Linear Recovery from Gaussian Observations

joint work with

A. Nemirovski, Georgia Tech

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

Paris, January 9, 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: $\mathbb{R}^n \to \mathbb{R}^k$.

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

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

$\Rightarrow$ 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*

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

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

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

Note: ϕ is convex $\Rightarrow$ *building the minimum risk linear estimate reduces to solving convex minimization problem*

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

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 $S \succ 0$

$$\max_{x^T S x \leq 1} \mathrm{Tr}(xx^T W) = \lambda_{\max}\left(S^{-1}W\right).$$

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

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

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

• We are about to consider a family of sets $\mathcal{X}$ – *ellitopes* – for which reasonably tight bounds Ψ^+ 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 $\{x\in\mathbb{R}^n: -1\leq x_i\leq 1\}$ $(\mathcal{T}=\{t\in\mathbb{R}^n: 0\leq t_k\leq 1,\ k\leq K=n\},\ x^TS_kx=x_k^2)$

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

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
- $\bigcap_i \mathcal{X}_i$
- $\mathcal{X}_1\times...\times\mathcal{X}_I$
- $\mathrm{Conv}(\bigcup_i \mathcal{X}_i)$
- $\mathcal{X}_1+...+\mathcal{X}_I$
- ...

Observation

Let

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

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

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

where

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

We conclude that

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

and

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

This attracts our attention to the optimization problem

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

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

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

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

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

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 $\sigma > 0$) satisfies the risk bound*

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

Bayesian risks

• *Minimax risk* $\mathrm{Risk}[\widehat{x}|\mathcal{X}]$ 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 signal $x \in \mathbb{R}^n$, 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 $\|\cdot\|_2^2$ recovery 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

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

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, so that 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*

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

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

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

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

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

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

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

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

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

Now, all we need is a simple result as follows:
Lemma *Let S and Q be positive semidefinite $n\times n$ matrices with $\rho := \mathrm{Tr}(SQ) \leq 1$, and let $\eta \sim \mathcal{N}(0, Q)$. Then*

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

We arrive at the following

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

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

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*

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

satisfies the risk bound

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

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:

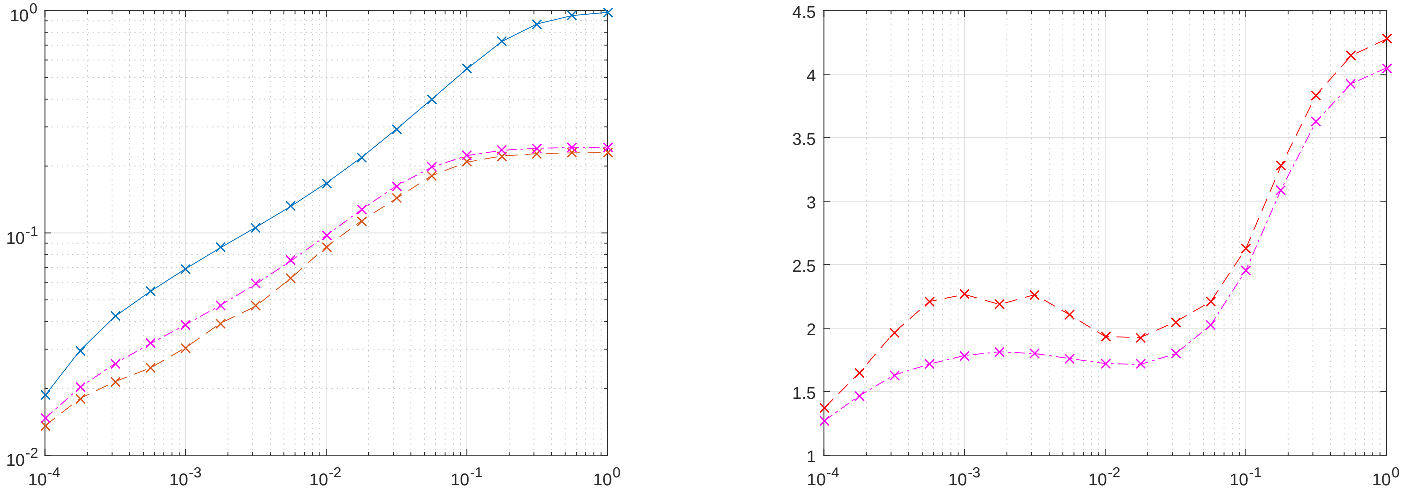
$$\mathcal{X}_1 = \{x \in \mathbb{R}^n : \sum_{j=1}^{n} j^2 x_j^2 \leq 1\},$$

