STAT 593
Robust optimization overview

Joseph Salmon
http://josephsalmon.eu

Télécom Paristech, Institut Mines-Télécom
&
University of Washington, Department of Statistics
(Visiting Assistant Professor)

Outline

Robust optimization point of view

Linear regression case: connexion with regularization

Table of Contents

General motivation

Optimization with uncertainty: in a statistical / corrupted scenario, computing standard estimators requires to take into account dataset corruptions inherent to statistical modeling/measures.

References on this fields include a long survey[(1)] and a more exhaustive book[(2)]

[(1)] D. Bertsimas, D. B. Brown, and C. Caramanis. "Theory and applications of robust optimization". In: *SIAM Rev.* 53.3 (2011), pp. 464–501.

[(2)] A. Ben-Tal, L. El Ghaoui, and A. Nemirovski. *Robust optimization*. Princeton Series in Applied Mathematics. Princeton University Press, Princeton, NJ, 2009, pp. xxii+542.

Example

Linear Programming (LP):

$$\begin{aligned} \min_x \quad & \langle\, x \,,\, c \,\rangle \\ \text{s.t.} \quad & Ax \le b \end{aligned}$$

In real scenarios such object are observed with noise: one would like to optimize such a problem with some robustness on the measures

Example

Linear Programming (LP):

$$\begin{aligned} \min_{x} \quad & \langle x , c \rangle \\ \text{s.t.} \quad & Ax \leq b \end{aligned}$$

In real scenarios such object are observed with noise: one would like to optimize such a problem with some robustness on the measures

Robust reformulation: for a set of **uncertainty** $\mathcal{U}$

$$\begin{aligned} \min_{x} \quad & \max_{A,b,c} \langle x , c \rangle \\ \text{s.t.} \quad & Ax \leq b, \quad (A, b, c) \in \mathcal{U} \\ \iff & \\ \min_{x,t} \quad & t \\ \text{s.t.} \quad & Ax \leq b \text{ and } \langle x , c \rangle \leq t, \quad \forall (A, b, c) \in \mathcal{U} \end{aligned}$$

Rem: this is a min-max point of view

Possible variants: stochastic optim. / chance constraint

Assume randomness on (A, b, c)

Chance constraint formulation:

$$\begin{aligned} \min_{x,t} \quad & t \\ \text{s.t.} \quad & \mathbb{P}_{(A,b,c)}\{\langle x , c \rangle \le t \text{ and } Ax \le b\} \ge 1 - \epsilon, \end{aligned}$$

- ϵ is a tolerance parameter
- $\mathbb{P}_{(A,b,c)}$ is a fixed associated probability

Difficulty: modeling the perturbation on the data might be difficult; also the associated problem could be hard to solve

Possible variants: stochastic optimization / ambiguous chance constraint

Assume randomness on (A, b, c)

Chance constraint formulation:

$$\begin{aligned} \min_{x,t} \quad & t \\ \text{s.t.} \quad & \mathbb{P}_{(A,b,c)}\{\langle\, x\,,\, c\,\rangle \leq t \text{ and } Ax \leq b\} \geq 1 - \epsilon, \forall P \in \mathcal{P} \end{aligned}$$

- ϵ is a tolerance parameter
- $\mathcal{P}$ is a family of (potentially parametric) probabilities

Difficulty: modeling the perturbation on the data might be difficult; also the associated problem could be hard to solve

Robust optimization framework

Solve the following optimization problem:

$$\begin{array}{ll} \min & f_0(x) \\ \text{s.t.} & f_j(x, u_j) \leq 0, \quad (u_1, \ldots, u_m)^\top \in \mathcal{U} \end{array}$$

- $x \in \mathbb{R}^n$: vector of decision variables,
- $f_0, f_j : \mathbb{R}^n \to \mathbb{R}$ are functions
- $\forall j, u_j \in \mathbb{R}^k$:**uncertainty parameters** assumed to take values in the (closed) **uncertainty set** $\mathcal{U} \subseteq (\mathbb{R}^k)^m$

Goal : compute solutions x among all solutions which are feasible for all realizations of the disturbances $(u_1, \ldots, u_m)^\top \in \mathcal{U}$.

