Risk Assessment for Sparse Optimization with Relaxation

Zhicheng Zhang and Yasumasa Fujisaki

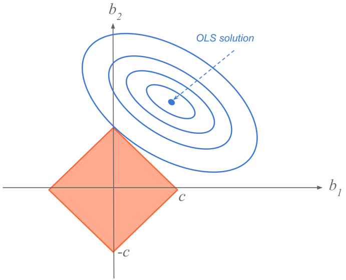
Graduate School of Information Science and Technology
Osaka University

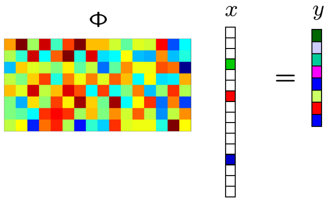
Nov. 18, 2023

@ISCIE SSS'23

Sparse Modeling

Lots of momentums for sparsity ... pictures from wiki





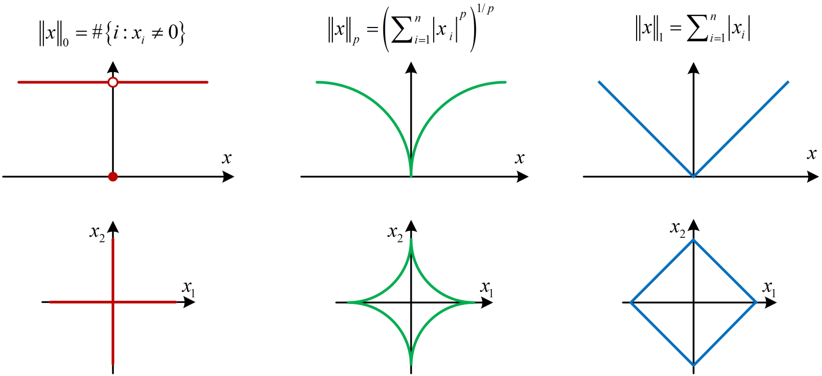
Lasso in statistics, compress sensing in signal processing

- In fact, uncertainty is ubiquitous: noise, model mismatch, data ...

■ Robustness plays an important role !

Sparsity: Exactness vs. Relaxation

- Exact ℓ_0 “quasi-norm”, approximated ℓ_p norm, convex ℓ_1 norm



■ Convex relaxation: nonconvex ℓ_0 “norm” $\Rightarrow$ convex ℓ_1 norm

Sparse Optimization Meets Uncertainty

Chance Constrained Sparse Optimization Problem

$$(\mathrm{CCSOP}_{\epsilon}^{0}): \quad \min_{x\in\mathcal{X}} \quad \|x\|_0$$
$$\text{s.t.} \quad \mathbb{P}\{q \in \mathcal{Q} : h(x,q) \le 0\} \ge 1-\epsilon,$$

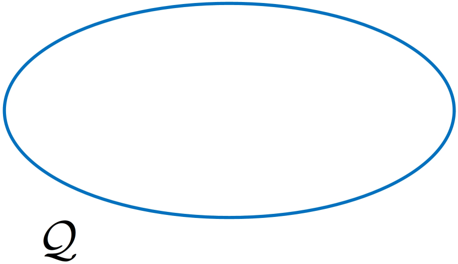
where $x \in \mathcal{X} \subseteq \mathbb{R}^n$, $q \in \mathcal{Q} \subseteq \mathbb{R}^{n_q}$, $h(u,q) : \mathbb{R}^n \times \mathbb{R}^{n_q} \to \mathbb{R}$

- $h(x,q)$ is *convex* in x for all $q \in \mathcal{Q}$, and *bounded* in q for all x.
- $\epsilon \in (0,1)$ is a stipulated risk level, e.g., $\epsilon = 0.05$.
  - The violation of probability (or risk) is defined as

$$V(x) \triangleq \mathbb{P}\{q \in \mathcal{Q} : h(x,q) > 0\} \quad \Rightarrow \quad V(x) \le \epsilon$$

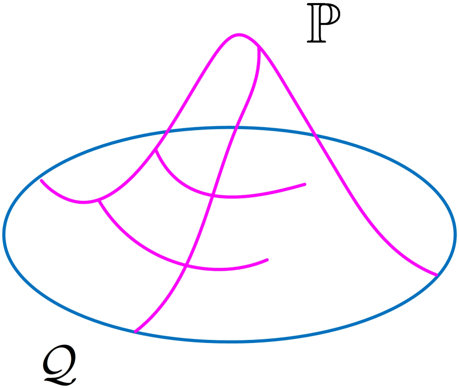
- Problem is well-defined, while it is computationally intractable...
  - ⊘ **Nonconvex** program, NP hard, Multiple integral calculation, etc.

