Robust Optimization

- definitions of robust optimization
- robust linear programs
- robust cone programs
- chance constraints

Robust optimization

convex objective $f_0 : \mathbf{R}^n \to \mathbf{R}$, *uncertainty set* $\mathcal{U}$, and $f_i : \mathbf{R}^n \times \mathcal{U} \to \mathbf{R}$,

$$x \mapsto f_i(x,u) \text{ convex for all } u \in \mathcal{U}$$

general form

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x,u) \leq 0 \text{ for all } u \in \mathcal{U}, i = 1, \ldots, m. \end{array}$$

equivalent to

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & \sup_{u \in \mathcal{U}} f_i(x,u) \leq 0, i = 1, \ldots, m. \end{array}$$

- Bertsimas, Ben-Tal, El-Ghaoui, Nemirovski (1990s–now)

Setting up robust problem

- can always replace objective f_0 with $\sup_{u \in \mathcal{U}} f_0(x,u)$, rewrite in epigraph form to

$$\begin{array}{ll} \text{minimize} & t \\ \text{subject to} & \sup\limits_u f_0(x,u) \le t, \sup\limits_u f_i(x,u) \le 0, \ \ i = 1, \ldots, m \end{array}$$

- equality constraints make no sense: a robust equality $a^T(x+u) = b$ for all $u \in \mathcal{U}$?

three questions:

- is robust formulation useful?
- is robust formulation computable?
- how should we choose $\mathcal{U}$?

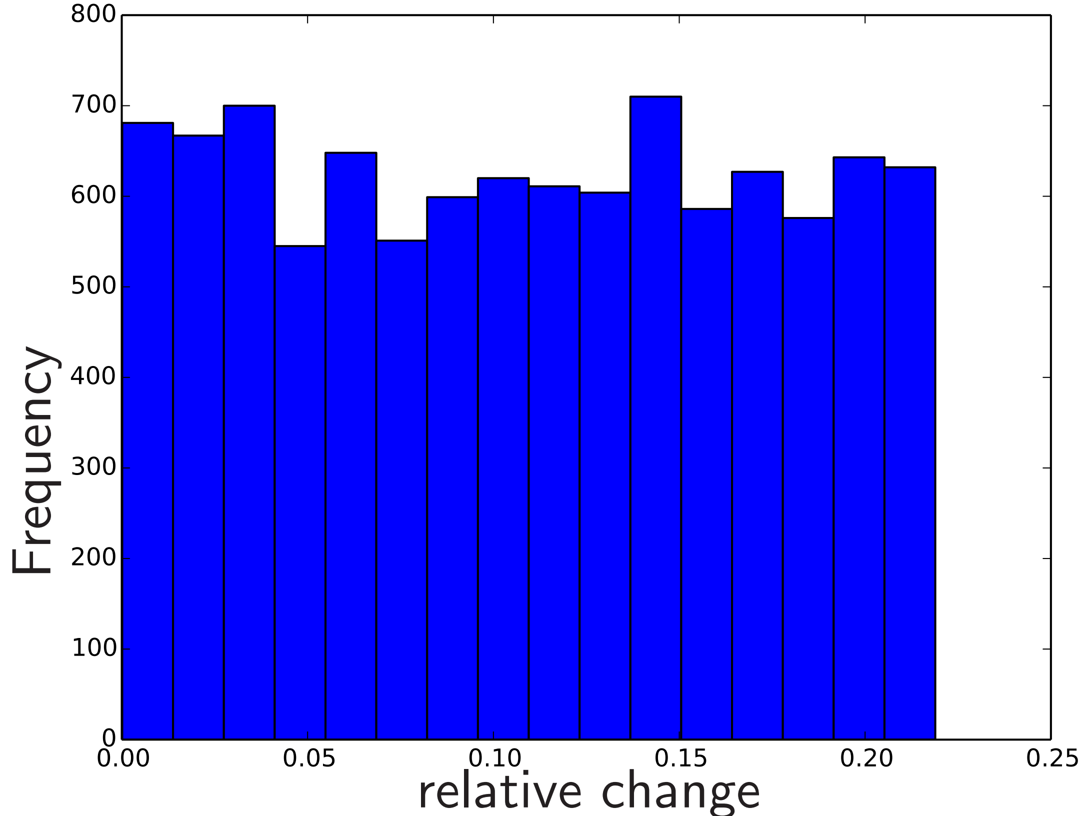
Example failure for linear programming

$$c = \begin{bmatrix} 100 \\ 199.9 \\ -5500 \\ -6100 \end{bmatrix} \quad A = \begin{bmatrix} -.01 & -.02 & .5 & .6 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 90 & 100 \\ 0 & 0 & 40 & 50 \\ 100 & 199.9 & 700 & 800 \\ & -I_4 & & \end{bmatrix} \quad \text{and} \quad b = \begin{bmatrix} 0 \\ 1000 \\ 2000 \\ 800 \\ 100000 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

