Some Applications of Convex Optimization in Control and Statistics

Anatoli Juditsky — University Grenoble-Alpes
Arkadi Nemirovski — Georgia Institute of Technology
Joint research with Y. Bekri & G. Kotsalis

Algorithmic Optimization: Tools for AI and Data Science
August 27-30, 2024, UCLouvain, Belgium
a.k.a.
Celebrating 50 years of Yuri Nesterov's life in Optimization

• When speaking about the operational side of Convex Optimization, the emphasis is usually on algorithms– their synthesis, complexity analysis, etc.

• **However:** the Convex Optimization Operational toolbox is more than just algorithms. While at the end of the day, we want to get a number, at the beginning of the day, we need a computationally tractable model responsible for this number.

Computation-friendly modeling of applied problems is often a highly nontrivial and challenging task. To resolve this task, one usually needs to utilize the Descriptive toolbox of Convex Optimization, in our experience – primarily Conic Duality.

• In this talk, we present several computation-friendly models (in our appreciation, interesting and somehow instructive) for applications coming from Control and Statistics. We focus on presenting our related results, skipping the underlying developments (not always trivial).

Preliminaries: Bounding induced norms of uncertain matrices

- "Working horse" of what follows addresses the following problem:

  Given unit balls $\mathcal{X} \subset \mathbb{R}^n, \mathcal{Y} \subset \mathbb{R}^m$ of norms $\|\cdot\|_{\mathcal{X}}$, $\|\cdot\|_{\mathcal{Y}}$ and uncertain matrix $\mathcal{A}$ with structured norm-bounded uncertainty – a parametric set

$$\mathcal{A} = \left\{ A_{\rm nom} + \sum_s \delta_s A_s + \sum_t L_t^\top \Delta_t R_t : |\delta_s| \le 1, s \le S, \underbrace{\|\Delta_t\|_{2,2}}_{\text{spectral norm}} \le 1, t \le T \right\}$$

  of $m \times n$ matrices – compute the robust $\mathcal{X}$,$\mathcal{Y}$-norm

$$\|\mathcal{A}\|_{\mathcal{X},\mathcal{Y}} = \max_{A \in \mathcal{A}} \|A\|_{\mathcal{X},\mathcal{Y}}$$

$$\left[\|A\|_{\mathcal{X},\mathcal{Y}} = \max_{x \in \mathcal{X}} \|Ax\|_{\mathcal{Y}} : \text{induced by } \|\cdot\|_{\mathcal{X}}, \|\cdot\|_{\mathcal{Y}} \text{ norm of } A \right]$$

  of $\mathcal{A}$.

- **Fact:** *Aside from a small number of unique cases, computing $\|\mathcal{A}\|_{\mathcal{X},\mathcal{Y}}$ is hard already when $\mathcal{A} = \{A_{\rm nom}\}$ is certain. However, $\|\mathcal{A}\|_{\mathcal{X},\mathcal{Y}}$ admits efficiently computable tight upper bound, provided $\mathcal{X}$ and the polar $\mathcal{Y}_*$ of $\mathcal{Y}$ possess nice geometry.*

- **Nice geometry sets:** *ellitopes* and *spectratopes*

Ellitopes and spectratopes

• **Basic ellitope** of e-size K: a bounded set in $\mathbb{R}^N$ represented as

$$\mathcal{V}=\{v:\exists t\in\mathcal{T}:v^\top T_k v\leq t_k, k\leq K\},$$

where $T_k\succeq 0$, and $\mathcal{T}\subset\mathbb{R}^K_+$ is a convex compact set containing a positive vector and *monotone:* $0\leq t'\leq t\in\mathcal{T}\Rightarrow t'\in\mathcal{T}$

Basic examples: (a) finite and bounded intersections of centered at the origin ellipsoids/elliptics cylinders, and (b) $\|\cdot\|_p$-balls, $2\leq p\leq\infty$.

Ellitope of e-size K: linear image of basic ellitope of e-size K.

• **Basic spectratope** of s-size D a bounded set in $\mathbb{R}^N$ represented as

$$\mathcal{V}=\{v:\exists t\in\mathcal{T}:T_k^2[v]\preceq t_k I_{d_k}, k\leq K\},\ T_k[v]=\sum_j v_j T^{kj},\ T^{kj}\in\mathbf{S}^{d_k},$$

where $\mathcal{T}\subset\mathbb{R}^K_+$ is as in the definition of basic ellitope, and $D=\sum_k d_k$

Basic examples: (a) basic ellitopes, and (b) matrix boxes $\{v\in\mathbb{R}^{p\times q}:\|v\|_{2,2}\leq 1\}$

Spectratope of s-size D: linear image of basic spectratope of s-size D.

• **Fact:** *As applied to ellitopes/spectratopes, basic operations preserving convexity and symmetry w.r.t. the origin result in ellitopes/spectratopes: when $\mathcal{V}_1,...,\mathcal{V}_N$ are ellitopes (spectratopes), so are their intersections, direct products, sums, linear images, and inverse images under linear embeddings.*

$$\mathcal{A}=\{A_{\rm nom}+\sum_s\delta_sA_s+\sum_t L_t^\top\Delta_tR_t : |\delta_s|\leq 1, s\leq S, \|\Delta_t\|_{2,2}\leq 1, t\leq T\}$$
$$\|\mathcal{A}\|_{\mathcal{X},\mathcal{Y}}:=\max_{A\in\mathcal{A},x\in\mathcal{X},u\in{\rm Polar}\,(\mathcal{Y})} u^\top Ax\leq ???$$

- **Theorem I** [J&Kotsalis&N'22,Bekri&J&N'23] *Let $\mathcal{X}$ and ${\rm Polar}\,(\mathcal{Y})$ be ellitopes of e-sizes K,L. Then $\|\mathcal{A}\|_{\mathcal{X},\mathcal{Y}}$ admits efficiently computable upper bound ${\rm Opt}[\mathcal{A}]$ such that*
  - *the bound is reasonably tight:*

$$\|\mathcal{A}\|_{\mathcal{X},\mathcal{Y}}\leq {\rm Opt}[\mathcal{A}]\leq [\varsigma(K,L)+\varkappa(K)\varkappa(L)\max[\vartheta(2\max_s{\rm Rank}(A_s)),\pi/2]]\|\mathcal{A}\|_{\mathcal{X},\mathcal{Y}}$$

$$\left[\begin{array}{l}\varsigma(K,L)=\left\{\begin{array}{ll}3\sqrt{\ln(3K)\ln(3L)} & ,\max[K,L]>1\\ 1 & ,K=L=1\end{array}\right. \quad \varkappa(J)=\left\{\begin{array}{ll}\frac{5}{2}\sqrt{\ln(2J)} & ,J>1\\ 1 & ,J=1\end{array}\right.\\ \vartheta(\cdot): \text{ universal function such that}\\ \qquad \vartheta(1)=1,\vartheta(2)=\pi/2,\vartheta(3)=1.7348...,\vartheta(4)=2\ \&\ \vartheta(k)\leq\frac{\pi}{2}\sqrt{k}, k\geq 1\end{array}\right]$$

  - *the bound is convex in the part $A_{\rm nom},\{A_s,s\leq S,L_t,t\leq T\}$ of the data of $\mathcal{A}$*
- **Similar fact** *with e-sizes replaced with s-sizes and modified absolute constants in the definitions of $\varsigma,\varkappa$, holds true when $\mathcal{X}$, ${\rm Polar}\,(\mathcal{Y})$ are spectratopes.*

Note: When $\mathcal{A}$ is certain (i.e., $S=T=0$), the tightness factor is just $\varsigma(K,L)$.

$$\mathcal{A} = \left\{A_{\rm nom} + \sum_s \delta_s A_s + \sum_t L_t^\top \Delta_t R_t : |\delta_s| \leq 1, s \leq S, \|\Delta_t\|_{2,2} \leq 1, t \leq T\right\}$$
$$\|\mathcal{A}\|_{\mathcal{X},\mathcal{Y}} := \max_{A\in\mathcal{A}, x\in\mathcal{X}, u\in{\rm Polar}\,(\mathcal{Y})} u^\top Ax \leq ???$$

• Theorem I is directly applicable when $\|\cdot\|_{\mathcal{X}}$ is a *simple ellitopic norm* – $\mathcal{X}$ is ellitope, and $\|\cdot\|_{\mathcal{Y}}$ is a *simple co-ellitopic norm* – Polar $(\mathcal{Y})$ is an ellitope.