Scenario Approach: Lifting Chance Constraints



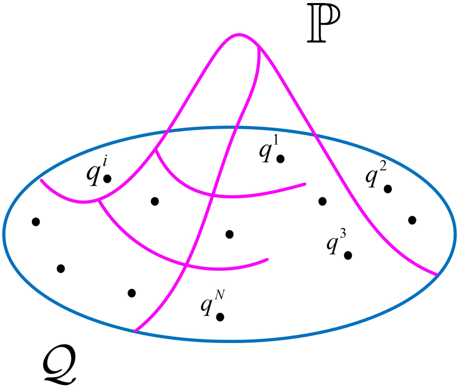
Worst-case uncertainty: $h(x, q) \leq 0$, for all $q \in \mathcal{Q}$

Scenario Approach: Lifting Chance Constraints



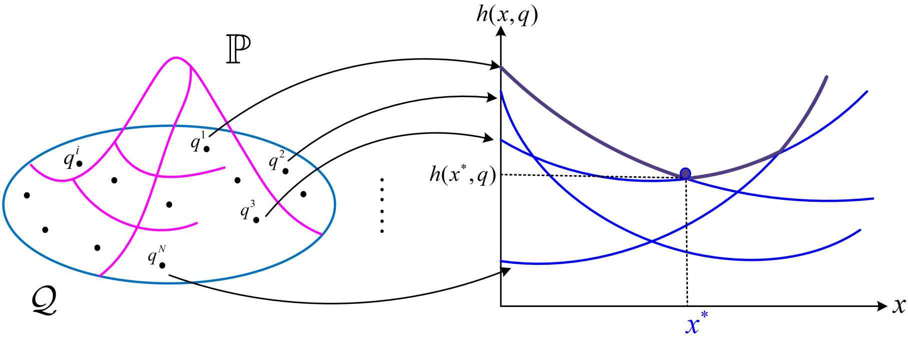
Chance-constrained uncertainty: $\mathbb{P}\{q \in \mathcal{Q} : h(x, q) \leq 0\}$

Scenario Approach: Lifting Chance Constraints



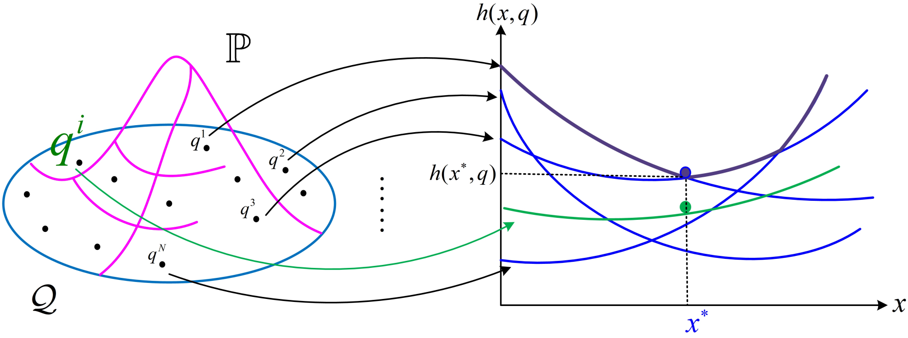
Sampled uncertainty: randomly i.i.d. generate N scenarios from the probability
e.g., scenario approach, sample average approximation, Monte-Carlo sampling