c vector of costs/profits for two drugs, constraints $Ax \preceq b$ on production

- what happens if we vary percentages $.01$, $.02$ (chemical composition of raw materials) by .5% and 2%, i.e. $.01 \pm .00005$ and $.02 \pm .0004$?

Example failure for linear programming



Frequently lose 15–20% of profits

Alternative robust LP

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & (A+\Delta)x \preceq b, \quad \text{all } \Delta \in \mathcal{U} \end{array}$$

where $|\Delta_{11}| \leq .00005$, $|\Delta_{12}| \leq .0004$, $\Delta_{ij} = 0$ otherwise

- solution x_{robust} has degradation *provably* no worse than 6%

How to choose uncertainty sets

- uncertainty set $\mathcal{U}$ a modeling choice
- common idea: let U be random variable, want constraints that

$$\mathbf{Prob}(f_i(x, U) \geq 0) \leq \epsilon \tag{1}$$

- typically hard (non-convex except in special cases)
- find set $\mathcal{U}$ such that $\mathbf{Prob}(U \in \mathcal{U}) \geq 1 - \epsilon$, then sufficient condition for (1)

$$f_i(x, u) \leq 0 \text{ for all } u \in \mathcal{U}$$

Uncertainty set with Gaussian data

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & \mathbf{Prob}(a_i^T x > b_i) \leq \epsilon, \;\; i = 1, \ldots, m \end{array}$$

coefficient vectors a_i i.i.d. $\mathcal{N}(\overline{a}, \Sigma)$ and failure probability ϵ

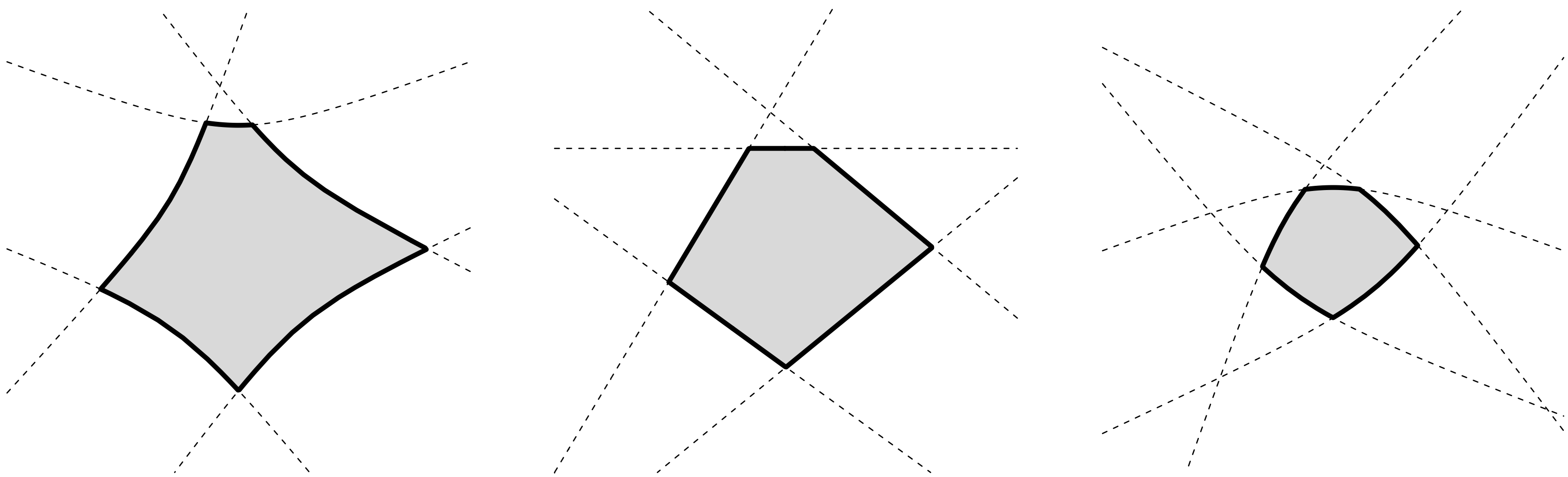
- marginally $a_i^T x \sim \mathcal{N}(\overline{a}_i^T x, x^T \Sigma x)$
- for $\epsilon = .5$, just LP

$$\text{minimize} \;\; c^T x \;\; \text{subject to} \; a_i^T x \leq b_i, \;\; i = 1, \ldots, m$$