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:

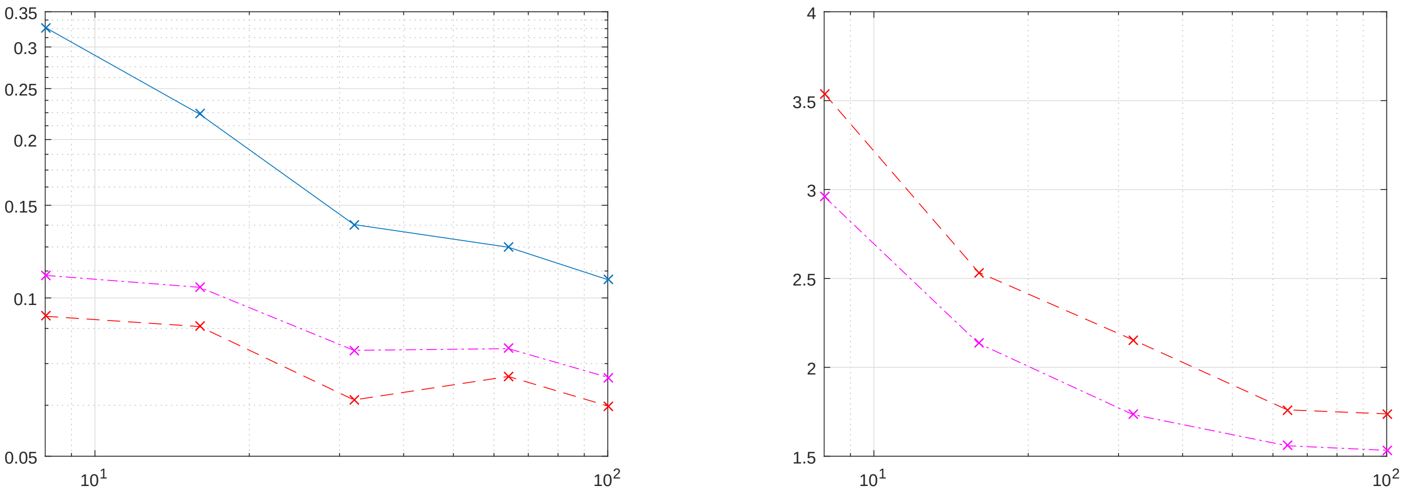
$$\mathcal{X} = \{x \in \mathbb{R}^n : j|x_j| \leq 1,\ 1 \leq j \leq n\} \quad [K = n, S_k = k^2 e_k e_k^T, k = 1, ..., K, \mathcal{T} = [0, 1]^K].$$