For example,

• *the norm* $\|x\|_{\mathcal{X}} = \left\|\left[\|S_1x\|_{\mathcal{X}_1}; ...; \|S_Kx\|_{\mathcal{X}_K}\right]\right\|_s$, *where* $\cap_k {\rm Ker} S_k = \{0\}$, $\mathcal{X}_k$, $k \leq K$, *are ellitopes, and* $s \in [2,\infty]$ *is simple ellitopic.* In particular, *the block* ℓ_r/ℓ_s *norm*

$$\|[x^1; ...; x^K]\| = \left\|\left[\|x^1\|_{r_1}; ...; \|x^K\|_{r_K}\right]\right\|_s \qquad (*)$$

with $s, r_1, ..., r_K \in [2,\infty]$, *is simple ellitopic.*

• *the norm* $\|[y^1; ...; y^K]\| = \left\|\left[\|S_1y^1\|_{\mathcal{Y}_1}; ...; \|S_Ky^K\|_{\mathcal{Y}_K}\right]\right\|_s$, *where* S_k *are invertible,* Polar $(\mathcal{Y}_k)$ *are ellitopes,* $k \leq K$, *and* $s \in [1,2]$ *is simple co-ellitopic.* In particular, *the block* ℓ_r/ℓ_s *norm* $(*)$ *with* $s, r_1, ..., r_K \in [1,2]$ *is simple co-ellitopic.*

$$\mathcal{A} = \left\{A_{\text{nom}} + \sum_s \delta_s A_s + \sum_t L_t^\top \Delta_t R_t : |\delta_s| \leq 1, s \leq S, \|\Delta_t\|_{2,2} \leq 1, t \leq T\right\}$$
$$\|\mathcal{A}\|_{\mathcal{X},\mathcal{Y}} := \max_{A\in\mathcal{A}, x\in\mathcal{X}, u\in\text{Polar}\,(\mathcal{Y})} u^\top A x \leq ???$$

• We can use Theorem I also in the case when $\|\cdot\|_{\mathcal{X}}$ is an *ellitopic norm* –

$$\mathcal{X} = \text{Conv}\big(\cup_{j\leq M} P_j \mathcal{X}_j\big) \text{ with ellitopes } \mathcal{X}_j \Leftrightarrow \|x\|_{\mathcal{X}} = \min_{x^j, j\leq N} \left\{\sum_j \|x^j\|_{\mathcal{X}_j} : \sum_j P_j x^j = x\right\}$$

and $\|\cdot\|_{\mathcal{Y}}$ is a *co-ellitopic norm* – the conjugate of ellitopic norm, or, equivalently,

$$\|y\|_{\mathcal{Y}} = \max_{i\leq N} \|Q_i^\top y\|_{\mathcal{Y}_i} \text{ where Polar}\,(\mathcal{Y}_i) \text{ are ellitopes}$$

In this case

$$\|\mathcal{A}\|_{\mathcal{X},\mathcal{Y}} = \max_{i\leq M, j\leq N} \|Q_i^\top \mathcal{A} P_j\|_{\mathcal{X}_j,\mathcal{Y}_i}$$
$$Q^\top \mathcal{A} P = \{Q^\top A_{\text{nom}} P + \sum_s \delta_s Q^\top A_s P + \sum_t [L_t Q]^\top \Delta_t [R_t P] : |\delta_s| \leq 1, \|\Delta_t\|_{2,2} \leq 1\}$$

and we can upper-bound $\|\mathcal{A}\|_{\mathcal{X},\mathcal{Y}}$ by the maximum of the upper bounds on $\|Q_i^\top \mathcal{A} P_j\|_{\mathcal{X}_j,\mathcal{Y}_i}$ given by Theorem I.

For example,

• the block ℓ_r/ℓ_1 norm $\|[x^1; ...; x^K]\| = \sum_k \|x^k\|_{r_k}$ with $r_k \in [2, \infty]$ is ellitopic

• the block ℓ_r/ℓ_∞ norm $\|[y^1; ...; y^K]\| = \max_k \|y^k\|_{r_k}$ with $r_k \in [1, 2]$ is co-ellitopic

Application A: Controlling Peak-to-Peak Gain in Discrete Time Linear Systems

[J&Kotsalis&N'22]

- **Situation:** *Given Discrete Time Linear Dynamical System*

[initial state]	x_0	$=$	z
[state equations]	x_{t+1}	$=$	$A_t x_t + B_t u_t + C_t d_t,\ 0 \le t < N$
[observable outputs]	y_t	$=$	$D_t x_t + E_t d_t,\ 0 \le t < N$

• N: time horizon • x_t: states • z: initial state • u_t: controls • y_t: outputs
• $d_0, ..., d_{N-1}$: disturbances

we want to design affine controller $u_t = g_t + \sum_{\tau=0}^{t} G_t^\tau y_\tau,\ 0 \le t < N$ *obeying given constraints on the dependence of the state-control-output trajectory*

$$[x_1; ...; x_N] = X_d \underbrace{[d_0; ...; d_{N-1}]}_{d} + X_z z + \overline{x},$$

$$[u_0; ...; u_{N-1}] = U_d d + U_z z + \overline{u},\ [y_0; y_1; ...; y_{N-1}] = Y_d d + Y_z z + \overline{y}$$

on the initial state z *and the disturbances* d_t*.*

- *Peak-to-Peak* specifications upper-bound the norms of the matrices X_d, U_d, Y_d induced by block-ℓ_∞ norms

$$\|[d_0; ...; d_{N-1}]\| = \max_{0\le t<N} \|d_t\|_{(d)}, \quad \|[x_1; ...; x_N]\| = \max_{1\le t\le N} \|x_t\|_{(x)},$$
$$\|[u_0; ...; u_{N-1}]\| = \max_{0\le t<N} \|u_t\|_{(u)}, \quad \|[y_0; ...; y_{N-1}]\| = \max_{0\le t<N} \|y_t\|_{(y)}$$

of disturbances, states, controls, and outputs. *These bounds become constraints on the parameters of the controller responsible for the matrices* X_d, U_d, Y_d.

• When the norm $\|\cdot\|_{(d)}$ and norms *conjugate* to $\|\cdot\|_{(x)}, \|\cdot\|_{(u)}, \|\cdot\|_{(y)}$ are ellitopic/spectratopic, the induced norms of X_d, U_d, Y_d admit tight efficient upper-bounding, making the *Analysis problem* "given affine controller, check whether the peak-to-peak specifications are met" – more or less tractable.

However: X_d, U_d, Y_d are highly nonlinear in the parameters of controller, making *synthesis* of an affine controller obeying peak-to-peak specifications heavily intractable.

• **Remedy:** *Smart nonlinear reparameterization of affine output-based nonanticipating controllers makes the trajectory bi-affine function of controller's parameters and of $[z; d]$. With this parameterization, X_d, U_d, Y_d become affine in the design parameters, making synthesis tractable.*

• **Purified outputs.** We augment the controlled system with its *model*

System:
$x_0 = z,\ x_{t+1} = A_t x_t + B_t u_t + C_t d_t,\ 0 \leq t < N,\ y_t = D_t x_t + E_t d_t,\ 0 \leq t < N$
Model:
$\overline{x}_0 = 0,\ \overline{x}_{t+1} = A_t \overline{x}_t + B_t u_t,\ 0 \leq t < N,\ \overline{y}_t = D_t x_t,\ 0 \leq t < N$

and run the model in parallel with the actual system, feeding both with the same controls u_t yielded by a (whatever nonanticipating) controller. At time t, after y_t is observed and before u_t is to be specified, we have at our disposal *purified output* $v_t = y_t - \overline{y}_t$

System:
$x_0 = z,\ x_{t+1} = A_t x_t + B_t u_t + C_t d_t,\ 0 \le t < N,\ y_t = D_t x_t + E_t d_t,\ 0 \le t < N$
Model:
$\overline{x}_0 = 0,\ \overline{x}_{t+1} = A_t \overline{x}_t + B_t u_t,\ 0 \le t < N,\ \overline{y}_t = D_t x_t,\ 0 \le t < N$
Purified outputs:
$v_t = y_t - \overline{y}_t$

Facts [Ben-Tal&Boyd&N.'05]

- *purified outputs are affine functions of $z, d_0, d_1, \ldots$ completely independent of how the controls are generated*
- *Passing from affine output-based controllers to affine purified-output-based ones*

$$u_t = h_t + \sum_{\tau=0}^{t} H_t^\tau v_\tau,$$

we preserve achievable behaviors of the controlled system: every mapping from the space of sequences $z, d_0, d_1, \ldots$ to the space of sequences $x_0, y_0, u_0, x_1, y_1, u_1, \ldots$ stemming from an affine output-based controller stems from an affine purified-output-based one as well, and vice versa.