- what about $\epsilon = .1, .9$?

Gaussian uncertainty sets

$$\{x \mid \mathbf{Prob}(a_i^T x > b_i) \le \epsilon\} = \{x \mid \overline{a}_i^T x - b_i - \Phi^{-1}(\epsilon)\sqrt{x^T \Sigma x} \le 0\}$$



$\epsilon = .9$ $\epsilon = .5$ $\epsilon = .1$

Problem is convex, so no problem?

not quite...
consider quadratic constraint

$$\|Ax + Bu\|_2 \le 1 \quad \text{for all} \quad \|u\|_\infty \le 1$$

- convex quadratic *maximization* in u
- solutions on extreme points $u \in \{-1, 1\}^n$
- and NP-hard to maximize (even approximately [Håstad]) convex quadratics over hypercube

Robust LPs

Important question: when is a robust LP still an LP (robust SOCP an SOCP, robust SDP an SDP)

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & (A+U)x \preceq b \ \text{ for } U \in \mathcal{U}. \end{array}$$

can always represent formulation constraint-wise, consider only one inequality

$$(a+u)^T x \leq b \ \text{ for all } u \in \mathcal{U}.$$

- Simple example: $\mathcal{U} = \{u \in \mathbf{R}^n \mid \|u\|_\infty \leq \delta\}$, then

$$a^T x + \delta \|x\|_1 \leq b$$

Polyhedral uncertainty

for matrix $F \in \mathbf{R}^{m \times n}$, $g \in \mathbf{R}^m$,

$$(a+u)^T x \leq b \;\; \text{for } u \in \mathcal{U} = \{u \in \mathbf{R}^n \mid Fu + g \succeq 0\}.$$

duality essential for transforming (semi-)infinite inequality into tractable problem

- Lagrangian for maximizing $u^T x$:

$$L(u, \lambda) = x^T u + \lambda^T (Fu + g), \quad \sup_u L(u, \lambda) = \begin{cases} +\infty & \text{if } F^T \lambda + x \neq 0 \\ \lambda^T g & \text{if } F^T \lambda + x = 0. \end{cases}$$

- gives equivalent inequality constraints

$$a^T x + \lambda^T g \leq b, \quad F^T \lambda + x = 0, \quad \lambda \succeq 0.$$

Portfolio optimization (with robust LPs)

- n assets $i = 1, \ldots, n$, random multiplicative return R_i with $\mathbf{E}[R_i] = \mu_i \geq 1$, $\mu_1 \geq \mu_2 \geq \cdots \geq \mu_n$
- "certain" problem has solution $x_{\rm nom} = e_1$,

$$\text{maximize} \;\; \mu^T x \;\; \text{subject to} x^T \mathbf{1} = 1 \; x \succeq 0$$

- if asset i varies in range $\mu_i \pm u_i$, robust problem

$$\text{maximize} \;\; \sum_{i=1}^{n} \inf_{u \in [-u_1, u_i]} (\mu_i + u) x_i \;\; \text{subject to } \mathbf{1}^T x = 1, x \succeq 0$$

and equivalent

$$\text{maximize} \;\; \mu^T x - u^T x \;\; \text{subject to } \mathbf{1}^T x = 1, x \succeq 0$$

Robust LPs as SOCPs

norm-based uncertainty on data vectors a,

$$(a + Pu)^T x \leq b \quad \text{for } u \in \mathcal{U} = \{u \in \mathbf{R}^m \mid \|u\| \leq 1\},$$

gives dual-norm constraint

$$a^T x + \left\|P^T x\right\|_* \leq b$$

Portfolio optimization (tigher control)

