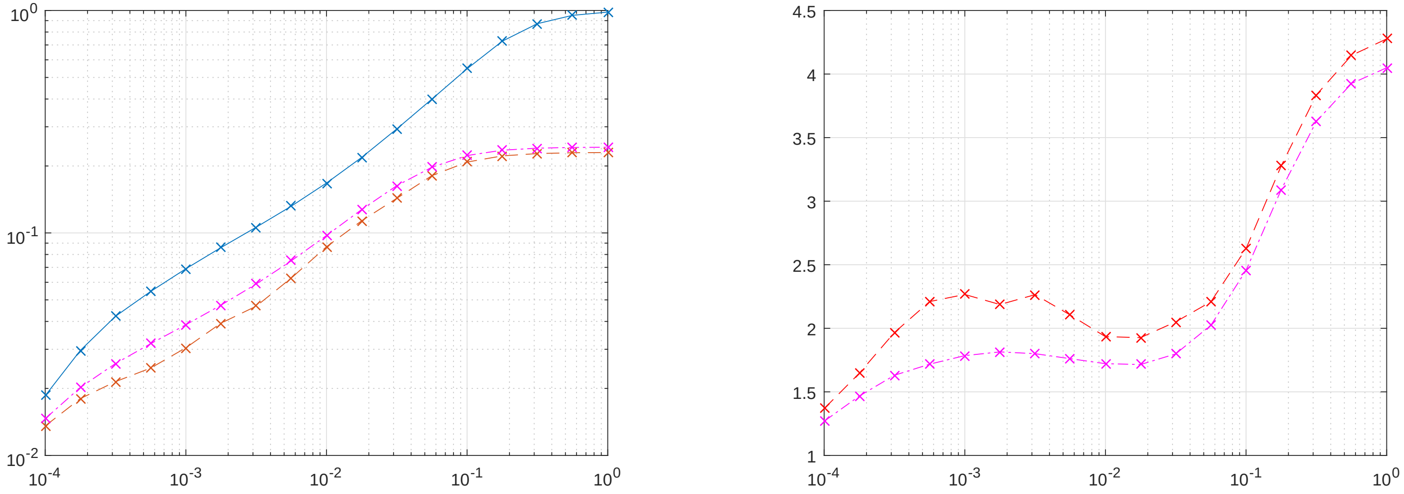
$$\mathcal{X}_1 = \{x \in \mathbb{R}^n : \sum_{j=1}^{n} j^2 x_j^2 \leq 1\},$$

that is, $K = 1$, $S_1 = \sum_{j=1}^{n} j^2 e_j e_j^T$ (e_j are basic orths), and $\mathcal{T} = [0, 1]$.
Theoretical "suboptimality factor" in the interval [31.6, 73.7] *in this experiment.*

- In the second experiment, the signal set $\mathcal{X}$ is the box:

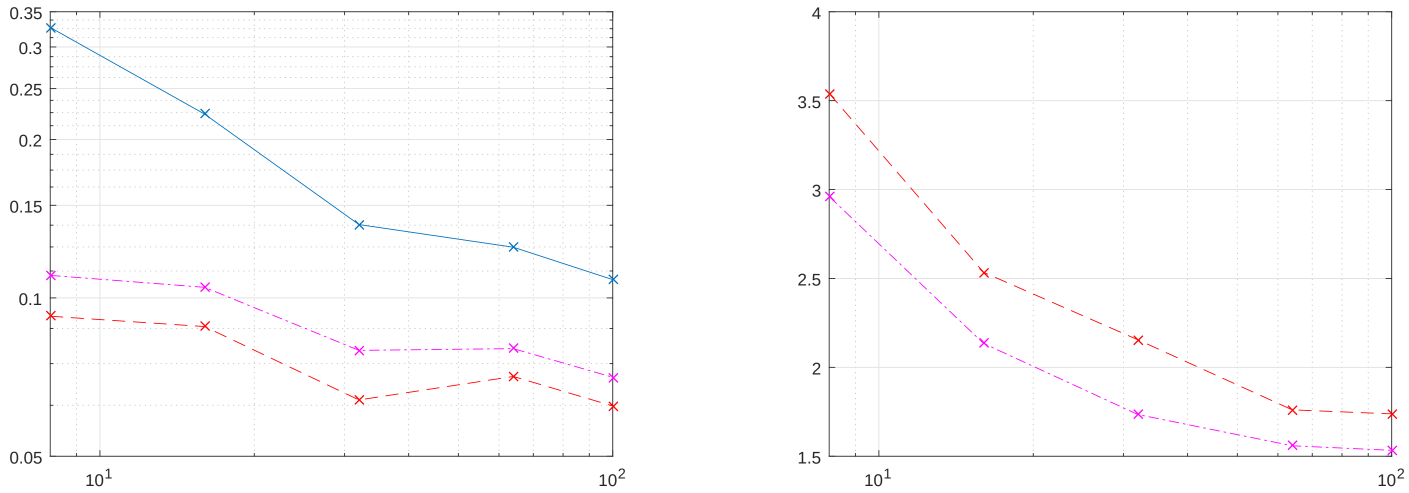
$$\mathcal{X} = \{x \in \mathbb{R}^n : j|x_j| \leq 1,\ 1 \leq j \leq n\} \quad [K = n, S_k = k^2 e_k e_k^T, k = 1, ..., K, \mathcal{T} = [0, 1]^K].$$