- *For an affine purified-output-based controller, the matrices $X_z, X_d, U_z, U_d, Y_z, Y_d$ become affine in the controller's parameters $\{h_t, H_t^\tau, 0 \le \tau \le t\}$, paving road to computationally efficient synthesis of controllers under a wide spectrum of design specifications.*

$x_0 = z,\ x_{t+1} = A_t x_t + B_t u_t + C_t d_t,\ \ y_t = D_t x_t + E_t d_t,\ \ 0 \le t < N$ $y_0, \ldots, y_t \Rightarrow v_0, \ldots, v_t \Rightarrow u_t = h_t + \sum_{\tau=0}^{t} H_t^\tau v_\tau$
$\begin{array}{rcl} [x_1; \ldots; x_N] &=& X_d[\eta]\overbrace{[d_0; \ldots; d_{N-1}]}^{d} + X_z[\eta]z + \overline{x} \\ [u_0; \ldots; u_{N-1}] &=& U_d[\eta]d + U_z z + \overline{u} \\ [y_0; y_1; \ldots; y_{N-1}] &=& Y_d[\eta]d + Y_z[\eta]z + \overline{y} \end{array}$
$X_d[\eta], \ldots, Y_z[\eta]$: affine in $\eta = \{h_t, H_\tau^t, 0 \le \tau \le t < N\}$

- The *peak-to-peak gain* from, say, d to x is the norm $\|X_d[\eta]\|_{\mathcal{D},\mathcal{X}}$ with

$$\|[d_0; \ldots; d_{N-1}]\|_{\mathcal{D}} = \max_{0 \le t < N} \|d_t\|_{(d)},\ \|[x_1; \ldots; x_N]\|_{\mathcal{X}} = \max_{1 \le t \le N} \|x_t\|_{(x)}$$

Assuming the norm $\|\cdot\|_{(d)}$ simple ellitopic, and the norm $\|\cdot\|_{(x)}$ co-ellitopic, with e-sizes of the participating ellitopes not exceeding K for $\|\cdot\|_{(d)}$ and L for $\|\cdot\|_{(x)}$, Theorem I provides efficiently computable convex in η upper bound $\Psi(\eta)$ on $\|X_d[\eta]\|_{\mathcal{D},\mathcal{X}}$, tight within the factor $O(1)\sqrt{\ln(KN+1)\ln(L+1)}$, and we can "safely approximate" bounding the peak-to-peak gain with bounding Ψ.

Assuming unit balls of the norm $\|\cdot\|_{(d)}$ and of the norm conjugate to $\|\cdot\|_{(x)}$ ellitopes, Theorem I provides efficiently computable convex in η upper bound $\Psi(\eta)$ on $\|X_d[\eta]\|_{\mathcal{D},\mathcal{X}}$, and we can "safely approximate" bounding the peak-to-peak gain with bounding Ψ.

Note: the approximation ratio $\Psi(\eta)/\|X_d[\eta]\|_{\mathcal{D},\mathcal{X}}$ is $\leq O(1)\sqrt{\ln(NK+1)\ln(L+1)}$. In some special cases it can be improved [Nesterov'98,Nesterov'00] to

- $\frac{\pi}{4-\pi} \approx 3.660$, when $\|\cdot\|_{(d)}$-unit ball is basic ellitope with mutually commuting matrices in the ellitopic representation, and similarly for the polar of the $\|\cdot\|_{(x)}$-unit ball
- $\frac{\pi}{2\sqrt{3}-2\pi/3} \approx 2.2936$, when $\|\cdot\|_{(d)} = \|\cdot\|_q$, $\|\cdot\|_{(x)} = \|\cdot\|_p$, $1 \leq p \leq 2 \leq q$
- $\sqrt{\pi/2} \approx 1.2533$, when $\|\cdot\|_{(d)} = \|\cdot\|_q$, $\|\cdot\|_{(x)} = \|\cdot\|_2$, $q \geq 2$
- **Note:** logarithmic in K, L dependence of the approximation ratio *is a must* in the general ellitopic case, same as in the special cases above, provided that we want to bound the norm of *the restriction* of a linear map onto a linear subspace in the argument space.

How It Works:
Optimizing Peak-to-Peak Gain for Boeing 747

• The system: linearized and discretized in time longitudinal dynamics model of Boeing 474's cruise flight in XZ plane [S. Boyd, Lecture Notes on Linear Dynamical Systems, 2007]

Dimensions: $\dim x = 4$, $\dim u = \dim y = \dim d = 2$.

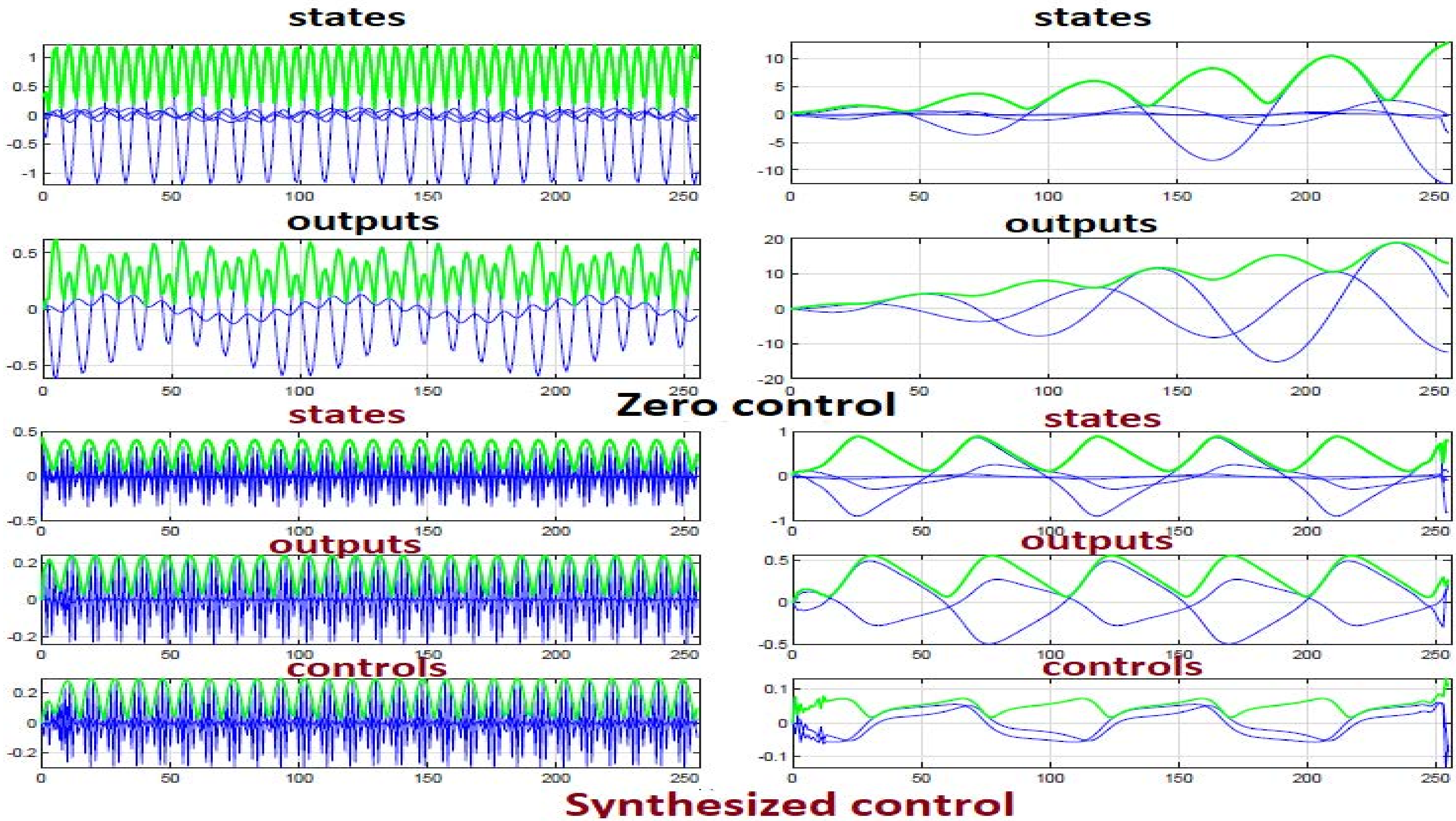
$$x_{t+1} = \left[\begin{array}{cccccccc} 0.996 & 0.034 & -0.021 & -0.321 & 0.014 & 0.989 & 0.004 & -0.033 \\ 0.008 & 0.470 & 4.664 & 0.002 & -3.437 & 1.665 & -0.008 & 0.528 \\ 0.017 & -0.060 & 0.404 & -0.003 & -0.822 & 0.438 & -0.017 & 0.060 \\ 0.009 & -0.037 & 0.719 & 0.999 & -0.473 & 0.249 & -0.009 & 0.037 \end{array}\right] \left[\begin{array}{c} x_t \\ u_t \\ d_t \end{array}\right]$$

$$y_t = \left[\begin{array}{cccc} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 7.74 \end{array}\right] x_t$$

• We use $\|\cdot\|_{(x)} = \|\cdot\|_{(d)} = \|\cdot\|_2$ and "training horizon" $N = 128$. The synthesized control is tested on horizon $N = 256$.

• Open loop system (i.e., all-zero controls) is stable with peak-to-peak disturbance-to-state gain bounded by ≈ 12 uniformly in N.

• Minimizing the (tight upper bound on the) peak-to-peak gain over the parameters of linear purified-output-based controller, the gain is reduced to ≈ 1.02.



Disturbances: Left pane – random harmonic oscillation, right pane – “bad” (both of unit ℓ_2/ℓ_∞-norm)

In green: $\|\cdot\|_2$-norms of states/outputs/controls.