- Returns $R_i \in [\mu_i - \mathsf{u}_i, \mu_i + \mathsf{u}_i]$ with $\mathbf{E}\, R_i = \mu_i$
- guarantee return with probability $1 - \epsilon$

$$\underset{\mu, t}{\text{maximize}} \;\; t \;\; \text{subject to} \;\; \mathbf{Prob}\left(\sum_{i=1}^{n} R_i x_i \geq t \right) \geq 1 - \epsilon$$

- *value at risk* is non-convex in x, approximate it?
- approximate with high-probability bounds
- less conservative than LP (certain returns) approach

Portfolio optimization: probability approximation

- Hoeffding's inequality

$$\mathbf{Prob}\left(\sum_{i=1}^{n}(R_i - \mu_i)x_i \le -t\right) \le \exp\left(-\frac{t^2}{2\sum_{i=1}^{n} x_i^2 \mathsf{u}_i^2}\right).$$

- written differently

$$\mathbf{Prob}\left[\sum_{i=1}^{n} R_i x_i \le \mu^T x - t\left(\sum_{i=1}^{n} \mathsf{u}_i^2 x_i^2\right)^{\frac{1}{2}}\right] \le \exp\left(-\frac{t^2}{2}\right)$$

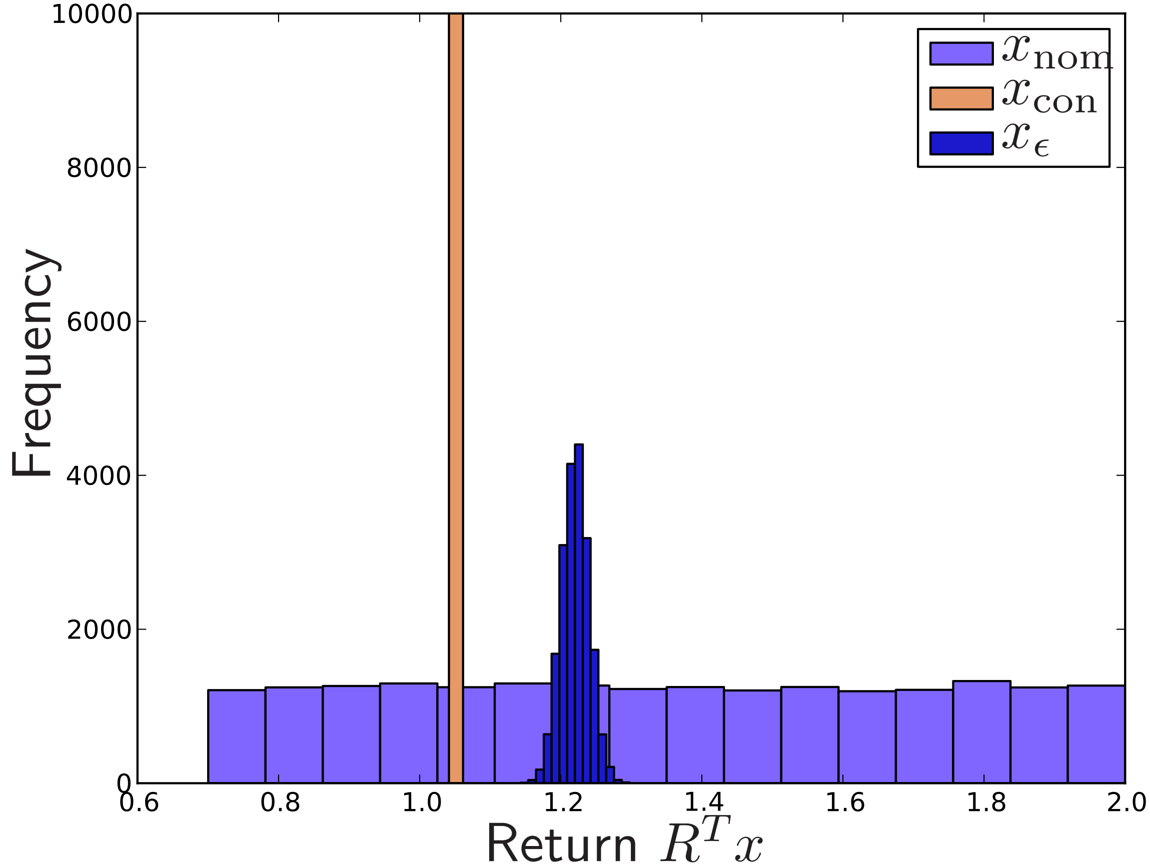
- set $t = \sqrt{2\log(1/\epsilon)}$, gives robust problem

$$\text{maximize } \mu^T x - \sqrt{2\log\frac{1}{\epsilon}}\,\|\mathbf{diag}(u)x\|_2 \quad \text{subject to } \mathbf{1}^T x = 1,\ x \succeq 0.$$