Rem: $\mathcal{U}$ a continuous set $\implies$ infinite number of constraints

Intuitively: offers some protection for optimization problems containing parameters which are not known exactly.

Robust linear optimization

linear programming case:

$$\begin{aligned} \min \quad & \langle x \, , \, c \rangle \\ \text{s.t.} \quad & Ax \leq b \quad \forall a_j \in \mathcal{U}_j, j = 1, \ldots, m \end{aligned}$$

where a_j represents the j-th row of A

Rem: a coupling on the constraint of a_i might exist

Reformulation of the constraint as a **sub-problem**:

$$\max_{a_j \in \mathcal{U}_j} \langle a_j \, , \, x \rangle \leq b_j, \quad \forall j \iff Ax \leq b \quad \forall a_j \in \mathcal{U}_j, j = 1, \ldots, m$$

Details for interval uncertainty

Consider

$$\{\langle\, a\, ,\, x\,\rangle \leq b\}_{[a;b]\in\mathcal{U}} \text{ where } \mathcal{U} = \{[a;b] = [a^0;b^0] + \sum_{\ell=1}^{L} \zeta_\ell [a^\ell;b^\ell], \zeta \in \mathcal{Z}\}$$

$$\text{and } \mathcal{Z} = \{\zeta \in \mathbb{R}^L : \|\zeta\|_\infty \leq 1\}$$

Details for interval uncertainty

Consider

$$\{\langle\, a\,,\, x\,\rangle \le b\}_{[a;b]\in\mathcal{U}} \text{ where } \mathcal{U} = \{[a;b] = [a^0;b^0] + \sum_{\ell=1}^{L} \zeta_\ell [a^\ell; b^\ell], \zeta \in \mathcal{Z}\}$$

$$\text{and } \mathcal{Z} = \{\zeta \in \mathbb{R}^L : \|\zeta\|_\infty \le 1\}$$

The former is equivalent to

$$\left\langle\, a^0\,,\, x\,\right\rangle + \sum_{\ell=1}^{L} \zeta_\ell \left\langle\, a^\ell\,,\, x\,\right\rangle \le b^0 + \sum_{\ell=1}^{L} \zeta_\ell b^\ell, \quad \forall \zeta, \|\zeta\|_\infty \le 1$$

Details for interval uncertainty

Consider

$$\{\langle a, x \rangle \le b\}_{[a;b]\in\mathcal{U}} \text{ where } \mathcal{U} = \{[a;b] = [a^0;b^0] + \sum_{\ell=1}^{L} \zeta_\ell [a^\ell; b^\ell], \zeta \in \mathcal{Z}\}$$

$$\text{and } \mathcal{Z} = \{\zeta \in \mathbb{R}^L : \|\zeta\|_\infty \le 1\}$$

The former is equivalent to

$$\left\langle a^0, x \right\rangle + \sum_{\ell=1}^{L} \zeta_\ell \left\langle a^\ell, x \right\rangle \le b^0 + \sum_{\ell=1}^{L} \zeta_\ell b^\ell, \quad \forall \zeta, \|\zeta\|_\infty \le 1$$

$$\iff \sum_{\ell=1}^{L} \zeta_\ell \left(\left\langle a^\ell, x \right\rangle - b^\ell\right) \le b^0 - \left\langle a^0, x \right\rangle, \quad \forall \zeta, \|\zeta\|_\infty \le 1$$

Details for interval uncertainty

Consider

$$\{\langle a, x \rangle \le b\}_{[a;b]\in\mathcal{U}} \text{ where } \mathcal{U} = \{[a;b] = [a^0;b^0] + \sum_{\ell=1}^{L} \zeta_\ell [a^\ell; b^\ell], \zeta \in \mathcal{Z}\}$$

$$\text{and } \mathcal{Z} = \{\zeta \in \mathbb{R}^L : \|\zeta\|_\infty \le 1\}$$