Application B: System Identification under Uncertain-But-Bounded Observation Errors

[J&Kotsalis&N'22]

• **Problem:** *Given noisy observations of states* $v_t \in \mathbb{R}^d$ *and inputs* $r_t \in \mathbb{R}^h$ *of linear time-invariant dynamical system*

$$v_{t+1} = X\left[\begin{array}{c} v_t \\ r_t \end{array}\right], \, 0 \leq t < N$$

on finite time horizon, we want to recover the image $B(X)$ *of* X *under a given linear mapping.*

• Our observations are

$$\begin{array}{rcl} \overline{u}_{ti} & = & [v_t]_i - \xi_{ti}, \, 1 \leq i \leq d, 0 \leq t \leq N, \\ \overline{u}_{ti} & = & [r_t]_{i-d} - \xi_{ti}, \, d < i \leq d+h, 0 \leq t \leq N-1, \end{array}$$

with observation errors ξ_{ti} obeing given bounds $\overline{\xi}_{ti}$: $|\xi_{ti}| \leq \overline{\xi}_{ti}$.

• After straightforward preprocessing of the problem, we arrive at the following

Situation: *Given matrix B, we want to recover the image $Bx \in \mathbb{R}^\nu$ of unknown vector $x \in \mathbb{R}^n$ known to satisfy, for some unknown $\zeta_s \in [-1,1]$, the system of linear equations*

$$\left[Q - \sum_{s=1}^{S} \zeta_s Q_s\right] x = q - \sum_{s=1}^{S} \zeta_s q_s \qquad (\mathcal{O})$$

$$\left[\begin{array}{l} \bullet\, Q,\, Q_s \in \mathbb{R}^{m\times n}: \text{ known matrices stemming from observations} \\ \bullet\, q, q_s \in \mathbb{R}^m: \text{ known vectors stemming from observations} \\ \bullet\, n = \dim X = d(d+h), \\ \bullet\, m = dN,\, S = (d+h)N + d \end{array}\right]$$

• We want to recover Bx by *linear estimate*

$$\widehat{w}_H = H^\top q,$$

and quantify the performance of a candidate estimate by its *$\mathcal{B}$-risk*

$$\mathrm{Risk}_{\mathcal{B}}[H] = \sup_x \{\|H^\top q - Bx\|_{\mathcal{B}} : x \text{ is compatible with } (\mathcal{O})\}$$

where $\|\cdot\|_{\mathcal{B}}$ is a given norm on $\mathbb{R}^\nu$, and compatibility means that x satisfies $(\mathcal{O})$ for some selection of $\zeta_s \in [-1,1]$.

Processing the problem

A. We associate with $m\times\nu$ matrix H uncertain vector and matrix

$$\begin{array}{rcl}\mathcal{V}_0[H] &=& \left\{\sum_s \zeta_s H^\top q_s : |\zeta_s|\leq 1\,\forall \leq S\right\}\subset \mathbb{R}^\nu,\\ \mathcal{V}[H] &=& \left\{[H^\top Q - B] + \sum_s \zeta_s H^\top Q_s : |\zeta_s|\leq 1\,\forall s\leq S\right\}\subset \mathbb{R}^{\nu\times n}\end{array}$$

Given a norm $\|\cdot\|_{\mathcal{X}}$ on $\mathbb{R}^n$, it is easily seen that

$$\boxed{\begin{array}{c}x \text{ is compatible with } (\mathcal{O})\\ \Rightarrow \|H^\top q - Bx\|_{\mathcal{B}} \leq \Theta[H]\|x\|_{\mathcal{X}} + \Theta_0[H]\end{array}}$$

$$\left[\begin{array}{rcl}\Theta_0[H] &=& \max\limits_{v\in\mathcal{V}_0[H]} \|v\|_{\mathcal{B}} \text{ – robust } \mathcal{B}\text{-norm of } \mathcal{V}_0[H]\\ \Theta[H] &=& \max\limits_{V\in\mathcal{V}[H]} \|V\|_{\mathcal{X},\mathcal{B}} \text{ – robust } \mathcal{X},\mathcal{B}\text{-norm of } \mathcal{V}[H]\end{array}\right]$$

B. We associate with a matrix $E\in\mathbb{R}^{m\times n}$ uncertain vector and matrix

$$\begin{array}{rcl}\mathcal{W}_0[E] &=& \left\{\sum_s\zeta_sE^\top q_s : |\zeta_s|\leq 1\,\forall s\leq S\right\}\subset\mathbb{R}^n,\\ \mathcal{W}[E] &=& \left\{[E^\top Q-I_n]+\sum_s\zeta_sE^\top Q_s : |\zeta_s|\leq 1\,\forall s\leq S\right\}\subset\mathbb{R}^{n\times n}\end{array}$$

and set

$$\begin{array}{rcl}\Upsilon_0[E] &=& \max\limits_{w\in\mathcal{W}_0[E]}\|w\|_{\mathcal{X}} \text{ – robust } \mathcal{X}\text{-norm of } \mathcal{W}_0[E]\\ \Upsilon[E] &=& \max\limits_{W\in\mathcal{W}[E]}\|W\|_{\mathcal{X},\mathcal{X}} \text{ – robust } \mathcal{X},\mathcal{X}\text{-norm of } \mathcal{W}[E]\end{array}$$

• **Theorem II** (i) *Let E be such that $\Upsilon[E]<1$. Then for all x compatible with $\mathcal{O}$ it holds*

$$\|x\|_{\mathcal{X}}\leq\frac{\Upsilon[E]}{1-\Upsilon[E]}\|E^\top q\|_{\mathcal{X}}+\frac{\Upsilon_0[E]}{1-\Upsilon[E]}.$$

Invoking **A**, we conclude that

(ii) *For every $H\in\mathbb{R}^{\nu\times n}$ the $\mathcal{B}$-risk of the linear estimate $H^\top q$ satisfies*

$$\begin{array}{c}\mathrm{Risk}_{\mathcal{B}}[H]\leq\mathrm{Opt}=\min_H\{\Gamma\Theta[H]+\Theta_0[H]\}\\ \Gamma=\min_E\left\{\frac{\Upsilon[E]}{1-\Upsilon[E]}\|E^\top q\|_{\mathcal{X}}+\frac{\Upsilon_0[E]}{1-\Upsilon[E]} : \Upsilon[E]<1\right\}\end{array}$$

- Assume that $\|x\|_{\mathcal{X}} = \max\limits_{k\leq K}\|P_kx\|_2$ and the norm *conjugate to* $\|y\|_{\mathcal{B}}$ is $\|y\|_{\mathcal{B}}^* = \max\limits_{\ell\leq L}\|Q_\ell^\top y\|_2$. Then our machinery for upper-bounding robust norms provides efficiently computable convex in H, E upper bounds $\overline{\Theta}_0[H]$, $\overline{\Theta}[H]$, $\overline{\Upsilon}_0[E]$, $\overline{\Upsilon}[E]$ on $\Theta_0[H]$,...,$\Upsilon[E]$. The bounds are tight within the factor $\leq O(1)\sqrt{d\ln(K+1)\ln(L+1)}$.
- Utilizing bounds $\overline{\Theta}_0[H]$,...,$\overline{\Upsilon}[E]$, we arrive at the following strategy:
  - Solving convex optimization problems

$$\min_E\left\{\frac{\gamma}{1-\gamma}\|E^\top q\|_{\mathcal{X}} + \frac{\overline{\Upsilon}_0[E]}{1-\gamma} : \overline{\Upsilon}[E] \leq \gamma\right\}$$

  along a grid of values of $\gamma\in[0,1)$, we, with luck, build an upper bound $\overline{\Gamma}$ on Γ. Luck is guaranteed when observation noises are small and observation horizon N is not too small.
  - We solve convex optimization problem

$$\mathrm{Opt} = \min_H\left\{\overline{\Gamma}\,\overline{\Theta}[H] + \overline{\Theta}_0[H]\right\}.$$

  An optimal solution H_* yields a “presumably good” linear estimate $H_*^\top q$ of $B(X)$, with $\|\cdot\|_{\mathcal{B}}$-magnitude of the recovery error $\leq \mathrm{Opt}$.

How It Works
Recovering Boeing 747 dynamics

- **The goal** is to recover full dynamics ($B = I$) of Boeing 747 from observations of states $x_t \in \mathbb{R}^4$ for $0 \leq t \leq N = 12$ and inputs $r_t \in \mathbb{R}^4$ for $0 \leq t < 12$.