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



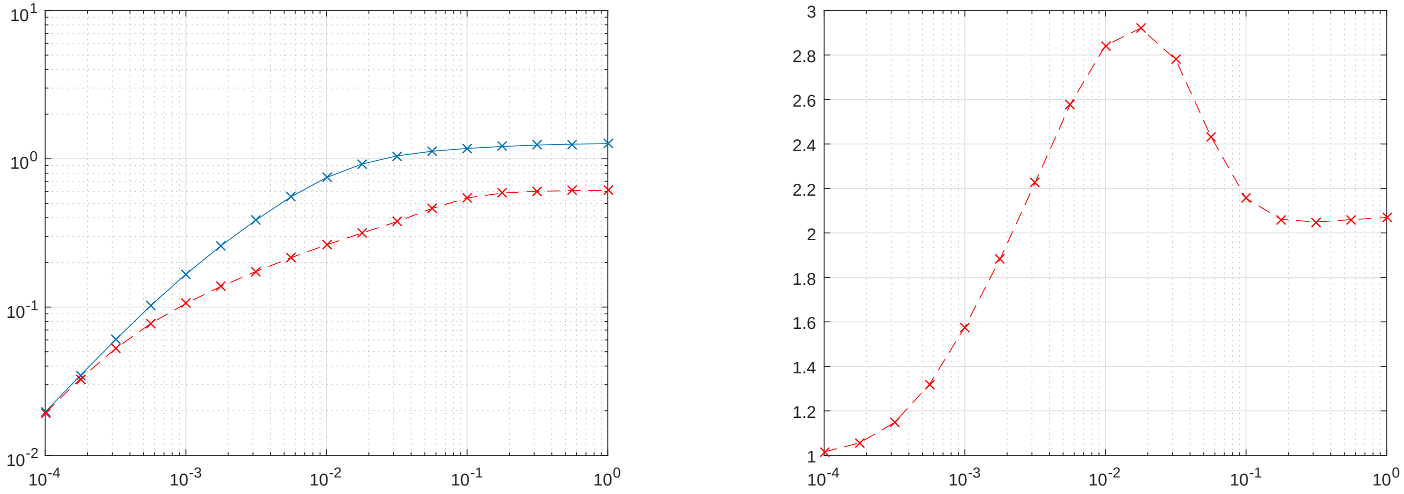
Recovery on ellipsoids: risk bounds as functions of the noise level σ, dimension $n = 32$.

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



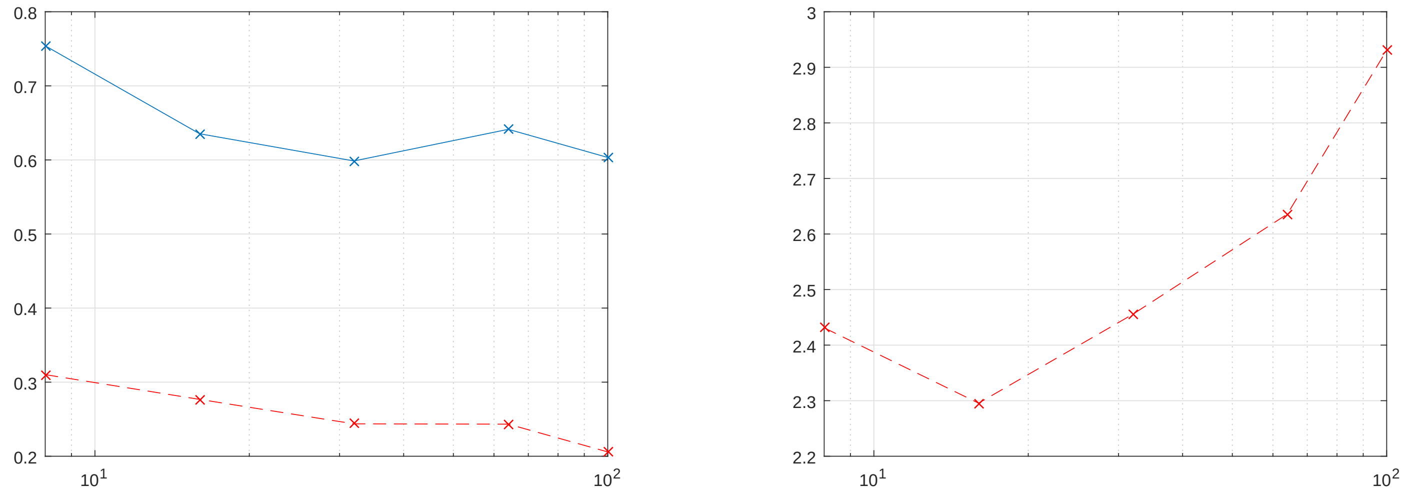
Recovery on ellipsoids: risk bounds as functions of problem dimension n, noise level $\sigma = 0.01$.