Scenario Approach: Lifting Chance Constraints



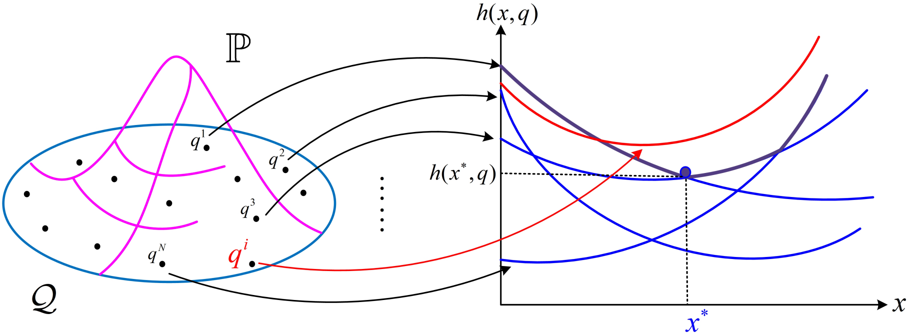
$$\{q^1, \cdots, q^N\} \rightsquigarrow \{h(x, q^1), \ldots, h(x, q^N)\} \Leftrightarrow \bigcap_{i=1}^{N} h(x, q^i) \text{ i.e., Feasibility}$$

Scenario Approach: Lifting Chance Constraints



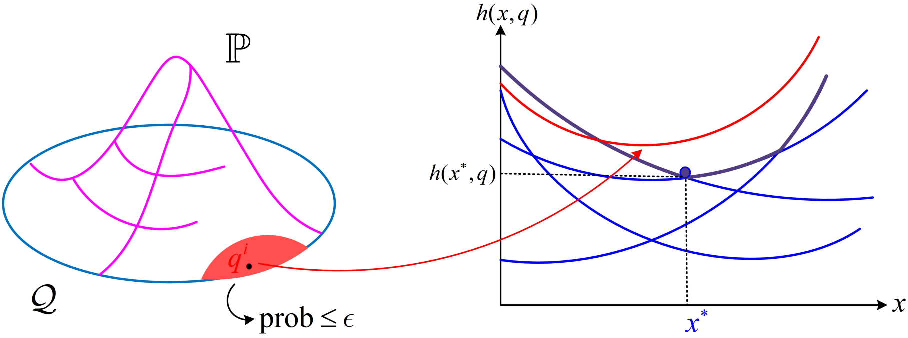
In general, for most of scenarios q^i's $\rightsquigarrow$ $h(x, q^i) \leq h(x^*, q)$

Scenario Approach: Lifting Chance Constraints



While, for some of scenarios q^i's $\rightsquigarrow$ $h(x, q^i) > h(x^*, q)$, i.e., violation

Scenario Approach: Lifting Chance Constraints



A small violation is allowed, e.g., $\epsilon = 1\%$, $V(x) \leq \epsilon$

Sparse Random Convex Program

Scenario Sparse Convex Optimization Problem [Zhang & Fujisaki, ISCIE SSS'21]

$$(\mathrm{SSCOP}_N^1): \quad \min_{x \in \mathcal{X}} \quad \|x\|_1 \quad \text{s.t.} \quad h(x, q^i) \leq 0, \quad i = 1, \ldots, N.$$

Mild Assumptions

- Assumption 1: Optimal solution x_N^* exists and is unique.
- Assumption 2: For every $x \in \mathcal{X}$, $\mathbb{P}\{q : h(x, q) = 0\} = 0$.

- $V(x_N^*)$ is dominated by a Beta distribution, i.e., [Calafiore & Campi, IEEE TAC'08]

$$\mathbb{P}^N\{V(x_N^*) \leq \epsilon\} \geq 1 - \beta, \quad \beta = \sum_{i=0}^{n-1} \binom{N}{i} \epsilon^i (1-\epsilon)^{N-i}$$

- A priori and explicit sample complexity: $N \geq \frac{2}{\epsilon}\left(\ln \frac{1}{\beta} + n\right)$

Sparse Optimization with Relaxation

This talk concerns on:

- Balance the "*Sparse Performance*" and the associated "*Risk*".
- A *posterioi* probabilistic robustness guarantees

Relaxed Scenario-based Sparse Convex Optimization Problem

$$
(\mathrm{SSCOP}_N^\rho): \quad \min_{x\in\mathcal{X},\xi^i\geq 0} \quad \|x\|_1 + \rho\sum_{i=1}^{N}\xi^i
$$
$$
\text{s.t.} \quad h(x,q^i)\leq \xi^i, \quad i=1,\ldots,N,
$$

- (SSCOP_N^ρ) has $n+N$ decision variables.
- The "relaxed" slack variables $\xi^i\geq 0$, $i=1,\cdots,N$ are "regrets".
- Weight ρ: balance the sparse cost and the violated constraints
  - $\rho\to 0$ non regret , $\rho = 1/N$ empirical regret, $\rho\to\infty$ infinite regret