In our experiments,

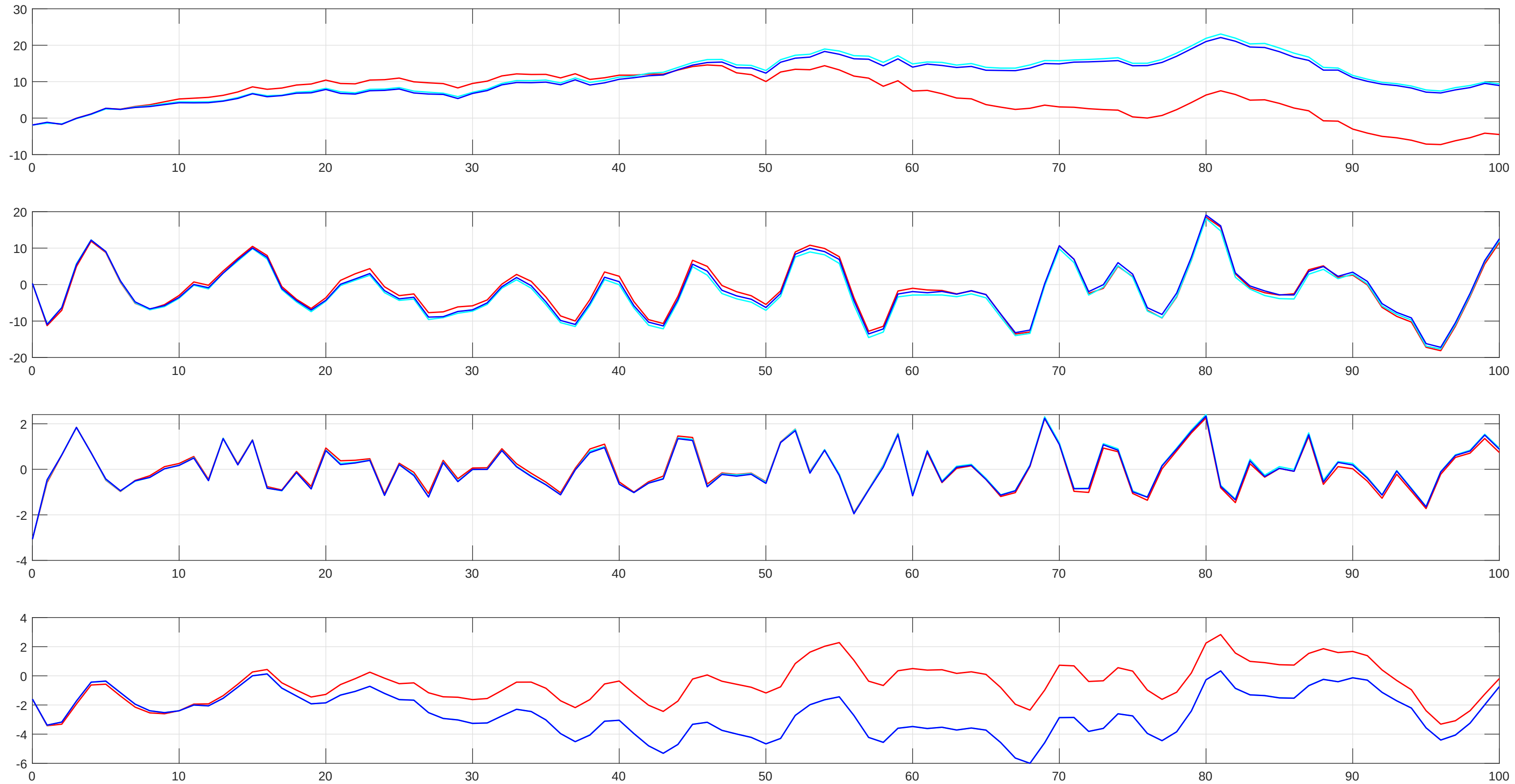
- initial state and inputs were drawn at random
- observation noises were bounded in a semi-relative scale: observation r' of a real r satisfies $|r' - r| \leq \epsilon \max[|r|, 1]$, with known to us error level ϵ
- We used $B = I$, $\|\cdot\|_{\mathcal{B}} = \|\cdot\|_2$, $\|\cdot\|_{\mathcal{X}} = \|\cdot\|_2$.

Results: Simple LS recovery (*as if* there were no observation errors) vs robust linear recovery

ε	0.001	0.002	0.003	0.004	0.005
Least Squares	0.017/12.1	0.035/10.0	0.071/12.6	0.078/11.0	0.061/12.1
Robust Recovery	0.021/1.9	0.042/1.5	0.085/1.6	0.096/1.5	0.074/1.7
ε	0.006	0.007	0.008	0.009	0.010
Least Squares	0.118/12.2	0.124/11.4	0.123/10.4	0.124/12.1	0.145/13.2
Robust Recovery	0.131/1.4	0.163/1.4	0.162/1.6	0.157/1.8	0.189/1.6

Blue: $\|\cdot\|_2$ recovery error Red: ratio of upper error bound to actual error
median over 10 simulations per value of ε

States of the actual and of the recovered systems vs. time.



Actual dynamics (red), robust recovery (blue), and LS recovery (cyan).
Relative observation error $\varepsilon = 0.01$.
Actual $\|\cdot\|_2$ recovery errors are 0.1968 for LS and 0.1394 for robust recovery.

0.20

Application C: Signal Recovery from Indirect Observations under Uncertainty in Sensing Matrix

- **Situation and goal:** *Given observation*

$$\omega = A[\eta]x + \xi$$

we want to recover linear image $Bx \in \mathbb{R}^\nu$ *of unknown signal* x *known to belong to a given signal set* $\mathcal{X}$.

Here

- $\mathcal{X} = \{x \in \mathbb{R}^n : \exists t \in \mathcal{T} : x^\top T_k x \le t_k, k \le K\}$ is a basic ellitope
- $A[\eta] = A_{\mathrm{nom}} + \sum_{s=1}^S \eta_s A_s + \sum_{t=1}^T L_t^\top \eta^t R_t \in \mathbb{R}^{m\times n}$ is sensing matrix affected by non-observable *perturbation* $\eta = \{\eta_s \in \mathbb{R}, s \le S, \eta^t \in \mathbb{R}^{p_t \times q_t}, t \le T\}$
- ξ *is random observation noise with distribution* P_x *assumed to be sub-Gaussian with parameters* $0, \sigma^2 I_m$ *for every* $x \in \mathcal{X}$: $\mathbf{E}_{\xi\sim P_x}\{\mathrm{e}^{h^\top \xi}\} \le \mathrm{e}^{\frac{\sigma^2}{2}h^\top h}$

• **We assume that**

either

- (random perturbations) $T=0$ and $\eta\in\mathbb{R}^S$ is sub-Gaussian with parameters $0, I_S$,

or

- (structured norm-bounded perturbations) η runs through *perturbation set*

$$\mathcal{U}=\{\eta: |\eta_s|\leq 1, s\leq S, \|\eta^t\|_{2,2}\leq 1, t\leq T\}$$

- the recovery error is measured in the norm $\|u\|=\max_{\ell\leq L}\sqrt{u^\top\Pi_\ell u}$ with $\Pi_\ell\succeq 0$, $\sum_\ell \Pi_\ell\succ 0$
- the performance of a candidate estimate $\widehat{w}(\omega)$ is quantified by its ϵ-*risk*

Random perturbations: $\mathrm{Risk}_\epsilon[\widehat{w}\mid\mathcal{X}] = \sup\limits_{x\in\mathcal{X}}\inf\{\rho: \mathrm{Prob}_{\xi,\eta}\{\|Bx-\widehat{w}(A[\eta]x+\xi)\|>\rho\}\leq\epsilon\}$
Uncertain-but-bounded perturbations: $\mathrm{Risk}_\epsilon[\widehat{w}\mid\mathcal{X}] = \sup\limits_{x\in\mathcal{X},\eta\in\mathcal{U}}\inf\{\rho: \mathrm{Prob}_{\xi\sim P_x}\{\|\widehat{w}(A[\eta]x+\xi)-Bx\|>\rho\}\leq\epsilon\}.$

Previous research:

• Linear estimates $\widehat{w}(\omega) = H^\top\omega$ originate from Kuks&Olman'71-72 and were studied, usually in the "no uncertainty in the sensing matrix" case, by many authors (Pinsker'81, Donoho et al'90, Donoho'94, Efroimovich&Pinsker'96, Efroimovich'08, Ibragimov&Khasminskii'81, Tsybakov'09, Wasserman'06, J&N'20,...)

• Random perturbations: significant research on linear regression with random errors in regressors (Bennani et al'88, Carroll&Ruppert'96, Fan&Truong'93, Gleser'81, Kukush et al'05, Stewart'90, Van Huffel&Lemmerling'13,...) usually addressed via total least squares, or signal processing with stochastic errors in sensing matrix (Cavalier&Hengratner'05, Cavalier&Raimondo'07, Efroimovich&Koltchinskii'01, Hall&Horowitz'05, Hoffman&Reiss'08, Marteau'06,...)

• Uncertain-but-bounded perturbations: an extension of the intensively studied problem of solving linear systems with uncertain data (Cope&Rust'79, Higham'02, Kreinovich et al'93, Nazin&Polyak'05, Neumaier'90, Oettli&Prager'64, Polyak'03,...) and system identification with imprecisely measured states (Bertcekas&Rodes'71, Casini et al'14, Cerone'93, J&Kotsalis&N'22, Kurzhansky&Valyi'97, Matasov'98, Milanese et al'13, Nazin&Polyak'07, Tempo&Vicino'9,...)

• Our contributions: a complementary computation-friendly approach to synthesis of "presumably good" linear estimates equipped with reasonably tight risk bounds.

✓ While we do *not* know whether our "presumably good" linear estimates are near-minimax-optimal among *all* estimates, linear and nonlinear alike, this provably is the case when the observation noise ξ is Gaussian, and uncertainty in sensing matrix is small.