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.

Application: Inverse Heat Equation

Situation: square plate is heated at time 0 and is rest to cool; the temperature at the plate's boundary is kept 0 all the time.
Given noisy measurements of plate's temperature at time t_1, taken along m points, we want to recover distribution of temperature at a given time t_0, $0 < t_0 < t_1$.

The Model: the temperature field $u(t;p,q)$ evolves according to *Heat Equation*

$$\frac{\partial}{\partial t}u(t;p,q) = \left[\frac{\partial}{\partial p^2} + \frac{\partial^2}{\partial q^2}\right] u(t;p,q),\ t \geq 0, (p,q) \in S$$

- t: time
- $S = \{(p,q) : -1 \leq p,q \leq 1\}$: the plate

with boundary conditions

$$u(t;p,q)\big|_{(p,q)\in\partial S} \equiv 0.$$

It is convenient to represent $u(t;p,q)$ by its expansion

$$\begin{array}{rcl} u(t;p,q) & = & \sum\limits_{k,\ell} x_{k\ell}(t)\phi_k(p)\phi_\ell(q), \qquad\qquad (*) \\ \phi_k(s) & = & \left\{\begin{array}{lll} \cos(\omega_{2i-1}s), & \omega_{2i-1} = (i-1/2)\pi, & k = 2i-1 \\ \sin(\omega_{2i}s), & \omega_{2i} = i\pi, & k = 2i \end{array}\right. \end{array}$$

Note:

- $\{\phi_{k\ell}(p,q) = \phi_k(p)\phi_\ell(q)\}_{k,\ell}$ form an orthonormal basis in $L_2(S)$
- $\phi_{k\ell}(\cdot)$ meet the boundary conditions $\phi_{k\ell}(p,q)\big|_{(p,q)\in\partial S} = 0$
- in terms of the coefficients $x_{k\ell}(t)$, the Heat Equation becomes

$$\frac{d}{dt}x_{k\ell}(t) = -(\omega_k^2 + \omega_\ell^2)x_{k\ell}(t) \Rightarrow x_{k\ell}(t) = \mathrm{e}^{-(\omega_k^2+\omega_\ell^2)}x_{k\ell}(0).$$

We select integer discretization parameter N and

- restrict $(*)$ to terms with $1 \leq k, \ell \leq 2N - 1$
- discretize the spatial variable (p, q) to reside on the grid