The former is equivalent to

$$\begin{aligned}
& \left\langle a^0, x \right\rangle + \sum_{\ell=1}^{L} \zeta_\ell \left\langle a^\ell, x \right\rangle \le b^0 + \sum_{\ell=1}^{L} \zeta_\ell b^\ell, \quad \forall \zeta, \|\zeta\|_\infty \le 1 \\
\iff \; & \sum_{\ell=1}^{L} \zeta_\ell \left(\left\langle a^\ell, x \right\rangle - b^\ell \right) \le b^0 - \left\langle a^0, x \right\rangle, \quad \forall \zeta, \|\zeta\|_\infty \le 1 \\
\iff \; & \max_{-1 \le \zeta_\ell \le 1} \sum_{\ell=1}^{L} \zeta_\ell \left(\left\langle a^\ell, x \right\rangle - b^\ell \right) \le b^0 - \left\langle a^0, x \right\rangle,
\end{aligned}$$

Details for interval uncertainty

Consider

$$\{\langle a , x \rangle \leq b\}_{[a;b]\in\mathcal{U}} \text{ where } \mathcal{U} = \{[a;b] = [a^0;b^0] + \sum_{\ell=1}^{L} \zeta_\ell [a^\ell;b^\ell], \zeta \in \mathcal{Z}\}$$

$$\text{and } \mathcal{Z} = \{\zeta \in \mathbb{R}^L : \|\zeta\|_\infty \leq 1\}$$

The former is equivalent to

$$\left\langle a^0 , x \right\rangle + \sum_{\ell=1}^{L} \zeta_\ell \left\langle a^\ell , x \right\rangle \leq b^0 + \sum_{\ell=1}^{L} \zeta_\ell b^\ell, \quad \forall \zeta, \|\zeta\|_\infty \leq 1$$

$$\iff \sum_{\ell=1}^{L} \zeta_\ell \left(\left\langle a^\ell , x \right\rangle - b^\ell\right) \leq b^0 - \left\langle a^0 , x \right\rangle, \quad \forall \zeta, \|\zeta\|_\infty \leq 1$$

$$\iff \max_{-1\leq\zeta_\ell\leq 1} \sum_{\ell=1}^{L} \zeta_\ell \left(\left\langle a^\ell , x \right\rangle - b^\ell\right) \leq b^0 - \left\langle a^0 , x \right\rangle,$$

$$\iff \sum_{\ell=1}^{L} \left|\left\langle a^\ell , x \right\rangle - b^\ell\right| \leq b^0 - \left\langle a^0 , x \right\rangle,$$

Details for ellipsoid uncertainty

Consider

$$\{\langle\, a\,,\, x\,\rangle \le b\}_{[a;b]\in\mathcal{U}} \text{ where } \mathcal{U} = \{[a;b] = [a^0;b^0] + \sum_{\ell=1}^{L} \zeta_\ell [a^\ell;b^\ell], \zeta \in \mathcal{Z}\}$$

$$\text{and } \mathcal{Z} = \{\zeta \in \mathbb{R}^L : \|\zeta\|_2 \le \rho\}$$

Details for ellipsoid uncertainty

Consider

$$\{\langle\, a \,,\, x \,\rangle \leq b\}_{[a;b]\in\mathcal{U}} \text{ where } \mathcal{U} = \{[a;b] = [a^0; b^0] + \sum_{\ell=1}^{L} \zeta_\ell [a^\ell; b^\ell], \zeta \in \mathcal{Z}\}$$

$$\text{and } \mathcal{Z} = \{\zeta \in \mathbb{R}^L : \|\zeta\|_2 \leq \rho\}$$

The former is equivalent to (conic quadratic inequality)

$$\left\langle\, a^0 \,,\, x \,\right\rangle + \sum_{\ell=1}^{L} \zeta_\ell \left\langle\, a^\ell \,,\, x \,\right\rangle \leq b^0 + \sum_{\ell=1}^{L} \zeta_\ell b^\ell, \quad \forall \zeta, \|\zeta\|_2 \leq \rho$$