Theoretical "suboptimality factor" in the interval [73.2, 115.4].



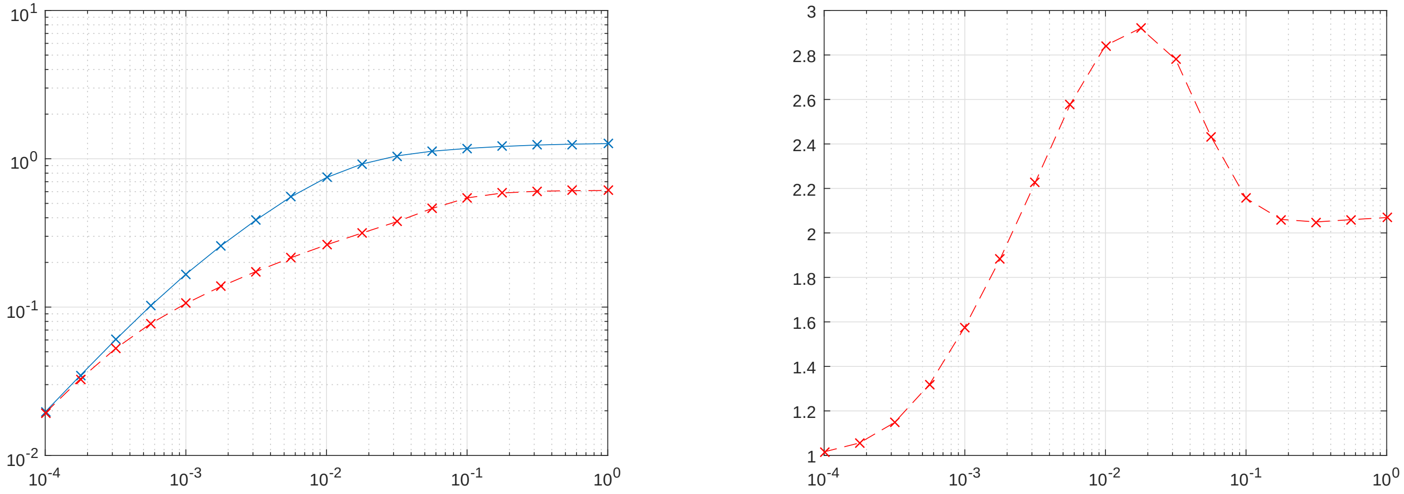
Recovery on ellipsoids: risk bounds as functions of the noise level σ, dimension $n = 32$.

Left plot: upper and lower bounds of the risk; right plot: suboptimality ratios.