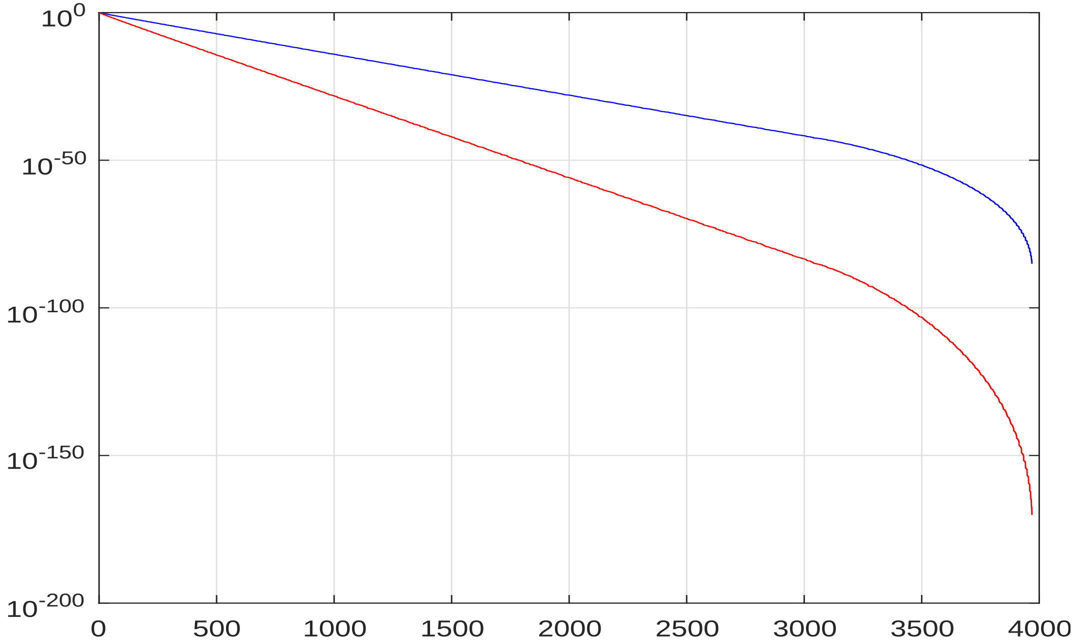
$$G_N = \{P_{ij} = (p_i, p_j) = (i/N - 1, j/N - 1),\ 1 \leq i, j \leq 2N - 1\}$$

Restricting functions $\phi_{k\ell}(\cdot,\cdot)$, $1 \leq k, \ell \leq 2N-1$ on the grid G_N, we get orthogonal basis in $\mathbb{R}^{(2N-1)\times(2N-1)}$.

The model

- The signal x underlying observation is $x = \{x_{k\ell}(0), 1 \leq k, \ell \leq 2N-1\} \in \mathbb{R}^{(2N-1)\times(2N-1)}$
- The observation is $\omega = A(x) + \sigma\xi \in \mathbb{R}^m$, $\xi \sim \mathcal{N}(0, I_m)$, where
$$[A(x)]_\nu = \sum_{k,\ell=1}^{2N-1} x_{k\ell} \mathrm{e}^{-(\omega_k^2+\omega_\ell^2)t_1} \phi_k(p_i(\nu))\phi_\ell(p_j(\nu))$$
  - $(p_{i(\nu)}, p_{j(\nu)}) \in S$, $1 \leq \nu \leq m$: measurement points
- We want to recover the restriction $B(x)$ of $u(t_0; p, q)$ of some grid, say, square grid
$$G_K = \left\{ r_i = \frac{i}{K} - 1, r_j = \frac{j}{K} - 1,\ 1 \leq i, j \leq 2K-1 \right\} \subset S,$$
resulting in
$$[B(x)]_{ij} = \sum_{k,\ell=1}^{2N-1} x_{k\ell}(0) \mathrm{e}^{-(\omega_k^2+\omega_\ell^2)t_0} \phi_k(r_i)\phi_\ell(r_j)$$
- We assume that the initial distribution of temperatures $[u(0; p_i, p_j)]_{i,j=1}^{2N-1}$ satisfies $\|u\|_2 \leq R$, for some given R, implying that x resides in a centered at the origin ball
$$\mathcal{X} \subset \mathbb{R}^{(2N-1)\times(2N-1)}$$

$$u(t; p_i, p_j) = \sum_{k,\ell} \mathrm{e}^{-(\omega_k^2+\omega_\ell^2)t} x_{k\ell}(0)\phi_k(p_i)\phi_\ell(p_j)$$



Plots of $\mathrm{e}^{-(\omega_k^2+\omega_\ell^2)t}$ for $t = 0.02$ (red) and $t = 0.01$ (blue)