C.I: Random Perturbations in Sensing Matrix

$$\omega = A[\eta]x + \xi \;???\Rightarrow???\; \widehat{w} := H^\top \omega \approx Bx$$
$$A[\eta] = A_{\rm nom} + \sum_s \eta_s A_s,\;\; x \in \mathcal{X} = \{x \in \mathbb{R}^n : \exists t \in \mathcal{T} : x^\top T_k x \preceq t_k, k \leq K\}$$
$$\mathrm{Risk}_\epsilon[\widehat{w}|\mathcal{X}] = \sup_{x\in\mathcal{X}} \inf \left\{\rho :\; \mathrm{Prob}_{\xi,\eta}\{\|Bx - \widehat{w}(A[\eta]x+\xi)\| > \rho\} \leq \epsilon\right\}$$
$$\xi \sim \mathcal{SG}(0,\sigma^2 I_m),\; \eta \sim \mathcal{SG}(0, I_S),\; \|y\| = \max_{\ell\leq L}\sqrt{u^\top \Pi_\ell u}$$

Proposition. *Synthesis of presumably good linear estimate* $\widehat{w}^H(\omega) = H^\top\omega$ *reduces to solving the convex optimization problem*

$$\min_{H\in\mathbb{R}^{m\times\nu}} \mathfrak{R}[H]$$

$$\mathfrak{R}[H] = \min_{\substack{\lambda_\ell,\mu^\ell,\kappa^\ell,\\ \varkappa^\ell,\rho,\varrho}} \left\{ \left[1+\sqrt{2\ln(2L/\epsilon)}\right]\left[\sigma \max_{\ell\leq L}\|H\Pi_\ell^{1/2}\|_{\rm Fro} + \rho\right] + \varrho : \begin{array}{l} \mu^\ell \geq 0, \varkappa^\ell \geq 0,\; \lambda_\ell + \phi_{\mathcal{T}}(\mu_\ell) \leq \rho, \kappa_\ell + \phi_{\mathcal{T}}(\varkappa^\ell) \leq \varrho,\; \ell \leq L \\ \left[\begin{array}{c|c} \lambda_\ell I_{\nu S} & \frac{1}{2}\left[\Pi_\ell^{1/2}H^\top A_1; \ldots; \Pi_\ell^{1/2}H^\top A_S\right] \\ \hline \frac{1}{2}\left[A_1^\top H \Pi_\ell^{1/2}, \ldots, A_S^\top H\Pi_\ell^{1/2}\right] & \sum_k \mu_k^\ell T_k \end{array}\right] \succeq 0,\; \ell \leq L \\ \left[\begin{array}{c|c} \kappa^\ell I_\nu & \frac{1}{2}\Pi_\ell^{1/2}[B - H^\top A_{\rm nom}] \\ \hline \frac{1}{2}[B - H^\top A_{\rm nom}]^\top \Pi_\ell^{1/2} & \sum_k \varkappa_k^\ell T_k \end{array}\right] \succeq 0,\; \ell \leq L \end{array} \right\}$$

where $\phi_{\mathcal{T}}(\lambda) = \max_{t\in\mathcal{T}} \lambda^\top t$ *is the support function of* $\mathcal{T}$. *For a candidate* H, *one has* $\mathrm{Risk}_\epsilon[\widehat{w}^H|\mathcal{X}] \leq \mathfrak{R}[H]$.

$$A[\eta] = A_{\rm nom} + \sum_s \eta_s A_s,\ x \in \mathcal{X} = \{x \in \mathbb{R}^n : \exists t \in \mathcal{T} : x^\top T_k x \preceq t_k, k \leq K\}$$
$$\xi \sim \mathcal{SG}(0, \sigma^2 I_m),\ \eta \sim \mathcal{SG}(0, I_S),\ \|u\| = \max_{\ell \leq L} \sqrt{u^\top \Pi_\ell u}$$

A modification. To make risk small, the observation noise ξ and the influence of uncertainty η should be small.

• With $\xi \sim \mathcal{SG}(0, \sigma^2 I_m)$, "small ξ" means small σ. With our normalization $\eta \sim \mathcal{SG}(0, I_S)$, "small η" means small perturbation matrices A_s.

However: In some situations, "small uncertainty" does *not* translate into small A_s's. For example, assume that random perturbation of sensing matrix zeros out some columns of the matrix, with small probability γ to zero out a particular column, as when taking picture through a window with frost pattern.

• **Remedy:** To replace sub-Gaussianity with second moment boundedness:

$$\mathbf{E}\{\xi\xi^\top\} \preceq \sigma^2 I_m,\ \mathbf{E}\{\eta\eta^\top\} \preceq I_S \tag{$*$}$$

and allow for M*-repeated observations*

$$\omega^M = \{\omega_\mu = A[\eta_\mu]x + \xi_\mu,\ \mu \leq M\},$$

with i.i.d. pairs $[\xi_\mu, \eta_\mu]$ obeying $(*)$.

The construction (utilizing, in particular, the results of [Minsker'15] on geometric medians) is as follows:

• For each $\ell \leq L$, we compute optimal solutions $H_\ell \in \mathbb{R}^{m\times\nu}$ to the convex optimization problems

$$\widetilde{\mathfrak{R}}_\ell[H] \;=\; \min_{\lambda,\upsilon,\kappa,\varkappa}\left\{\sigma\|H\Pi_\ell^{1/2}\|_{\rm Fro}+\lambda+\phi_{\mathcal{T}}(\upsilon)+\kappa+\phi_{\mathcal{T}}(\varkappa):\right.$$
$$\upsilon\geq 0,\varkappa\geq 0,\left[\begin{array}{c|c}\kappa I_\nu & \frac{1}{2}\Pi_\ell^{1/2}[B-H^\top A_{\rm nom}]\\ \hline \frac{1}{2}[B-H^\top A_{\rm nom}]^\top\Pi_\ell^{1/2} & \sum_k\varkappa_kT_k\end{array}\right]\succeq 0$$
$$\left.\left[\begin{array}{c|c}\lambda I_{\nu S} & \frac{1}{2}\left[\Pi_\ell^{1/2}H^\top A_1;\dots;\Pi_\ell^{1/2}H^\top A_S\right]\\ \hline \frac{1}{2}\left[A_1^\top H\Pi_\ell^{1/2},\dots,A_S^\top H\Pi_\ell^{1/2}\right] & \sum_k\upsilon_kT_k\end{array}\right]\succeq 0\right\}$$

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

  • Given H_ℓ and observations ω_μ we compute linear estimates $w_\ell(\omega_\mu)=H_\ell^\top\omega_\mu$, $\ell=1,\dots,L$, $\mu=1,\dots,M$;

  • We define vectors $z_\ell\in\mathbb{R}^\nu$ as geometric medians of $w_\ell(\omega_\mu)$:

$$z_\ell(\omega^M)\in\mathop{\rm Argmin}_z\sum_{\mu=1}^M\|\Pi_\ell^{1/2}(w_\ell(\omega_\mu)-z)\|_2,\;\ell=1,\dots,L.$$

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

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

[$\widehat{w}^{(r)}(\omega^M)=0$ when $\mathcal{W}(\omega^M)=\emptyset$].

$$\omega^M\ \{\omega_\mu = A[\eta_\mu]x + \xi_\mu\}_{\mu\le M}\ ???\Rightarrow???\ \widehat{w}(\omega^M)\approx Bx$$
$$A[\eta] = A_{\rm nom} + \sum_s \eta_s A_s,\ x\in\mathcal{X} = \{x\in\mathbb{R}^n : \exists t\in\mathcal{T} : x^\top T_k x \preceq t_k, k\le K\}$$
$$\mathrm{Risk}_\epsilon[\widehat{w}|\mathcal{X}] = \sup_{x\in\mathcal{X}} \inf\left\{\rho:\ \mathrm{Prob}_{\xi^M,\eta^M}\{\|Bx - \widehat{w}(\{[A[\eta_\mu]x+\xi_\mu.\mu\le M\})\| > \rho\}\le\epsilon\right\}$$
$$\mathbf{E}\{\xi\xi^\top\}\preceq\sigma^2 I_m,\ \mathbf{E}\{\eta\eta^\top\}\preceq I_S,\ \|y\| = \max_{\ell\le L}\sqrt{u^\top\Pi_\ell u}$$

Proposition *One has*

$$\sup_{x\in\mathcal{X}} \mathbf{E}_{\eta_\mu,\xi_\mu}\left\{\|\Pi_\ell^{1/2}(w_\ell(\omega_\mu) - Bx)\|_2^2\right\} \le \widetilde{\mathfrak{R}}_\ell^2[H_\ell],\ \ell\le L,$$

and

$$\mathrm{Prob}\left\{\|\Pi_\ell^{1/2}(z_\ell(\omega^M) - Bx)\|_2 \ge 4\widetilde{\mathfrak{R}}_\ell[H_\ell]\right\} \le e^{-0.1070M},\ \ell\le L.$$