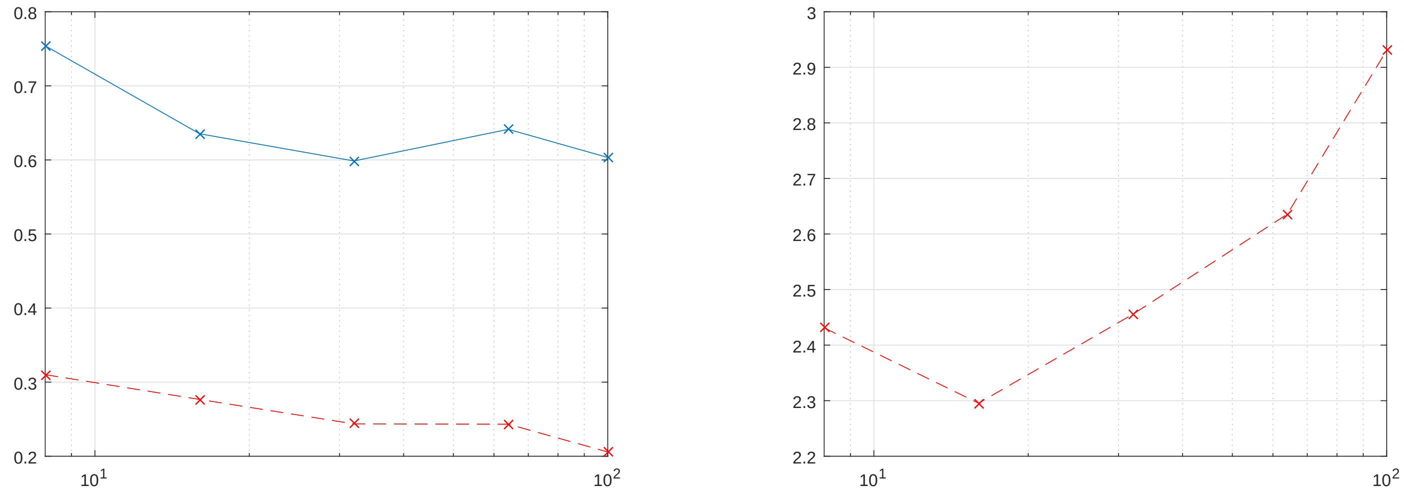
Recovery on ellipsoids: risk bounds as functions of problem dimension n, noise level $\sigma = 0.01$.

Left plot: upper and lower risk bounds; right plot: suboptimality ratios.



Recovery on a box: risk bounds as functions of the noise level σ, dimension $n = 32$.

Left plot: upper and lower bounds of the risk; right plot: suboptimality ratios.



Recovery on a box: risk bounds as functions of problem dimension n, noise level $\sigma = 0.01$.

Left plot: upper and lower risk bounds; right plot: suboptimality ratios.

Extensions

1. Relative risks

When "very large" signals are allowed, one may switch from the usual risk to its *relative version* – "S-risk" defined as follows:

- Given a positive semidefinite "risk calibrating matrix" S we set

$$\mathrm{RiskS}[\widehat{x}|\mathcal{X}] = \min\left\{\sqrt{\tau} : \mathbf{E}_{\xi}\left\{\|Bx - \widehat{x}(Ax + \sigma\xi)\|_2^2\right\} \leq \tau[1 + x^T Sx]\,\forall x \in \mathcal{X}\right\}$$

Note: setting $S = 0$ recovers the usual "plain" risk.

- Results on design of near-optimal, in terms of plain risk, linear estimates extend directly to the case of S-risk.

Design of near optimal linear estimate $\widehat{x}_{H_*}(\omega) = H_*^T\omega$ is given by an optimal solution (H_*, τ_*, λ_*) to the convex optimization problem

$$\mathrm{Opt} = \min_{H,\tau,\lambda}\left\{\tau : \left[\begin{array}{c|c}\sum_k \lambda_k S_k + \tau S & B^T - A^T H \\ \hline B - H^T A & I_k\end{array}\right] \succeq 0,\ \sigma^2 \mathrm{Tr}(HH^T) + \phi_{\mathcal{T}}(\lambda) \leq \tau,\ \lambda \geq 0\right\}$$

For the resulting estimate, it holds

$$\mathrm{RiskS}[\widehat{x}_{H_*}|\mathcal{X}] \leq \sqrt{\mathrm{Opt}},$$

provided ξ is zero mean with unit covariance matrix.