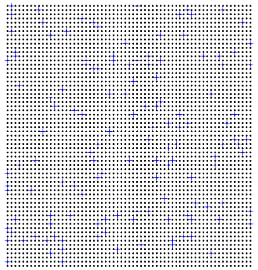
Bad news: *contributions of high frequency* (with large $\omega_k^2+\omega_\ell^2$) *components* $x_{k\ell}(0)$ *to* $A(x)$ *decrease exponentially fast with high decay rate as* $t_!$ *grows*
$\Rightarrow$ *High frequency components* $x_{k\ell}(0)$ *are impossible to recover from observations at time* t_1, *unless* t_1 *is very small* (red curve, $t=0.02$)

Good news: *Contributions of high frequency components* $x_{k\ell}(0)$ *to* $B(x)$ *are very small, provided* t_0 *is not too small*
$\Rightarrow$ *There is no need to recover well high frequency components, provided they are not huge* (blue curve, $t=0.01$)

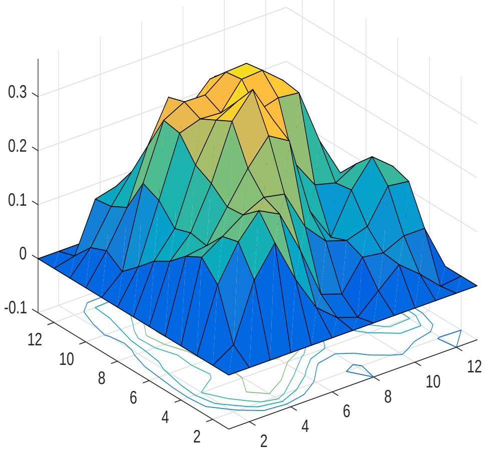
Numerical results $N=32$, $m=125$, $K=6$, $t_0=0.01$, $t_1=0.02$, $\sigma=0.001$,

$$\mathcal{X}=\{x\in\mathbb{R}^{63\times 63}:\|\{u(0;p_i,p_j)\}_{i,j=1}^{63}\|_2\leq 15\}.$$

- *Minimax risk of optimal linear estimate: 0.1707*



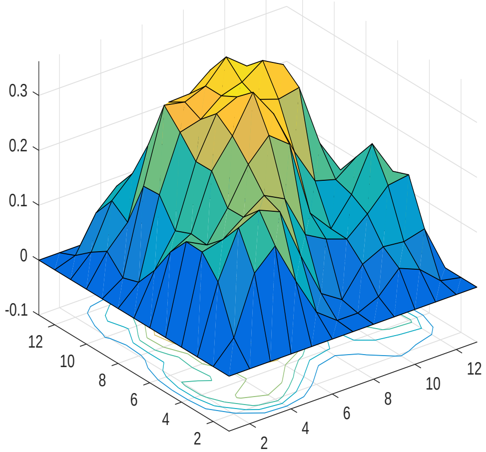
63×63 grid G_{63} and $m = 125$ measurement points



B

$\|B\|_2 = 2.128$

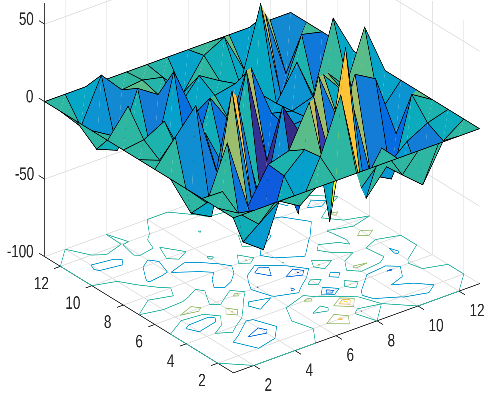
$\|B\|_\infty = 0.433$



$\widehat{B}$

$\|B - \widehat{B}\|_2 = 0.1535$

$\|B - \widehat{B}\|_\infty = 0.0313$



$\widetilde{B}$

$\|B - \widetilde{B}\|_2 = 266.3$

$\|B - \widetilde{B}\|_2 = 63.63$

Left: true signal, center: linear recovery, right: naive LS recovery

Extension I: 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

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

For the resulting estimate, it holds

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

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*

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

where

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

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 optimal solution to the convex problem