A General Theory for Risk Assessments

Lemma 1 (Violation) [Garatti & Campi, Math. Program.'22]

Under Assumptions 1 and 2, consider problem (SSCOP_N^{ρ}), given a confidence $\beta \in (0,1)$, the risk $V(x_N^*)$ is evaluated as follows

$$\mathbb{P}^N\{\underline{\epsilon}(s_N^*) \leq V(x_N^*) \leq \overline{\epsilon}(s_N^*)\} \geq 1-\beta$$

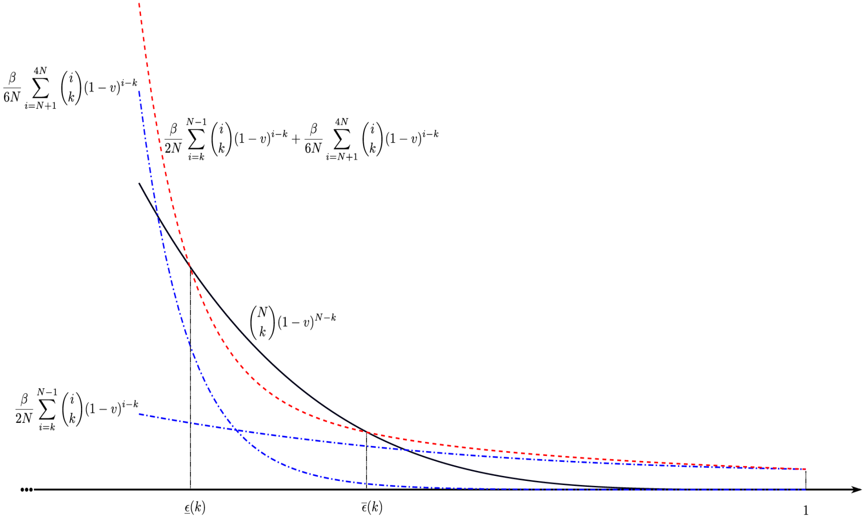
where $\underline{\epsilon}(\cdot) = \max\{0, 1-\overline{t}(k)\}$, $\overline{\epsilon}(\cdot) = \max\{0, 1-\underline{t}(k)\}$, s_N^* is the number of samples q^i for which $h(x, q^i) \geq 0$ at $x = x_N^*$, and for $k = 0, 1, \cdots, N-1$ the pair $\{\underline{t}(k), \overline{t}(k)\}$ is the solution of the polynomial equation in t variable

$$\mathfrak{B}_N(t;k) = \frac{\beta}{2N}\sum_{j=k}^{N-1}\mathfrak{B}_j(t;k) + \frac{\beta}{6N}\sum_{j=N+1}^{4N}\mathfrak{B}_j(t;k), \quad \mathfrak{B}_j(t;k) = \binom{j}{k} t^{j-k}$$

For $k = N$, the upper bound is set to $\overline{\epsilon}(k) = 1$ and the lower bound is as

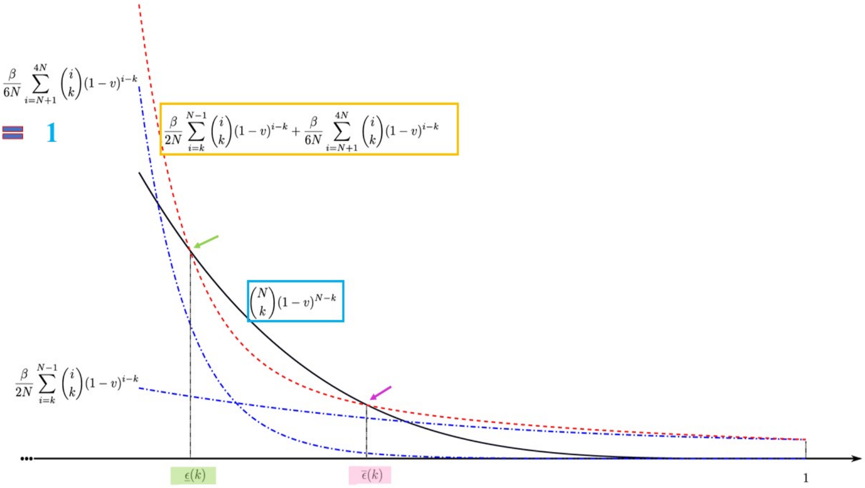
$$1 = \frac{\beta}{6N}\sum_{j=N+1}^{4N}\mathfrak{B}_j(t;N).$$