Near-optimality properties of the estimate $\widehat{x}_{H_*}$ remain the same as in the case of plain risk: *when $\xi \sim \mathcal{N}(0, I_m)$, one has*

$$\mathrm{RiskS}[\widehat{x}_{H_*}|\mathcal{X}] \leq \sqrt{6\ln\left(\frac{8KM_*^2}{\mathrm{RiskS}^2_{\mathrm{Opt}}[\mathcal{X}]}\right)}\mathrm{RiskS}_{\mathrm{Opt}}[\mathcal{X}],$$

where

$$M_* = \max_Q\left\{\sqrt{\mathrm{Tr}(BQB^T)} : Q \succeq 0, \exists t \in \mathcal{T} : \mathrm{Tr}(QS_k) \leq t_k,\ 1 \leq k \leq K\right\},$$

and

$$\mathrm{RiskS}_{\mathrm{Opt}}[\mathcal{X}] = \inf_{\widehat{x}(\cdot)} \mathrm{RiskS}[\widehat{x}|\mathcal{X}].$$

In the case $\mathcal{X} = \mathbb{R}^n$, the best linear estimate is yielded by the optimal solution to the convex problem

$$\text{Opt} = \min_{H,\tau} \left\{ \tau : \left[\begin{array}{c|c} \tau S & B^T - A^T H \\ \hline B - H^T A & I_k \end{array} \right] \succeq 0, \ \sigma^2 \text{Tr}(HH^T) \leq \tau \right\} \qquad (*)$$

A feasible solution τ, H to $(*)$ gives rise to linear estimate $\widehat{x}_H(\omega) = H^T \omega$ such that

$$\text{RiskS}[\widehat{x}_H | \mathbb{R}^n] \leq \sqrt{\tau},$$

provided ξ is zero mean with unit covariance matrix.

Proposition *Assume that $B \neq 0$ and $(*)$ is feasible. Then the problem is solvable, and its optimal solution* Opt, H_* *gives rise to the linear estimate*

$$\widehat{x}_{H_*}(\omega) = H_*^T \omega$$

with S-risk $\sqrt{\text{Opt}}$.

When $\xi \sim \mathcal{N}(0, I_m)$, this estimate is minimax optimal:

$$\text{RiskS}[\widehat{x}_{H_*} | \mathbb{R}^n] = \sqrt{\text{Opt}} = \text{RiskS}_{\text{Opt}}[\mathbb{R}^n].$$

2. Spectratopes

We say that a set $\mathcal{X} \subset \mathbb{R}^n$ is a *basic spectratope*, if it can be represented in the form

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

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 $x \in \mathbb{R}^n$ (i.e., "matrix coefficients" R^{ki} belong to $\mathbf{S}^n$)

[S$_2$] $\mathcal{T} \in \mathbb{R}^K_+$ is a convex compact subset of $\mathbb{R}^K_+$ which contains a positive vector and is monotone:

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

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

A *spectratope* is a linear image $\mathcal{Y} = P\mathcal{X}$ of a basic spectratope.

We refer to $D = \sum_k d_k$ as *size of the spectratope* $\mathcal{Y}$.

Examples

[A.] Any ellitope is a spectratope.

[B.] Let L be a positive definite $d \times d$ matrix. Then the *"matrix box"*

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

is a basic spectratope. As a result, a *bounded* set $\mathcal{X} \subset \mathbb{R}^n$ given by a system of "two-sided" LMI's, specifically,

$$\mathcal{X} = \{x \in \mathbb{R}^n : \exists t \in \mathcal{T} : -t_k L_k \preceq S_k[x] \preceq t_l L_k, \, 1 \leq k \leq K\}$$

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:

$$\mathcal{X} = \{x \in \mathbb{R}^n : \exists t \in \mathcal{T} : R_k^2[x] \leq t_k I_{d_k}, \, k \leq K\} \qquad [R_k[x] = L_k^{-1/2} S_k[x] L_k^{-1/2}]$$

Same as ellitopes, *spectratopes admit fully algorithmic calculus*.

Bounding quadratic forms over ellitopes

Proposition *Let G be a symmetric $n \times n$ matrix, $\mathcal{X} \subset \mathbb{R}^n$ be given by spectratopic representation, and let*

$$\mathrm{Opt}_* = \max_{x \in \mathcal{X}} x^T G x$$

and

$$\mathrm{Opt} = \min_{\Lambda = \{\Lambda_k\}_{k \le K}} \left\{ \phi_{\mathcal{T}}(\lambda[\Lambda]) : \Lambda_k \succeq 0, P^T G P \preceq \sum_k \mathcal{R}_k^*[\Lambda_k] \right\} \qquad (QPR)$$

where $\mathcal{R}_k^*(\Lambda) : \mathbf{S}^{d_k} \to \mathbf{S}^n$ is the conjugate linear mapping,

$$[\mathcal{R}_k^*(\Lambda)]_{ij} = \tfrac{1}{2}\mathrm{Tr}\left(\Lambda[R^{ki}R^{kj} + R^{kj}R^{ki}]\right), \ 1 \le i, j \le n,$$