Details for ellipsoid uncertainty

Consider

$$\{\langle\, a\,,\, x\,\rangle \le b\}_{[a;b]\in\mathcal{U}} \text{ where } \mathcal{U} = \{[a;b] = [a^0;b^0] + \sum_{\ell=1}^{L} \zeta_\ell [a^\ell;b^\ell], \zeta \in \mathcal{Z}\}$$

$$\text{and } \mathcal{Z} = \{\zeta \in \mathbb{R}^L : \|\zeta\|_2 \le \rho\}$$

The former is equivalent to (conic quadratic inequality)

$$\left\langle\, a^0\,,\, x\,\right\rangle + \sum_{\ell=1}^{L} \zeta_\ell \left\langle\, a^\ell\,,\, x\,\right\rangle \le b^0 + \sum_{\ell=1}^{L} \zeta_\ell b^\ell, \quad \forall \zeta, \|\zeta\|_2 \le \rho$$

$$\iff \sum_{\ell=1}^{L} \zeta_\ell \left(\left\langle\, a^\ell\,,\, x\,\right\rangle - b^\ell\right) \le b^0 - \left\langle\, a^0\,,\, x\,\right\rangle, \quad \forall \zeta, \|\zeta\|_2 \le \rho$$

Details for ellipsoid uncertainty

Consider

$$\{\langle a, x \rangle \le b\}_{[a;b]\in\mathcal{U}} \text{ where } \mathcal{U} = \{[a;b] = [a^0;b^0] + \sum_{\ell=1}^{L} \zeta_\ell [a^\ell; b^\ell], \zeta \in \mathcal{Z}\}$$

$$\text{and } \mathcal{Z} = \{\zeta \in \mathbb{R}^L : \|\zeta\|_2 \le \rho\}$$

The former is equivalent to (conic quadratic inequality)

$$\begin{aligned}
& \left\langle a^0, x \right\rangle + \sum_{\ell=1}^{L} \zeta_\ell \left\langle a^\ell, x \right\rangle \le b^0 + \sum_{\ell=1}^{L} \zeta_\ell b^\ell, \quad \forall \zeta, \|\zeta\|_2 \le \rho \\
\iff & \sum_{\ell=1}^{L} \zeta_\ell \left(\left\langle a^\ell, x \right\rangle - b^\ell \right) \le b^0 - \left\langle a^0, x \right\rangle, \quad \forall \zeta, \|\zeta\|_2 \le \rho \\
\iff & \max_{\|\zeta\|_2 \le \rho} \sum_{\ell=1}^{L} \zeta_\ell \left(\left\langle a^\ell, x \right\rangle - b^\ell \right) \le b^0 - \left\langle a^0, x \right\rangle,
\end{aligned}$$

Details for ellipsoid uncertainty

Consider

$$\{\langle a, x \rangle \le b\}_{[a;b]\in\mathcal{U}} \text{ where } \mathcal{U} = \{[a;b] = [a^0;b^0] + \sum_{\ell=1}^{L} \zeta_\ell [a^\ell; b^\ell], \zeta \in \mathcal{Z}\}$$

$$\text{and } \mathcal{Z} = \{\zeta \in \mathbb{R}^L : \|\zeta\|_2 \le \rho\}$$

The former is equivalent to (conic quadratic inequality)

$$\left\langle a^0, x \right\rangle + \sum_{\ell=1}^{L} \zeta_\ell \left\langle a^\ell, x \right\rangle \le b^0 + \sum_{\ell=1}^{L} \zeta_\ell b^\ell, \quad \forall \zeta, \|\zeta\|_2 \le \rho$$

$$\iff \sum_{\ell=1}^{L} \zeta_\ell \left(\left\langle a^\ell, x \right\rangle - b^\ell\right) \le b^0 - \left\langle a^0, x \right\rangle, \quad \forall \zeta, \|\zeta\|_2 \le \rho$$

$$\iff \max_{\|\zeta\|_2 \le \rho} \sum_{\ell=1}^{L} \zeta_\ell \left(\left\langle a^\ell, x \right\rangle - b^\ell\right) \le b^0 - \left\langle a^0, x \right\rangle,$$

$$\iff \rho \sqrt{\sum_{\ell=1}^{L} (\langle a^\ell, x \rangle - b^\ell)^2} \le b^0 - \left\langle a^0, x \right\rangle$$