Portfolio optimization comparison

- data $\mu_i = 1.05 + \frac{3(n-i)}{10n}$, uncertainty $|u_i| \leq \mathsf{u}_i = .05 + \frac{n-i}{2n}$ and $\mathsf{u}_n = 0$
- nominal minimizer $x_{\rm nom} = e_1$
- conservative (LP) minimizer $x_{\rm con} = e_n$ (guaranteed 5% return),
- robust (SOCP) minimizer x_ϵ for value-at risk $\epsilon = 2 \times 10^{-4}$

Portfolio optimization comparison



Returns chosen randomly in $\mu_i \pm u_i$, 10,000 experiments

LPs with conic uncertainty

- convex cone K, dual cone $K^* = \{v \in \mathbf{R}^m \mid v^T x \geq 0, \text{ all } x \in K\}$
- recall $x \succeq_K y$ iff $x - y \in K$
- robust inequality

$$(a+u)^T x \leq b \text{ for all } u \in \mathcal{U} = \{u \in \mathbf{R}^n \mid Fu + g \succeq_K 0\}$$

- under constraint qualification, equivalent to

$$a^T x + \lambda^T g \leq b, \quad \lambda \succeq_{K^*} 0, \quad x + F^T \lambda = 0$$

Example calculation: LP with semidefinite uncertainty

- symmetric matrices $A_0, A_1, \ldots, A_m \in \mathbf{S}^k$, robust counterpart to $a^T x \leq b$

$$(a + Pu)^T x \leq b \quad \text{for all } u \text{ s.t. } A_0 + \sum_{i=1}^{m} u_i A_i \succeq 0$$

- cones $K = \mathbf{S}_+^k$, $K^* = \mathbf{S}_+^k$
- Slater condition: $\bar{u}$ such that $A_0 + \sum_i A_i \bar{u}_i \succ 0$
- duality gives equivalent representation

$$a^T x + \mathbf{Tr}(\Lambda A_0) \leq b, \quad P^T x + \begin{bmatrix} \mathbf{Tr}(\Lambda A_1) \\ \vdots \\ \mathbf{Tr}(\Lambda A_m) \end{bmatrix} = 0, \quad \Lambda \succeq 0.$$

Robust second-order cone problems

- Lorentz/SOCP cone, nominal inequality

$$\|Ax + b\|_2 \leq c^T x + d$$

- $A = [a_1 \ \cdots \ a_n]^T \in \mathbf{R}^{m \times n}$, allow A, c to vary

- interval uncertainty

- ellipsoidal uncertainty

- matrix uncertainty

SOCPs with interval uncertainty

entries A_{ij} perturbed by Δ_{ij} with $|\Delta_{ij}| \leq \delta$, c by cone:

$$\|(A+\Delta)x+b\|_2 \leq (c+u)^T x + d \quad \text{all} \quad \|\Delta\|_\infty \leq \delta, \; u \in \mathcal{U}$$

- split into two inequalities (first is robust LP)

$$\|(A+\Delta)x+b\|_2 \leq t, \quad t \leq (c+u)^T x + d$$

second

$$\begin{aligned}
\sup_{\Delta:|\Delta_{ij}|\leq\delta} \|(A+\Delta)x+b\|_2 &= \sup_{\Delta:|\Delta_{ij}|\leq\delta} \left(\sum_{i=1}^m [(a_i+\Delta_i)^T x + b_i]^2 \right)^{1/2} \\
&= \sup_{\Delta\in\mathbf{R}^{m\times n}} \left\{ \|z\|_2 \mid z_i = a_i^T x + \Delta_i^T x + b_i, \|\Delta_i\|_\infty \leq \delta \right\} \\
&= \inf \left\{ \|z\|_2 \mid z_i \geq |a_i^T x + b| + \delta \|x\|_1 \right\}.
\end{aligned}$$