$\phi_{\mathcal{T}}$ is the support function of $\mathcal{T}$, and for $\Lambda = \{\Lambda_k \in \mathbf{S}^{d_k}\}_{k \le K}$,

$$\lambda[\Lambda] = [\mathrm{Tr}[\Lambda_1]; ...; \mathrm{Tr}[\Lambda_K]].$$

Then (QPR) *is solvable, and*

$$\mathrm{Opt}_* \le \mathrm{Opt} \le 2\max[\ln(2D), 1]\mathrm{Opt}.$$

Remark

The result of the proposition has some history.

- Nemirovski, Roos and Terlaky, 1999 – $\mathcal{X}$ is an intersection of centered at the origin ellipsoids/elliptic cylinders
- J. and Nemirovski 2016 – $\mathcal{X}$ is ellitope, with tighter bound

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

Note that in the case of an ellitope, (QPR) results in a somewhat worse "suboptimality factor" $O(1)\ln(\sum_{k=1}^K \mathrm{Rank}(S_k))$.

Building linear estimate

Proposition *Consider convex optimization problem*

$$\mathrm{Opt} = \min_{H,\Lambda,\tau}\left\{\tau : \begin{array}{l}(B-H^TA)^T(B-H^TA) \preceq \sum_k \mathcal{R}_k^*(\Lambda_k)\\ \sigma^2\mathrm{Tr}(H^TH)+\phi_{\mathcal{T}}(\lambda[\Lambda])\leq\tau\end{array}\right\} \qquad (*)$$

Problem $(*)$ *is solvable, and its feasible solution* (H,λ,τ) *induces a linear estimate* $\widehat{x}_H = H^T\omega$ *of* Bx*,* $x\in\mathcal{X}$*, via observation*

$$\omega = Ax + \sigma\xi,\ \xi\sim\mathcal{N}(0,I)$$

with the maximal over $\mathcal{X}$ *risk not exceeding* $\sqrt{\tau}$*.*

Proposition *Let $\mathcal{X}$ be a spectratope, and let*

$$\mathcal{Q} = \{Q \in \mathbf{S}^n_+ : \exists t \in \mathcal{T} : \mathcal{R}_k[Q] \preceq t_k I_{d_k}, k \leq K\}.$$

The set $\mathcal{Q}$ is a nonempty convex compact set containing a neighbourhood of the origin, so that the quantity

$$M_* = \sqrt{\max_{Q \in \mathcal{Q}} \mathrm{Tr}(BQB^T)},$$

is well defined and positive.

The efficiently computable linear estimate $\widehat{x}_{H_*}(\omega) = H_*^T \omega$ yielded by the optimal solution of $(*)$ is nearly optimal in terms of the risk:

$$\mathrm{Risk}[\widehat{x}_{H_*}|\mathcal{X}] \leq 2\sqrt{2\ln\left(\frac{8DM_*^2}{\mathrm{RiskS}^2_{\mathrm{Opt}}[\mathcal{X}]}\right)}\mathrm{Risk}_{\mathrm{Opt}}[\mathcal{X}],$$

where

$$\mathrm{Risk}_{\mathrm{Opt}}[\mathcal{X}] = \inf_{\widehat{x}(\cdot)} \mathrm{Risk}[\widehat{x}|\mathcal{X}]$$

is the minimax risk associated with $\mathcal{X}$, and $D = \sum_k d_k$.

3. Norms

We say that the norm $\|\cdot\|$ is spectratopic-representable if the unit ball $\mathcal{B}_*$ of the conjugate norm $\|\cdot\|_*$ is a spectratope:

$$\mathcal{B}_* = M\mathcal{Y},\ \ \mathcal{Y} = \left\{x \in \mathbb{R}^n : \exists r \in \mathcal{R} : S_\ell^2[x] \preceq r_\ell I_{f_\ell}, 1 \leq \ell \leq L\right\},$$

where $S_\ell \in \mathbb{R}^{f_\ell \times f_\ell}$, r_ℓ, $\ell = 1, ..., L$ and $\mathcal{R}$ is a "valid spectratopic data." We denote $F = \sum_\ell f_\ell$ the size of $\mathcal{B}_*$.