Type of constraints sets: polyhedron

Here we consider **polyhedron uncertainty**:

Theorem

The constraints $\langle a_j \, , \, x \rangle \leq b_j, \forall j$ and $(a_1, \ldots, a_m)^\top \in \mathcal{U}$ and $(a_1, \ldots, a_m)^\top \in \mathcal{U}$ where
$$\mathcal{U} = \left\{ (a_1, \ldots, a_m)^\top \in (\mathbb{R}^k)^m \; : \; \forall j, D_j a_j \leq d_j \right\}$$
is equivalent to
$$\begin{cases} \langle p_j \, , \, d_j \rangle \leq b_j, & j = 1, \ldots, m \\ p_j D_j = x, & j = 1, \ldots, m \\ p_j \geq 0, & j = 1, \ldots, m \end{cases}$$

Type of constraints sets: ellipsoids

Here we consider **ellipsoids uncertainty**:

Theorem

The constraints $\langle a_j , x \rangle \leq b_j, \forall j$ and $(a_1, \ldots, a_m)^\top \in \mathcal{U}$ and $(a_1, \ldots, a_m)^\top \in \mathcal{U}$ where
$\mathcal{U} = \left\{ (a_1, \ldots, a_m)^\top \in (\mathbb{R}^k)^m \ : \ \forall j, a_j = a_j^0 + \Delta_j u_j, \|u\|_2^2 \leq \rho \right\}$
is equivalent to

$$\left\langle a_j^0 , x \right\rangle \leq b_j - \rho \left\| \Delta_j x \right\|_2 , \forall j$$

Proof: see[3]

[3] A. Ben-Tal, L. El Ghaoui, and A. Nemirovski. *Robust optimization*. Princeton Series in Applied Mathematics. Princeton University Press, Princeton, NJ, 2009, pp. xxii+542.

Type of constraints sets: cardinality constraints

Given an uncertain matrix, $A = (a_{i,j})$, suppose each component $a_{i,j}$ lies in $[a_{i,j} - \hat{a}_{i,j}, a_{i,j} + \hat{a}_{i,j}]$.
Rather than protect against the case when every parameter can deviate, one can allow at most Γ_i coefficients of row i to deviate.

Rem: in this case of LP, one can show that the natural convex relaxation associated would be exact[4]

[4] D. Bertsimas, D. B. Brown, and C. Caramanis. "Theory and applications of robust optimization". In: *SIAM Rev.* 53.3 (2011), pp. 464–501.

Table of Contents

Link: robust optimization and regularization

Adversarial point of view on regularization in regression (later generalization to any loss),[5] but could be seen as a robust optimization point of view[6]

Main message: regularization $\iff$ robust optimization

[5] H. Xu, C. Caramanis, and S. Mannor. "Robust regression and Lasso". In: *IEEE Trans. Inf. Theory* 56.7 (2010), pp. 3561–3574.

[6] A. Ben-Tal, L. El Ghaoui, and A. Nemirovski. *Robust optimization*. Princeton Series in Applied Mathematics. Princeton University Press, Princeton, NJ, 2009, pp. xxii+542.