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

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

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

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 linear estimate*

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

with S*-risk* $\sqrt{\mathrm{Opt}}$.
When $\xi \sim \mathcal{N}(0, I_m)$, *this estimate is minimax optimal:*

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

Application: estimating the input of a dynamical system

Let $[r;v]$ be the state of a pendulum – a 2-dimensional dynamical system satisfying

$$\dot{r} = v, \quad \dot{v} = -\nu^2 r - \kappa v + w,$$

where w is the external input. When discretizing the system, and assuming that the input is constant on the discretization interval Δ, we come to finite-difference equation

$$z_\tau = Pz_{\tau-1} + Qw_\tau, \; \tau = 1, 2, ...$$

where

$$P = \exp\{\Delta\vartheta\}, \; Q = \int_0^\Delta \exp\{s\vartheta\} \begin{bmatrix} 0 \\ 1 \end{bmatrix} ds, \quad \vartheta = \begin{bmatrix} 0 & 1 \\ -\nu^2 & -\kappa \end{bmatrix}.$$

We assume that noisy observations of z_τ are available for $\tau = 1, ..., T$, and we want to recover "signals" – collections

$$w^k = [w_{T-k+1}, \; ..., w_T] = B^{(k)}x, \quad x = [z_0; w_1; ...; w_T].$$

We want to build the best in terms of S-risk estimate.

Building linear estimate

In our design S is not fixed in advance, but satisfies $S \succeq 0$, $\mathrm{Tr}(S) \le 1$.

To compute the estimate $\widehat{w}^k = H_*^T \omega$ of w^k we solve

$$\mathrm{Opt}[B] = \min_{\tau,H,S} \left\{ \tau : \left[\begin{array}{cc} \tau S & B^T - A^T H \\ B - H^T A & I_T \end{array} \right] \succeq 0,\ \sigma^2 \mathrm{Tr}(H^T H) \le \tau,\ S \succeq 0,\ \mathrm{Tr}(S) \le 1 \right\}$$

with $B = B^{(k)}$.

Interpretation: recall that $\mathrm{RiskS}[\widehat{w}^k|\mathbb{R}^n] \le \sqrt{\tau}$

$$\Leftrightarrow\ \mathbf{E}_{\xi\sim\mathcal{N}(0,I)}\{\|\widehat{w}^k(Ax+\sigma\xi) - B^k x\|_2^2\} \le \tau(1 + x^T S x)\ \ \forall x.$$