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

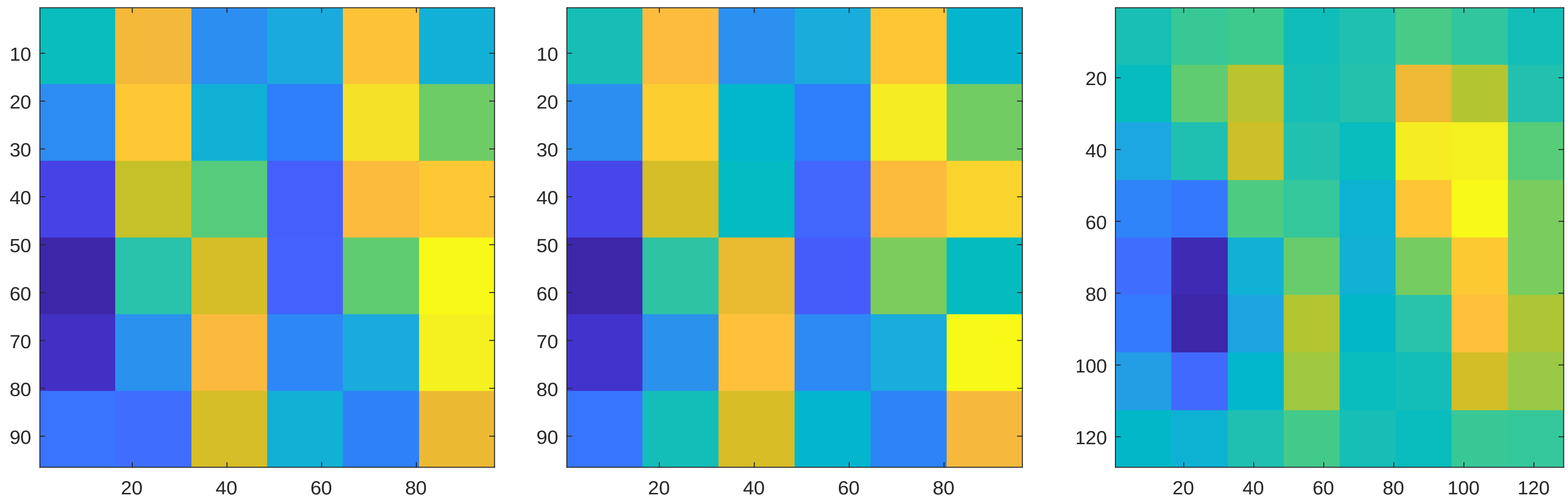
$$\mathrm{Risk}_\epsilon[\widehat{w}^{(r)}|\mathcal{X}] \le 8\max_{\ell\le L}\widetilde{\mathfrak{R}}_\ell[H_\ell].$$

How It Works: Looking through Frosty Glass

- **Problem:** Entries in 6×6 image x known to satisfy

$$\|\Delta x\|_\infty \le 1 \qquad [\Delta: \text{ discrete Laplace operator}]$$

are multiplied by i.i.d. Bernoulli r.v.'s taking value 0 with probability γ. The resulting image is convolved with 3×3 Gaussian kernel, and the resulting 8×8 image is observed in white Gaussian noise of intensity σ. Given the observation, we want to recover the original image.



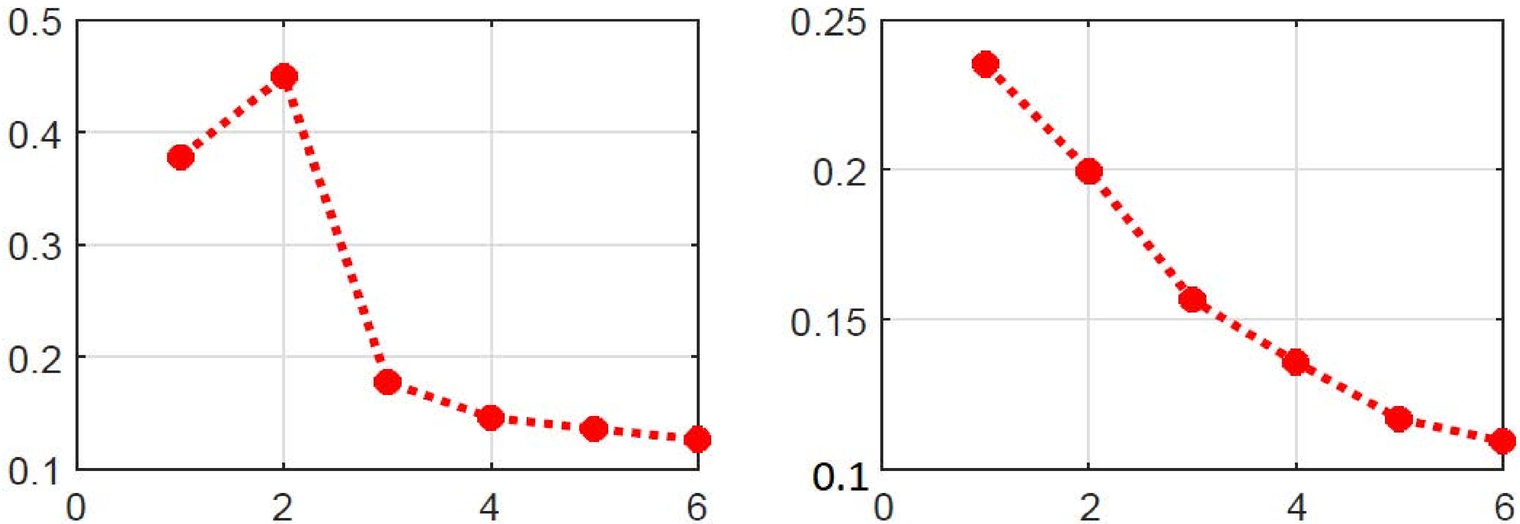
True image x, $\|x\|_{\text{Fro}} = 5.80$ Recovery $\widehat{w}$, $\|x - \widehat{w}\|_{\text{Fro}} \approx 2.03$ Observation

In this single-observation expieriment, multiplication by $\mathcal{D}$ zeroed out 3 of 36 entries in x

0.28

Looking through Frosty Glass (continued)

Effect of M-repeated observations



Frobenius norms of recovery errors vs M, data over 100 simulations; Left: means Right: medians

- Noise intensity $\sigma = 0.005$, $\mathrm{Cond}(A_{\mathrm{nom}}) \approx 164$, average Frobenius norm of signal 9.06
- Probability to zero out a particular entry in signal $\gamma = 0.005$ (0.6 suppressed pixels per signal at average)

C.II: Uncertain-But-Bounded Perturbations in Sensing Matrix

$$\omega = A[\eta]x + \xi \;??? \Rightarrow ???\; \widehat{w} := H^\top \omega \approx Bx$$
$$A[\eta] = A_{\rm nom} + \sum_{s\le S}\eta_s A_s + \sum_{t\le T} L_t^\top \eta^t R_t,\;\; x\in\mathcal{X} = \{x\in\mathbb{R}^n : \exists t\in\mathcal{T} : x^\top T_k x \preceq t_k, k\le K\}$$
$$\mathrm{Risk}_\epsilon[\widehat{w}|\mathcal{X}] = \sup_{x\in\mathcal{X},\eta\in\mathcal{U}} \inf\left\{\rho : \mathrm{Prob}_{\xi\sim P_x}\{\|\widehat{w}(A[\eta]x+\xi) - Bx\| > \rho\} \le \epsilon\right\}$$
$$\xi\sim\mathcal{SG}(0,\sigma^2 I_m),\, \mathcal{U} = \{\eta : |\eta_s|\le 1, s\le S, \eta^t\in\mathbb{R}^{p_t\times q_t}, \|\eta^t\|_{2,2}\le 1, t\le T\},\; \|u\| = \max_{\ell\le L}\sqrt{u^\top \Pi_\ell u}$$
$$\phi_{\mathcal{T}}(\lambda) = \max_{t\in\mathcal{T}} \lambda^\top t$$

Proposition. *The ϵ-risk of a linear estimate $\widehat{w}^H(\omega) = H^\top\omega$ can be tightly upper-bounded by the sum of 3 functions of H:*

A. Upper $(1-\epsilon)$-quantile of $\|H^\top\xi\|$: $\alpha(H) = \min\{t : \mathrm{Prob}\{\|H^\top\xi\| > t\} \le \epsilon\}$

B. Nominal bias: $\beta(H) = \max_{x\in\mathcal{X}} \|(B - H^\top A_{\rm nom})x\|$

C. Uncertainty-induced bias: $\gamma(H) = \max\limits_{\eta\in\mathcal{U}, x\in\mathcal{X}} \|[\sum_s \eta_s H^\top A_s + \sum_t [L_t H]^\top \eta^t R_t]x\|$.