Problem formulation

- n: number of samples
- p: number of features
- $y \in \mathbb{R}^n$: signal observed
- $X \in \mathbb{R}^{n \times p}$: design matrix

Goal: find a "good" weight β

$$\min_{\beta \in \mathbb{R}^p} \max_{\Delta \in \mathcal{C}} \|y - (X + \Delta)\beta\| \qquad (\star)$$

Interpretation: find good weights β such that under the worst (adversarial?) corruption of the design matrix authorized by the budget set $\mathcal{C} \subset \mathbb{R}^{n \times p}$, your estimate is still good

Regularization interpretation

Theorem

For $\mathcal{C} = \{\Delta = [\delta_1, \dots, \delta_p] \in \mathbb{R}^{n \times p} : \forall j \in [p], \|\delta_j\| \leq \lambda\}$

$$\min_{\beta \in \mathbb{R}^p} \max_{\Delta \in \mathcal{C}} \|y - (X + \Delta)\beta\| \qquad (\star)$$

is equivalent to solving the $\sqrt{\text{Lasso}}$[7] (aka Scaled Lasso[8]):

$$\min_{\beta \in \mathbb{R}^p} \|y - X\beta\| + \lambda \|\beta\|_1 \qquad (\star\star)$$

Rem: the set $\mathcal{C}$ is the (columnwise) $\ell_{2,\infty}$ unit ball

Interpretation: bounding the possible "corruption"/margin or error on the design is equivalent to adding an ℓ_1 penalty term

[7] A. Belloni, V. Chernozhukov, and L. Wang. "Square-root Lasso: pivotal recovery of sparse signals via conic programming". In: *Biometrika* 98.4 (2011), pp. 791–806.

[8] T. Sun and C.-H. Zhang. "Scaled sparse linear regression". In: *Biometrika* 99.4 (2012), pp. 879–898.

Proof

Reminder: $\mathcal{C} = \{\Delta = [\delta_1, \ldots, \delta_p] \in \mathbb{R}^{n\times p} : \forall j \in [p], \|\delta_j\| \leq \lambda\}$

$$V = \max_{\Delta \in \mathcal{C}} \|y - (X + \Delta)\beta\|$$

Proof

Reminder: $\mathcal{C} = \{\Delta = [\delta_1, \ldots, \delta_p] \in \mathbb{R}^{n\times p} : \forall j \in [p], \|\delta_j\| \leq \lambda\}$

$$
\begin{aligned}
V &= \max_{\Delta \in \mathcal{C}} \|y - (X + \Delta)\beta\| \\
&= \max_{\substack{[\delta_1,\ldots,\delta_p] \in \mathbb{R}^{n\times p} \\ \forall j \in [p], \|\delta_j\| \leq \lambda}} \left\| y - X\beta - \sum_{j=1}^{p} \delta_j \beta_j \right\|
\end{aligned}
$$

Proof

Reminder: $\mathcal{C} = \{\Delta = [\delta_1, \ldots, \delta_p] \in \mathbb{R}^{n \times p} : \forall j \in [p], \|\delta_j\| \leq \lambda\}$

$$
\begin{aligned}
V &= \max_{\Delta \in \mathcal{C}} \|y - (X + \Delta)\beta\| \\
&= \max_{\substack{[\delta_1,\ldots,\delta_p] \in \mathbb{R}^{n \times p} \\ \forall j \in [p], \|\delta_j\| \leq \lambda}} \left\| y - X\beta - \sum_{j=1}^{p} \delta_j \beta_j \right\| \\
&= \max_{\substack{[\delta_1,\ldots,\delta_p] \in \mathbb{R}^{n \times p} \\ \forall j \in [p], \|\delta_j\| \leq \lambda}} \max_{\substack{z \in \mathbb{R}^n \\ \|z\|_* \leq 1}} \left\langle y - X\beta - \sum_{j=1}^{p} \delta_j \beta_j \, , \, z \right\rangle
\end{aligned}
$$