SOCPs with ellipse-like uncertainty

- matrices $P_1, \ldots, P_m \in \mathbf{R}^{n \times n}$, $u \in \mathbf{R}^m$ with $\|u\| \leq 1$
- robust/uncertain inequality

$$\left(\sum_{i=1}^{m} [(a_i + P_i u)^T x + b_i]^2 \right)^{1/2} \leq t \quad \text{for all } u \text{ s.t. } \|u\|_2 \leq 1.$$

- rewrite $z_i \geq \sup_{\|u\| \leq 1} |a_i^T x + b_i + u^T P_i^T x|$, equivalent

$$\|z\|_2 \leq t, \quad z_i \geq |a_i^T x + b_i| + \left\| P_i^T x \right\|_*, \quad i = 1, \ldots, m.$$

SOCPs wtih matrix uncertainty

- Matrix $P \in \mathbf{R}^{m \times n}$ and radius δ, uncertain inequality

$$\|(A + P\Delta)x + b\|_2 \le t, \quad \text{for } \Delta \in \mathbf{R}^{n \times n} \text{ s.t. } \|\Delta\| \le \delta,$$

- tool one: Schur complements gives equivalence of

$$\|x\|_2 \le t \quad \text{and} \quad \begin{bmatrix} t & x^T \\ x & tI_n \end{bmatrix} \succeq 0.$$

- tool two: homogeneous S-lemma

$$x^T A x \ge 0 \text{ implies } x^T B x \ge 0 \quad \text{if and only if} \quad \exists\, \lambda \ge 0 \text{ s.t. } B \succeq \lambda A.$$

SOCPs with matrix uncertainty

$$\|(A+P\Delta)x+b\|_2 \leq t, \quad \text{for } \Delta \in \mathbf{R}^{n\times n} \text{ s.t. } \|\Delta\| \leq \delta,$$

equivalent to

$$\begin{bmatrix} t & ((A+P\Delta)x+b)^T \\ (A+P\Delta)x+b & tI_m \end{bmatrix} \succeq 0 \quad \text{for } \|\Delta\| \leq 1.$$

or

$$ts^2 + 2s((A+P\Delta)x+b)^T v + t\,\|v\|_2^2 \geq 0 \quad \text{for all } s \in \mathbf{R}, v \in \mathbf{R}^m, \quad \|\Delta\| \leq 1.$$

SOCPs with matrix uncertainty: final result

$$\|(A + P\Delta)x + b\|_2 \leq t, \quad \text{for } \Delta \in \mathbf{R}^{n\times n} \text{ s.t. } \|\Delta\| \leq \delta,$$

equivalent to

$$\begin{bmatrix} t & (Ax+b)^T & x^T \\ Ax+b & t - \lambda PP^T & 0 \\ x & 0 & \lambda I_n \end{bmatrix} \succeq 0.$$

Example: robust regression

$$\text{minimize} \; \|Ax - b\|_2$$

where A corrupted by Gaussian noise,

$$A = A_\star + \Delta \quad \text{for} \quad \Delta_{ij} \sim \mathcal{N}(0, 1)$$

decide to be robust to Δ by

- bounding individual entries Δ_{ij}
- bounding norms of rows Δ_i
- bounding (ℓ_2-operator) norm of Δ

Choice of uncertainty in robust regression

Theorem [e.g. Vershynin 2012] Let $\Delta \in \mathbf{R}^{m \times n}$ have i.i.d. $\mathcal{N}(0,1)$ entries. For all $t \geq 0$, the following hold:

- For each pair i, j

$$\mathbf{Prob}(|\Delta_{ij}| \geq t) \leq 2 \exp\left(-\frac{t^2}{2}\right).$$

- For each i

$$\mathbf{Prob}(\|\Delta_i\|_2 \geq \sqrt{n} + t) \leq \exp\left(-\frac{t^2}{2}\right).$$

- For the entire matrix Δ,

$$\mathbf{Prob}(\|\Delta\| \geq \sqrt{m} + \sqrt{n} + t) \leq \exp\left(-\frac{t^2}{2}\right).$$

Choice of uncertainty in robust regression

idea: choose bounds $t(\delta)$ to guarantee $\mathbf{Prob}(\text{deviation} \geq t(\delta)) \leq \delta$

- coordinate-wise: $t_\infty(\delta)^2 = 2\log\frac{2mn}{\delta}$,

$$\mathbf{Prob}(\max_{i,j} |\Delta_{ij}| \geq t_\infty(\delta)) \leq 2mn \exp\left(-\frac{t_\infty(\delta)^2}{2}\right) = \delta$$