Examples

- $\|\cdot\|_p$-norm with $1 \leq p \leq 2$ – $\mathcal{B}_*$ is the unit ball of $\|\cdot\|_q$-norm with $\frac{1}{p} + \frac{1}{q} = 1$
- $\|\cdot\|_1 + \|\cdot\|_2$ – B^* is an affine image of the the direct product of unit balls of norms $\|\cdot\|_\infty$ and $\|\cdot\|_2$
- "combined norm" $\min_{x=u+v} \|M_1 u\|_1 + \|M_2 v\|_2$ – $\mathcal{B}_*$ is the intersection $\mathcal{B}_*^\infty \cap \mathcal{B}_*^2$ of unit balls of norms $\|M_1^T \cdot\|_\infty$ and $\|M_2^T \cdot\|_2$
- nuclear norm $\|\cdot\|_{\mathrm{Sh},1}$ – $\mathcal{B}_*$ is the unit ball of the spectral norm $\|\cdot\|_{\mathrm{Sh},\infty}$
- ... spectral norm $\|\cdot\|_{\mathrm{Sh},\infty}$ is "difficult"

For an estimate $\widehat{x}$ of Bx, let

$$\mathrm{Risk}_{\|\cdot\|}[\widehat{x}|\mathcal{X}] = \sup_{x\in\mathcal{X}} \mathbf{E}_{\xi\sim\mathcal{N}(0,I_m)}\{\|Bx-\widehat{x}(Ax+\sigma\xi)\|\}.$$

Proposition *Consider the convex optimization problem*

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

Here for $\Lambda=\{\Lambda_i\in\mathbf{S}^{m_i}\}_{i\leq I}$

$$\lambda[\Lambda]=[\mathrm{Tr}[\Lambda_1];...;\mathrm{Tr}[\Lambda_I]],$$

$$\begin{array}{ll}[\mathcal{R}_k^*[\Lambda_k]]_{ij}=\frac{1}{2}\mathrm{Tr}(\Lambda_k[R_k^{ki}R_k^{kj}+R_k^{kj}R_k^{ki}]), & \textit{where } R_k[x]=\sum_i x_iR^{ki},\\ [\mathcal{S}_\ell^*[\Upsilon_\ell]]_{ij}=\frac{1}{2}\mathrm{Tr}(\Upsilon_\ell[S_\ell^{\ell i}S_\ell^{\ell j}+S_\ell^{\ell j}S_\ell^{\ell i}]), & \textit{where } S_\ell[y]=\sum_i y_iS^{\ell i},\end{array}$$

and $\phi_{\mathcal{T}}$ *and* $\phi_{\mathcal{R}}$ *are the support function of* $\mathcal{T}$ *and* $\mathcal{R}$*. 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*

$$\mathrm{Risk}_{\|\cdot\|}[\widehat{x}(\cdot)|\mathcal{X}]\leq\mathrm{Opt}.$$

Near-optimality of linear estimation on spectratopes

Proposition *Let*

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

Then there is an efficiently computable *linear estimate* $\widehat{x}_{H_*} = H_*\omega$ *which satisfies*

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

where C *is a positive absolute constant,*

$$\mathrm{Risk}_{\mathrm{Opt}}[\mathcal{X}] = \inf_{\widehat{x}(\cdot)}\left[\sup_{x\in\mathcal{X}} \mathbf{E}_{\xi\sim\mathcal{N}(0,I_m)}\{\|Bx-\widehat{x}(Ax+\sigma\xi)\|\}\right]$$

the infimum being taken over all estimates, and

$$D=\sum_k d_k,\ F=\sum_\ell f_\ell.$$

The key component

Lemma *Let Y be an $N \times \nu$ matrix, let $\|\cdot\|$ be a norm on $\mathbb{R}^\nu$ such that the unit ball $\mathcal{B}_*$ of the conjugate norm is the spectratope, and let $\zeta \sim \mathcal{N}(0, Q)$ for some positive semidefinte $N \times N$ matrix Q.*

Then the upper bound on

$$\phi_Q(Y) := \mathbf{E}\{\|Y^T\zeta\|\}$$

yielded by the SDP relaxation, that is, the optimal value $\mathrm{Opt}[Q]$ *of the convex optimization problem*

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

is tight, namely,

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

where $F = \sum_\ell f_\ell$ is the size of the spectratope $\mathcal{B}_*$.