Proof

Reminder: $\mathcal{C} = \{\Delta = [\delta_1, \ldots, \delta_p] \in \mathbb{R}^{n \times p} : \forall j \in [p], \|\delta_j\| \leq \lambda\}$

$$
\begin{aligned}
V &= \max_{\Delta \in \mathcal{C}} \|y - (X + \Delta)\beta\| \\
&= \max_{\substack{[\delta_1, \ldots, \delta_p] \in \mathbb{R}^{n \times p} \\ \forall j \in [p], \|\delta_j\| \leq \lambda}} \left\| y - X\beta - \sum_{j=1}^{p} \delta_j \beta_j \right\| \\
&= \max_{\substack{[\delta_1, \ldots, \delta_p] \in \mathbb{R}^{n \times p} \\ \forall j \in [p], \|\delta_j\| \leq \lambda}} \max_{\substack{z \in \mathbb{R}^n \\ \|z\|_* \leq 1}} \left\langle y - X\beta - \sum_{j=1}^{p} \delta_j \beta_j \,,\, z \right\rangle \\
&= \max_{\substack{z \in \mathbb{R}^n \\ \|z\|_* \leq 1}} \left(\langle y - X\beta \,,\, z \rangle + \max_{\substack{[\delta_1, \ldots, \delta_p] \in \mathbb{R}^{n \times p} \\ \forall j \in [p], \|\delta_j\| \leq \lambda}} \sum_{j=1}^{p} \langle -\delta_j \beta_j \,,\, z \rangle \right)
\end{aligned}
$$

Proof

Reminder: $\mathcal{C} = \{\Delta = [\delta_1, \ldots, \delta_p] \in \mathbb{R}^{n \times p} : \forall j \in [p], \|\delta_j\| \leq \lambda\}$

$$
\begin{aligned}
V &= \max_{\Delta \in \mathcal{C}} \|y - (X + \Delta)\beta\| \\
&= \max_{\substack{[\delta_1, \ldots, \delta_p] \in \mathbb{R}^{n \times p} \\ \forall j \in [p], \|\delta_j\| \leq \lambda}} \left\| y - X\beta - \sum_{j=1}^{p} \delta_j \beta_j \right\| \\
&= \max_{\substack{[\delta_1, \ldots, \delta_p] \in \mathbb{R}^{n \times p} \\ \forall j \in [p], \|\delta_j\| \leq \lambda}} \max_{\substack{z \in \mathbb{R}^n \\ \|z\|_* \leq 1}} \left\langle y - X\beta - \sum_{j=1}^{p} \delta_j \beta_j \, , \, z \right\rangle \\
&= \max_{\substack{z \in \mathbb{R}^n \\ \|z\|_* \leq 1}} \left(\langle y - X\beta \, , \, z \rangle + \max_{\substack{[\delta_1, \ldots, \delta_p] \in \mathbb{R}^{n \times p} \\ \forall j \in [p], \|\delta_j\| \leq \lambda}} \sum_{j=1}^{p} \langle -\delta_j \beta_j \, , \, z \rangle \right) \\
&= \max_{\substack{z \in \mathbb{R}^n \\ \|z\|_* \leq 1}} \langle y - X\beta \, , \, z \rangle + \lambda \sum_{j=1}^{p} |\beta_j| \, \|z\|_*
\end{aligned}
$$

Proof

Reminder: $\mathcal{C} = \{\Delta = [\delta_1, \ldots, \delta_p] \in \mathbb{R}^{n \times p} : \forall j \in [p], \|\delta_j\| \leq \lambda\}$

$$
\begin{aligned}
V &= \max_{\Delta \in \mathcal{C}} \|y - (X + \Delta)\beta\| \\
&= \max_{\substack{[\delta_1,\ldots,\delta_p] \in \mathbb{R}^{n\times p} \\ \forall j \in [p], \|\delta_j\| \leq \lambda}} \left\| y - X\beta - \sum_{j=1}^{p} \delta_j \beta_j \right\| \\
&= \max_{\substack{[\delta_1,\ldots,\delta_p] \in \mathbb{R}^{n\times p} \\ \forall j \in [p], \|\delta_j\| \leq \lambda}} \max_{\substack{z \in \mathbb{R}^n \\ \|z\|_* \leq 1}} \left\langle y - X\beta - \sum_{j=1}^{p} \delta_j \beta_j \, , \, z \right\rangle \\
&= \max_{\substack{z \in \mathbb{R}^n \\ \|z\|_* \leq 1}} \left(\langle y - X\beta \, , \, z \rangle + \max_{\substack{[\delta_1,\ldots,\delta_p] \in \mathbb{R}^{n\times p} \\ \forall j \in [p], \|\delta_j\| \leq \lambda}} \sum_{j=1}^{p} \langle -\delta_j \beta_j \, , \, z \rangle \right) \\
&= \max_{\substack{z \in \mathbb{R}^n \\ \|z\|_* \leq 1}} \langle y - X\beta \, , \, z \rangle + \lambda \sum_{j=1}^{p} |\beta_j| \, \|z\|_* \leq \|y - X\beta\| + \lambda \, \|\beta\|_1
\end{aligned}
$$