- row-wise: $t_2(\delta)^2 = 2\log\frac{m}{\delta}$,

$$\mathbf{Prob}(\max_{i} \|\Delta_i\|_2 \geq t_2(\delta)) \leq m \exp\left(-\frac{t_2(\delta)^2}{2}\right) = \delta$$

- matrix-norm: $t_{\mathrm{op}}(\delta)^2 = 2\log\frac{1}{\delta}$,

$$\mathbf{Prob}(\|\Delta\| \geq \sqrt{n} + \sqrt{m} + t_{\mathrm{op}}(\delta)) \leq \exp\left(-\frac{t_{\mathrm{op}}(\delta)^2}{2}\right) = \delta.$$

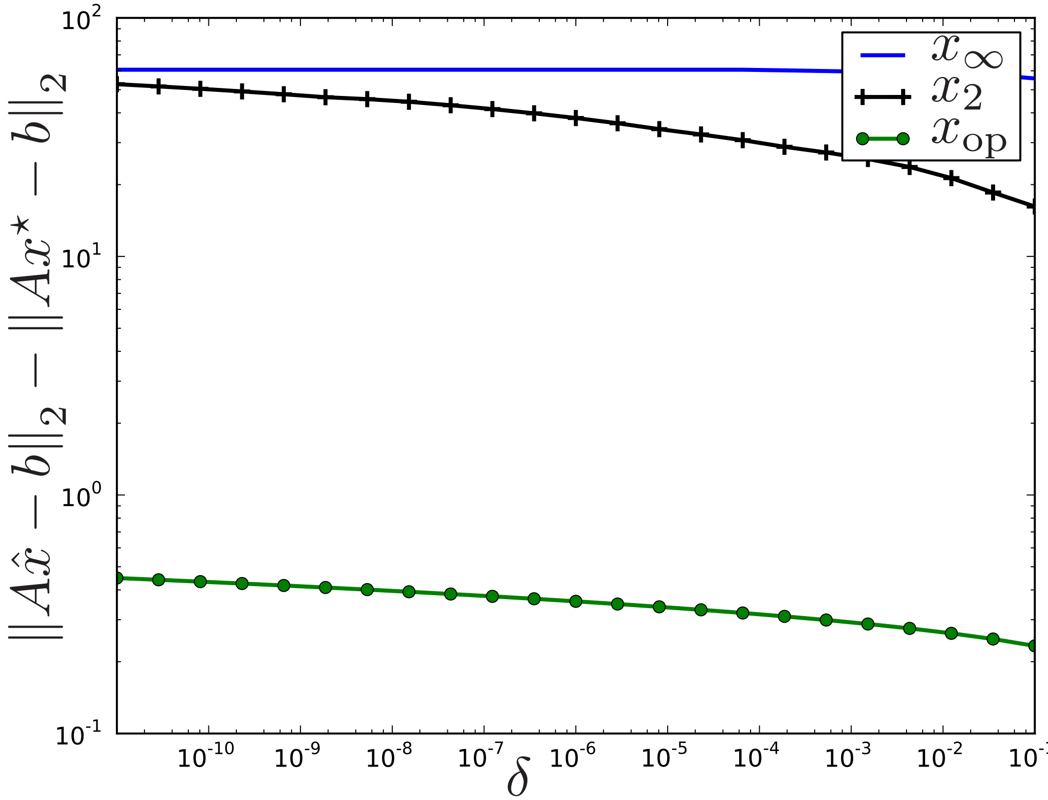
Robust regression results

$$\underset{x}{\text{minimize}} \quad \sup_{\Delta \in \mathcal{U}} \|(A+\Delta)x - b\|_2$$