All three functions admit reasonably tight efficiently computable convex in H upper bounds. Specifically,

- $\alpha(H) \le [1+\sqrt{2\ln(L/\epsilon)}]\sigma \max_{\ell\le L}\sqrt{\mathrm{Tr}(H\Pi_\ell H^\top)}$

– tight within the factor $O(\ln(L+1))$, at least for $\xi\sim\mathcal{N}(0,\sigma^2 I_m)$

- $\beta(H) \le \overline{\beta}(H) := \max\limits_{\ell\le L} \mathfrak{r}_\ell(H)$,

$$\mathfrak{r}_\ell(H) = \min_{\mu,\lambda}\left\{\lambda + \phi_{\mathcal{T}}(\mu) :\; \mu\ge 0,\; \left[\begin{array}{c|c} \lambda I_\nu & \frac{1}{2}\Pi_\ell^{1/2}[B - H^\top A_{\rm nom}] \\ \hline \frac{1}{2}[B - H^\top A_{\rm nom}]^\top \Pi_\ell^{1/2} & \sum_k \mu_k T_k \end{array}\right] \succeq 0\right\}$$

– tight within the factor $3\sqrt{\ln(3K)}$

• $\gamma(H) \leq \overline{\gamma}(H) := \max_{\ell \leq L} \mathfrak{s}_\ell(H)$,

$$\begin{array}{rcl}\mathfrak{s}_\ell(H) & = & \min\limits_{\mu,\upsilon,\lambda,U_s,V_s,U^t,V^t} \Big\{ \frac{1}{2}[\mu + \phi_{\mathcal{T}}(\upsilon)] : \ \mu \geq 0, \upsilon \geq 0, \lambda \geq 0 \\ & & \left[\begin{array}{c|c} U_s & \Pi_\ell^{1/2} H^\top A_s \\ \hline A_s^\top H \Pi_\ell^{1/2} & V_s \end{array}\right] \succeq 0, \ s \leq S, \ \left[\begin{array}{c|c} U^t & \Pi_\ell^{1/2} H^\top L_t^\top \\ \hline L_t H \Pi_\ell^{1/2} & \lambda_t I_{p_t} \end{array}\right] \succeq 0, \ t \leq T \\ & & \mu I_\nu - \sum_s U_s - \sum_t U^t \succeq 0, \ \ \sum_k \upsilon_k T_k - \sum_s V_s - \sum_t \lambda_t R_t^\top R_t \succeq 0 \Big\}\end{array}$$

– tight within the factor $\varkappa(K) \max[\vartheta(2\kappa), \pi/2]$,

$$\kappa = \max_{s \leq S} \mathrm{Rank}\,(A_s), \ \ \varkappa(K) = \left\{ \begin{array}{ll} 1, & K = 1, \\ \frac{5}{2}\sqrt{\ln(2K)}, & K > 1, \end{array} \right.$$

• *For every* H*, we have*

$$\mathrm{Risk}_\epsilon[\widehat{w}^H | \mathcal{X}] \leq \overline{\alpha}(H) + \overline{\beta}(H) + \overline{\gamma}(H).$$

Presumably good linear estimate is obtained by minimizing the right hand side in H*.*

How It Works

- **Situation:** *We observe the noisy image of imprecise convolution*

$$\omega_t = \sum_{0\leq\tau<d}[\chi_\tau + \rho\,\eta_{t,\tau}]x_{t-\tau} + \xi_t,\quad 1 \leq t < n+d$$

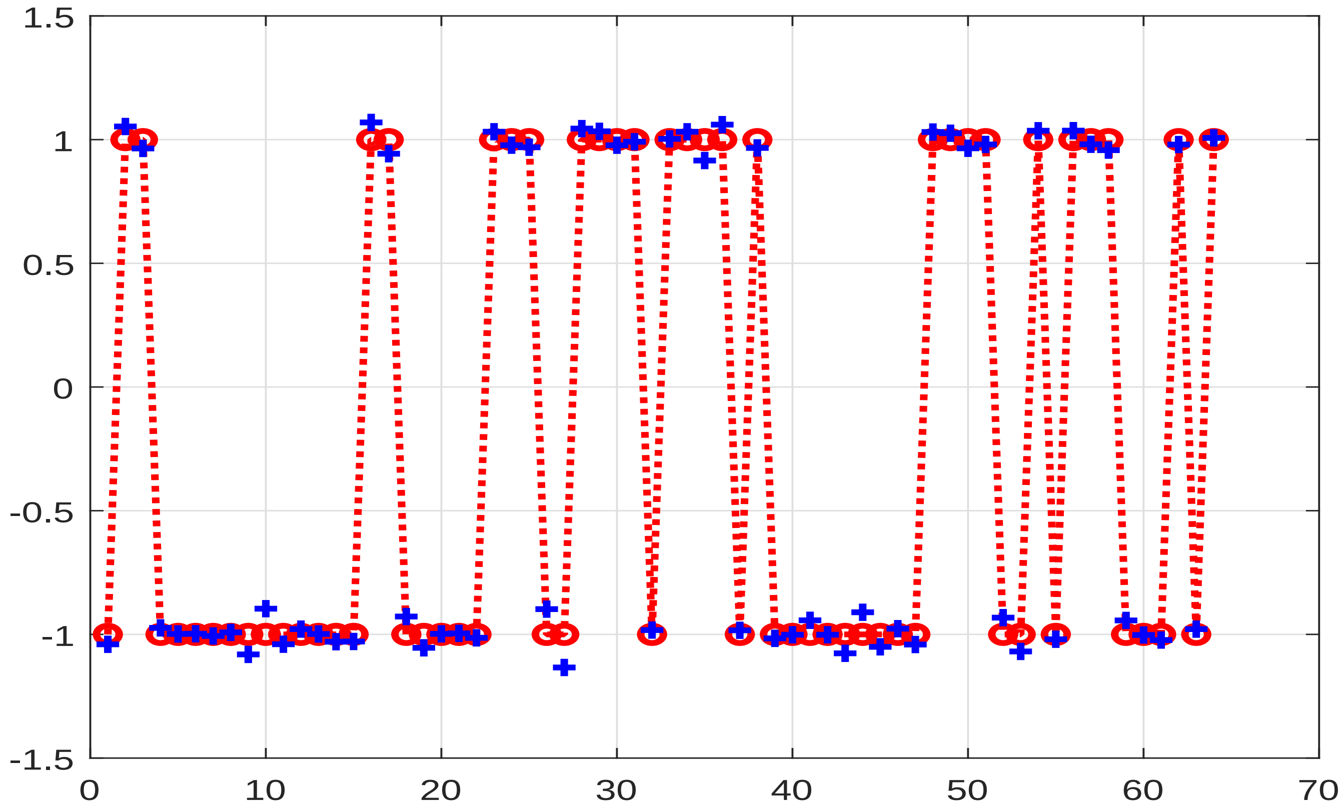
  - $x = \{x_s : s = 0, \pm 1, \pm 2, ...\}$: unknown signal known to have $\|x\|_\infty \leq 1$, with $x_t = 0\,\forall t \notin \{1, ..., n\}$,
  - $\chi \in \mathbb{R}^d$ – nominal kernel
  - $\eta_{t,\tau} \in [-1, 1]$ – perturbations
    $\Rightarrow$ *structured norm-bounded uncertainty with* $S = 0$, $T = nd$
  - ρ – perturbation level
  - $\xi_t \sim \mathcal{N}(0, \sigma^2)$ – independent across t observation noises

and want to recover x*, measuring the recovery error in* $\|\cdot\|_\infty$*.*

- **SetUp:** $n = 64$, $d = 8$, $\sigma = 0.001$, $\rho = 0.0005$, $\mathrm{Cond}(A_{\mathrm{nom}}) \approx 90$
- $\|\cdot\|_\infty$ **recovery errors**, data over 1,000 simulations

mean	median	max
0.119	0.116	0.224

- 0.01**-risk:** ✓ upper bound: 0.833 ✓ lower bound 0.189
- **Sample recovery:**



○ - signal, + - recovery

References

[Bekri&J&N'23] Bekri, Y., Juditski, A., Nemirovski, A. On robust recovery of signals from indirect observations
https://arxiv.org/pdf/2309.06563.pdf

[Ben-Tal&Boyd&N'06] Ben-Tal, A., Boyd, S., Nemirovski, A. Extending Scope of Robust Optimization: Comprehensive Robust Counterparts of Uncertain Problems. *Mathematical Programming Ser. B* **107** (2006), 63-89.

[J&Kotsalis&N'22] Juditsky, A., Kotsalis, G., Nemirovski, A. Tight computationally efficient approximation of matrix norms with applications – *Open Journal of Mathematical Optimization* **3:7** (2022), 1-38.
https://arxiv.org/pdf/2110.04389.pdf

[Minsker'15] Minsker, S. Geometric median and robust estimation in Banach spaces. *Bernoulli* **21(4)** 2015, 2308– 2335

[Nesterov'98] Nesterov, Yu. Semidefinite relaxation and nonconvex quadratic optimization. *Optimization methods and software* **9(1-3)** (1998), 141–160.

[Nesterov'00] Nesterov, Yu. Global quadratic optimization via conic relaxation. In *R. Saigal, H. Wolkowicz, and L. Vandenberghe, editors, Handbook on Semidefinite Programming* 363–387, Kluwer Academic Publishers, 2000.