Graphical Illustration for Lemma 1



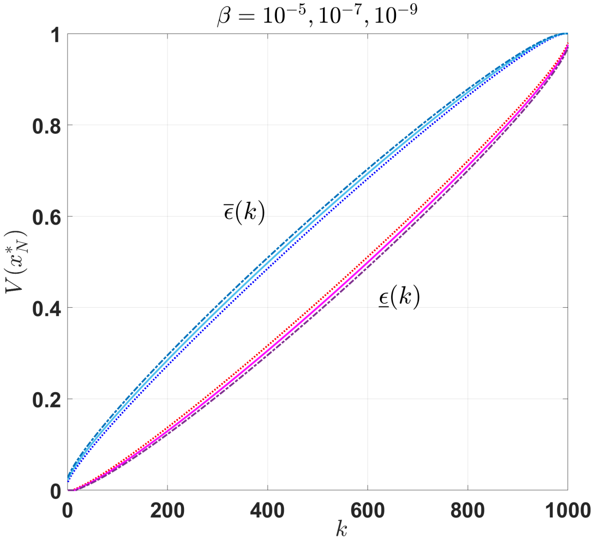
Let $t = 1 - v$ in $\mathfrak{B}_j(t;k)$ of Lemma 1

Graphical Illustration for Lemma 1



Let $t = 1 - v$ in $\mathfrak{B}_j(t; k)$ of Lemma 1

Graphical Illustration for Lemma 1



Curves $\underline{\epsilon}(k)$ and $\overline{\epsilon}(k)$ for violation $V(x_N^*)$ with $N = 1000$, and confidence β

Application: Robust Control

Consider an uncertain discrete LTI system

$$z_{l+1} = A(q)z_l + B(q)u_l, \quad l = 0, 1, \cdots, L-1, \tag{1}$$

- $z_l \in \mathbb{R}^n$ is the state with initial z_0, $u_l \in \mathbb{R}^m$ is the control input
- $A(q) \in \mathbb{R}^{n\times n}$, $B(q) \in \mathbb{R}^{n\times m}$ depend on the uncertainty $q \in \mathcal{Q} \subseteq \mathbb{R}^{n_q}$

Control Objective:

♠ Seek a control sequence $\{u_l\}_{l=0}^{L-1}$ with input sparsity that drives the system z_l from the initial state z_0 near to the terminal state $z_L(u, q)$.

$$z_L(u,q) = A(q)^L z_0 + \underbrace{\begin{bmatrix} A(q)^{L-1}B(q) & \dots & A(q)B(q) & B(q) \end{bmatrix}}_{\mathfrak{R}(q)\in\mathbb{R}^{n\times mL}} \underbrace{\begin{bmatrix} u_0 \\ \vdots \\ u_{L-1} \end{bmatrix}}_{u\in\mathbb{R}^{mL}}$$

Application: Robust Control

Consider an uncertain discrete LTI system

$$z_{l+1} = A(q)z_l + B(q)u_l, \quad l = 0, 1, \cdots, L-1, \tag{1}$$

- $z_l \in \mathbb{R}^n$ is the state with initial z_0, $u_l \in \mathbb{R}^m$ is the control input
- $A(q) \in \mathbb{R}^{n\times n}$, $B(q) \in \mathbb{R}^{n\times m}$ depend on the uncertainty $q \in \mathcal{Q} \subseteq \mathbb{R}^{n_q}$

Control Objective:

♠ Seek a control sequence $\{u_l\}_{l=0}^{L-1}$ with input sparsity that drives the system z_l from the initial state z_0 near to the terminal state $z_L(u, q)$.