where $\mathcal{U}$ is one of the three uncertainty sets

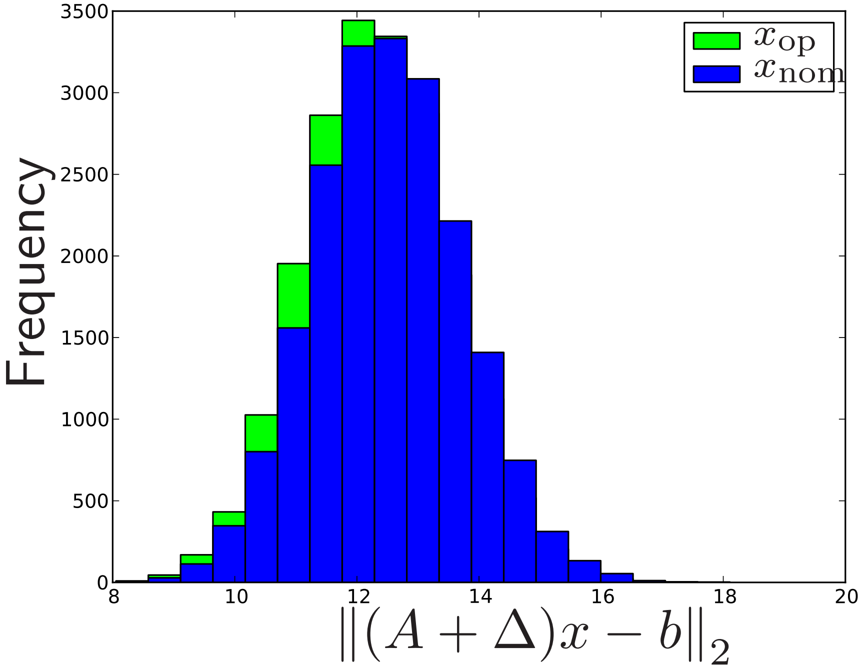
$$\begin{aligned}
\mathcal{U}_\infty &= \{\Delta \mid \|\Delta\|_\infty \leq t_\infty(\delta)\}, \\
\mathcal{U}_2 &= \{\Delta \mid \|\Delta_i\|_2 \leq \sqrt{n} + t_2(\delta) \text{ for } i = 1, \ldots m\}, \\
\mathcal{U}_{\text{op}} &= \{\Delta \mid \|\Delta\| \leq \sqrt{n} + \sqrt{m} + t_{\text{op}}(\delta)\}.
\end{aligned}$$

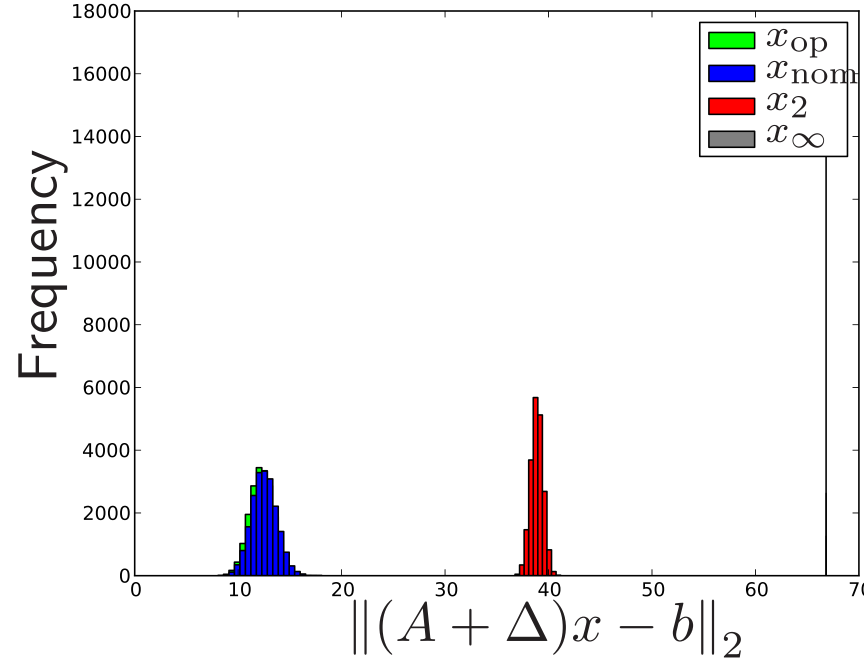
Robust regression results



Objective value $\|A\hat{x} - b\|_2 - \|Ax^\star - b\|_2$ versus δ, where $x^\star$ minimizes nominal objective and $\hat{x}$ denotes robust solution

Robust regression results





- residuals for the robust least squares problem $\|(A+\Delta)x-b\|_2$
- uncertainty sets $\mathcal{U}_{\mathrm{nom}} = \{0\}$ vs. $\mathcal{U}_\infty$, $\mathcal{U}_2$, $\mathcal{U}_{\mathrm{op}}$
- experiment with $N = 10^5$ random Gaussian matrices