Proof (continued)

Reminding[9] that the sub-differential of a norm $\|\cdot\|$ at x, is given by

$$\partial \|x\| = \begin{cases} \{z \in \mathbb{R}^d : \|z\|_* \leq 1\} = \mathcal{B}_{\|\cdot\|_*}, & \text{if } x = 0, \\ \{z \in \mathbb{R}^d : \|z\|_* = 1 \text{ and } \langle z , x \rangle = \|x\|\}, & \text{otherwise.} \end{cases}$$

Then, z achieves equality in

$$\max_{\substack{z \in \mathbb{R}^n \\ \|z\|_* \leq 1}} \langle y - X\beta , z \rangle + \lambda \sum_{j=1}^{p} |\beta_j| \, \|z\|_* \leq \|y - X\beta\| + \lambda \|\beta\|_1$$

if and only if $z \in \partial \|\cdot\| (y - X\beta)$

[9] F. Bach et al. “Convex optimization with sparsity-inducing norms”. In: *Foundations and Trends in Machine Learning* 4.1 (2012), pp. 1–106, Proposition 1.2.

The worst perturbation

Proposition

The choice of perturbation achieving the largest deviation in the previous theorem is $\Delta = [\delta_1, \ldots, \delta_p]$ s.t. for all $j \in [p]$

$$\delta_j \in -\lambda \partial \left\| \cdot \right\|_* (z \beta_j)$$

Proof: one needs in the previous proof: $\langle -\delta_j \, , \, \beta_j z \rangle = \lambda \left\| \beta_j z \right\|_*$, hence $-\delta_j \in \lambda \partial \left\| \cdot \right\|_* (\beta_j z)$

Generalization beyond norms

A similar approach could be adapted to consider:

$$\min_{\beta \in \mathbb{R}^p} \max_{\Delta \in \mathcal{C}} f(y - (X + \Delta)\beta)$$

instead of

$$\min_{\beta \in \mathbb{R}^p} \max_{\Delta \in \mathcal{C}} \|y - (X + \Delta)\beta\|$$

for a (close) convex function f.

Main tool: Fenchel duality + Fenchel-Young inequality

References I

- Bach, F. et al. "Convex optimization with sparsity-inducing norms". In: *Foundations and Trends in Machine Learning* 4.1 (2012), pp. 1–106.
- Belloni, A., V. Chernozhukov, and L. Wang. "Square-root Lasso: pivotal recovery of sparse signals via conic programming". In: *Biometrika* 98.4 (2011), pp. 791–806.
- Ben-Tal, A., L. El Ghaoui, and A. Nemirovski. *Robust optimization*. Princeton Series in Applied Mathematics. Princeton University Press, Princeton, NJ, 2009, pp. xxii+542.
- Bertsimas, D., D. B. Brown, and C. Caramanis. "Theory and applications of robust optimization". In: *SIAM Rev.* 53.3 (2011), pp. 464–501.
- Sun, T. and C.-H. Zhang. "Scaled sparse linear regression". In: *Biometrika* 99.4 (2012), pp. 879–898.
- Xu, H., C. Caramanis, and S. Mannor. "Robust regression and Lasso". In: *IEEE Trans. Inf. Theory* 56.7 (2010), pp. 3561–3574.