$h(u, q)$: measure the target $\bar{z}$ and final state $z_L(u, q)$ with radius $\gamma > 0$

$$\begin{aligned} h(u, q) &\triangleq \|z_L(u, q) - \bar{z}\|_2 - \gamma \\ &= \|A(q)^L z_0 + \mathfrak{R}(q)u - \bar{z}\|_2 - \gamma \leq 0 \end{aligned}$$

Risk-Aware Sparse Optimal Control Problem

Given a reachable system (1), and the parameters $z_0, \bar{z}, L, \rho, \gamma$, achieving a trade-off between the sparse control and the risk amounts to minimizing

$$
(\mathrm{RaSOCP}_N^\rho): \quad \begin{array}{cl} \displaystyle\min_{u \in \mathbb{R}^{mL}, \xi^i \geq 0} & \|u\|_1 + \rho \sum_{i=1}^{N} \xi^i \\ \text{s.t.} & h(u, q^i) \leq \xi^i, \quad i = 1, \cdots, N. \end{array}
$$

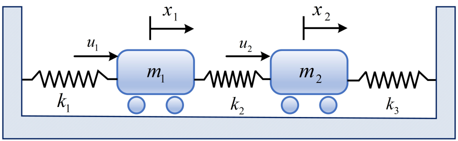
Theorem 1 (Violation of RaSOCP_N^ρ)

With $\underline{\epsilon}(\cdot)$ and $\overline{\epsilon}(\cdot)$ in Lemma 1, given a confidence $\beta \in (0,1)$, the risk assessment for violation $V(u_N^*)$ in problem (RaSOCP_N^ρ) is as follows

$$
\mathbb{P}^N\{\underline{\epsilon}(s_N^*) \leq V(u_N^*) \leq \overline{\epsilon}(s_N^*)\} \geq 1 - \beta,
$$

where s_N^* counts the number of the violated constraints $h(u_N^*, q^i) \geq 0$.

Numerical Example



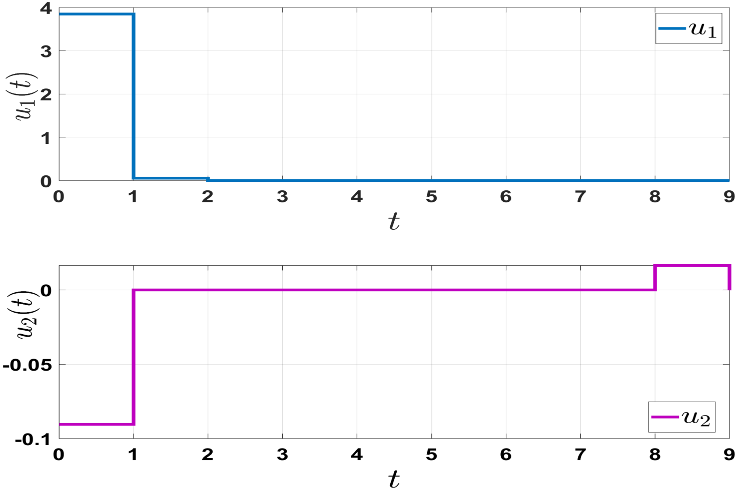
Consider a continuous fourth-order mass-spring system with two inputs

$$\dot{z}(t) = \begin{bmatrix} 0 & 1 & 0 & 0 \\ \frac{-(k_1+k_2)}{m_1} & 0 & \frac{k_2}{m_1} & 0 \\ 0 & 0 & 0 & 1 \\ \frac{k_2}{m_1} & 0 & \frac{-(k_2+k_3)}{m_2} & 0 \end{bmatrix} z(t) + \begin{bmatrix} 0 & 0 \\ \frac{1}{m_1} & 0 \\ 0 & 0 \\ 0 & \frac{1}{m_2} \end{bmatrix} u(t), \;\; z(0) = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix}$$

and discretize it with sampling time 0.05 s to be the discrete plant.