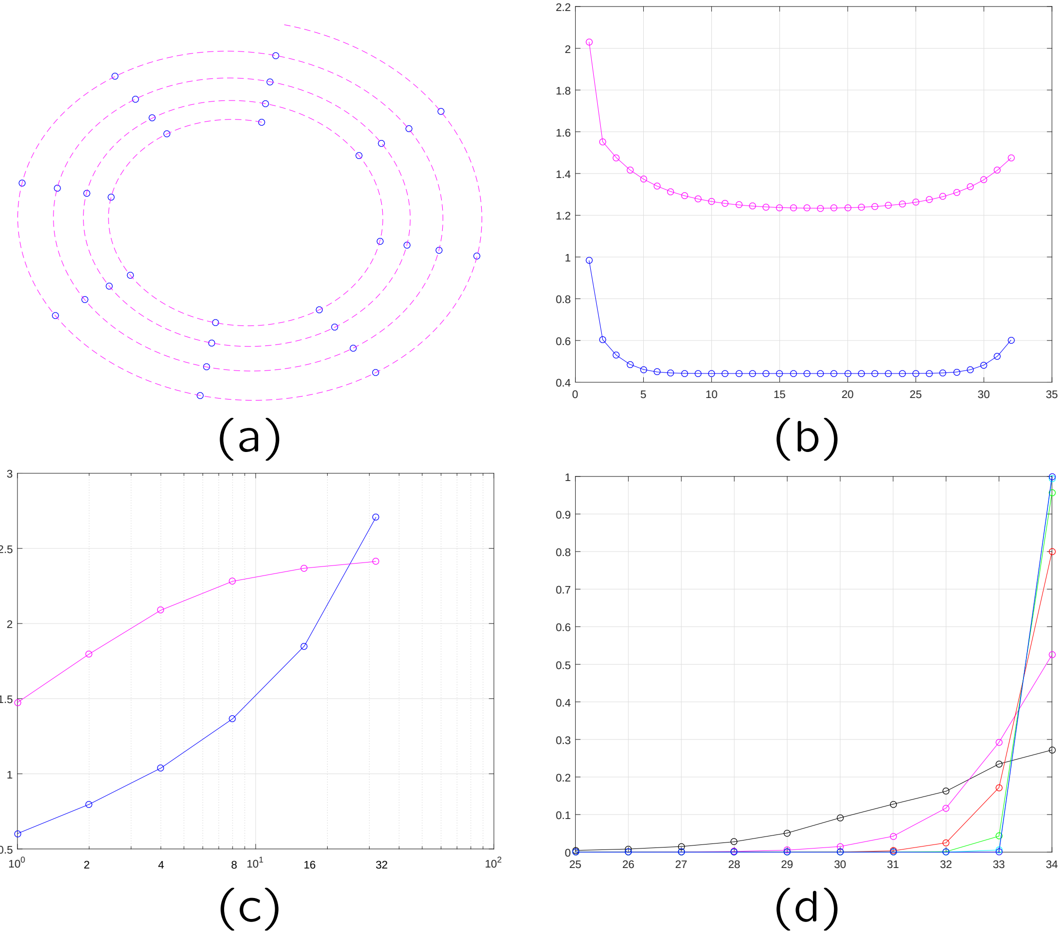
Assuming that x is random and distruibuted over a sphere $\mathcal{S}$ of radius

$$\sqrt{\dim x} = \sqrt{T+2},$$

we have

$$\mathbf{E}_{x,\xi}\{\|\widehat{w}^k(Ax+\sigma\xi) - B^k x\|_2^2\} \le 2\tau.$$

We refer to $\sqrt{\tau}$ as to *Bayesian risk* of $\widehat{x}^k$.



(a) free motion ($w \equiv 0$) in (r, v)-plane; (b) Bayesian and worst-case risks of recovering w_t vs. $t = 1, 2, ..., 32$; (c) Bayesian and worst-case risks of recovering $w^k := [w_{T-k+1}; w_{T-k+2}; ...; w_T]$ vs. k; (d) 10 largest eigenvalues $\lambda_i(S_k)$ of S_k ($k = 32$, 16, 8, 4, 2 and 1).

Extension II: linear estimation over 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 $\mathcal{Y} = PX$ 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 $\mathcal{X} \subset \mathbb{R}^\nu$ given by a system of "two-sided" LMI's, specifically,

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

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:

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

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

Building linear estimate

Proposition. *Consider convex optimization problem*

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

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

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

and for $\Lambda = \{\Lambda_k \in \mathbf{S}^{d_k}\}_{k\leq K}$, $\lambda[\Lambda] = [\mathrm{Tr}[\Lambda_1]; ...; \mathrm{Tr}[\Lambda_K]]$.

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 S*-risk not exceeding* $\sqrt{\tau}$.

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

$$\mathcal{Q} = \{Q \in \mathbf{S}^n_+ : \exists t \in \mathcal{T} : \mathcal{R}_k[Q] \preceq t_k I_{d_k}, k \le K\},\ \mathcal{Q}_\rho = \rho\mathcal{Q}, \rho > 0.$$

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 S-risk:

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

where

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

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