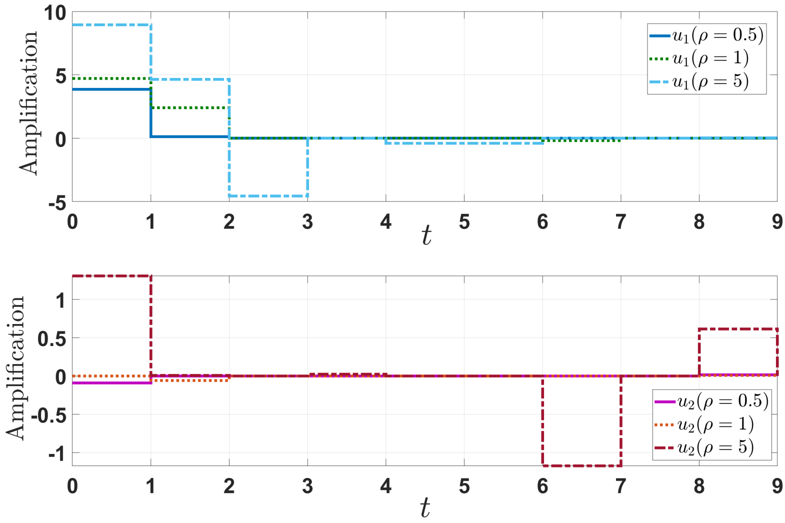
- $m_1 = 1,\ m_2 = 2,\ L = 10,\ q = [k_1\ \ k_2\ \ k_3] \overset{i.i.d.}{\sim} \mathcal{U}[0.1, 1]^3, \gamma = 0.5,\ \bar{z} = 0$
- Solve (RaSOCP_N^ρ) by using CVX and $N = 1000$.

Numerical Results



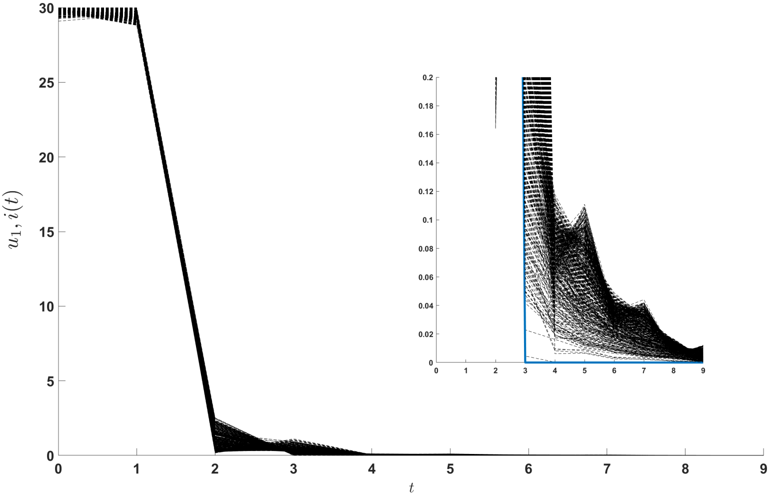
Optimal value $\|u_N^*\|_1 + \frac{1}{5}\sum_i \xi_*^i = 247.083$

Numerical Results



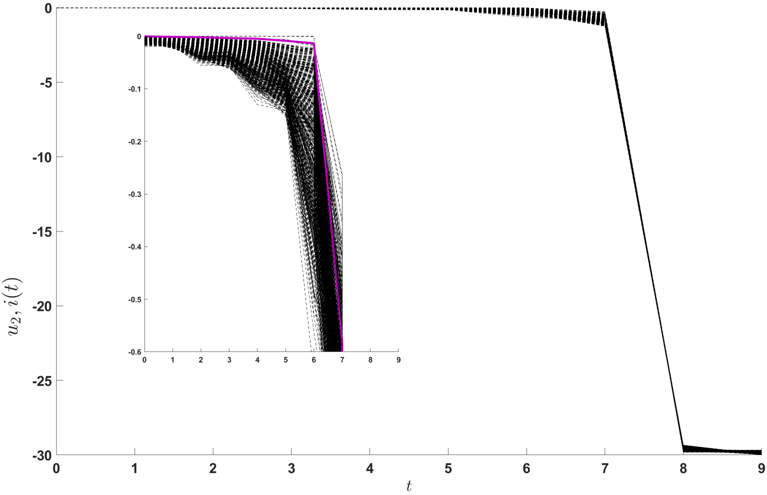
A lower value ρ improves the sparse cost (Non Free Lunch!)

Numerical Results



Additional samples testing $N_\mathrm{V} = 2000$ for $u_{1,i}(t)$, $i = 1, \ldots, 2000$

Numerical Results



Additional samples testing $N_{\rm V} = 2000$ for $u_{2,i}(t)$, $i = 1, \ldots, 2000$

Conclusions

The Take-Home Message

- Make a trade-off between the sparse cost performance and risk.
- ¶ Provide an interval risk assessment guarantees for sparse control.
- ★ Perform a posteriori testing for sparse optimal control.

Thank you for your attention !