Sensitivity Analysis With Degeneracy: Mirror Stratifiable Functions

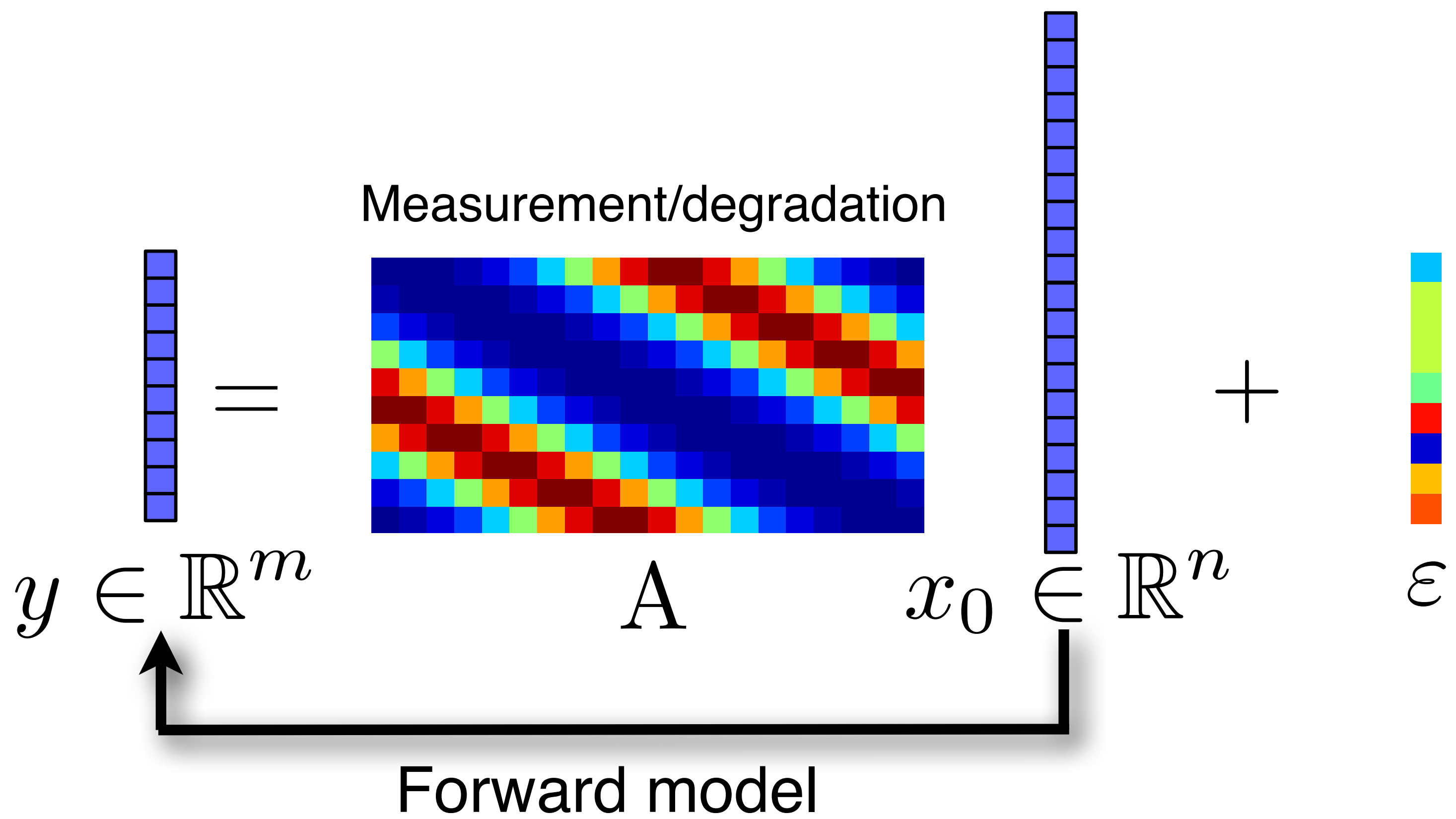
Jalal Fadili
Normandie Université-ENSICAEN, GREYC CNRS UMR 6072
Joint work with
Jérôme Malick and Gabriel Peyré

Journées GdR MOA-MIA Bordeaux 2017

Normandie Université

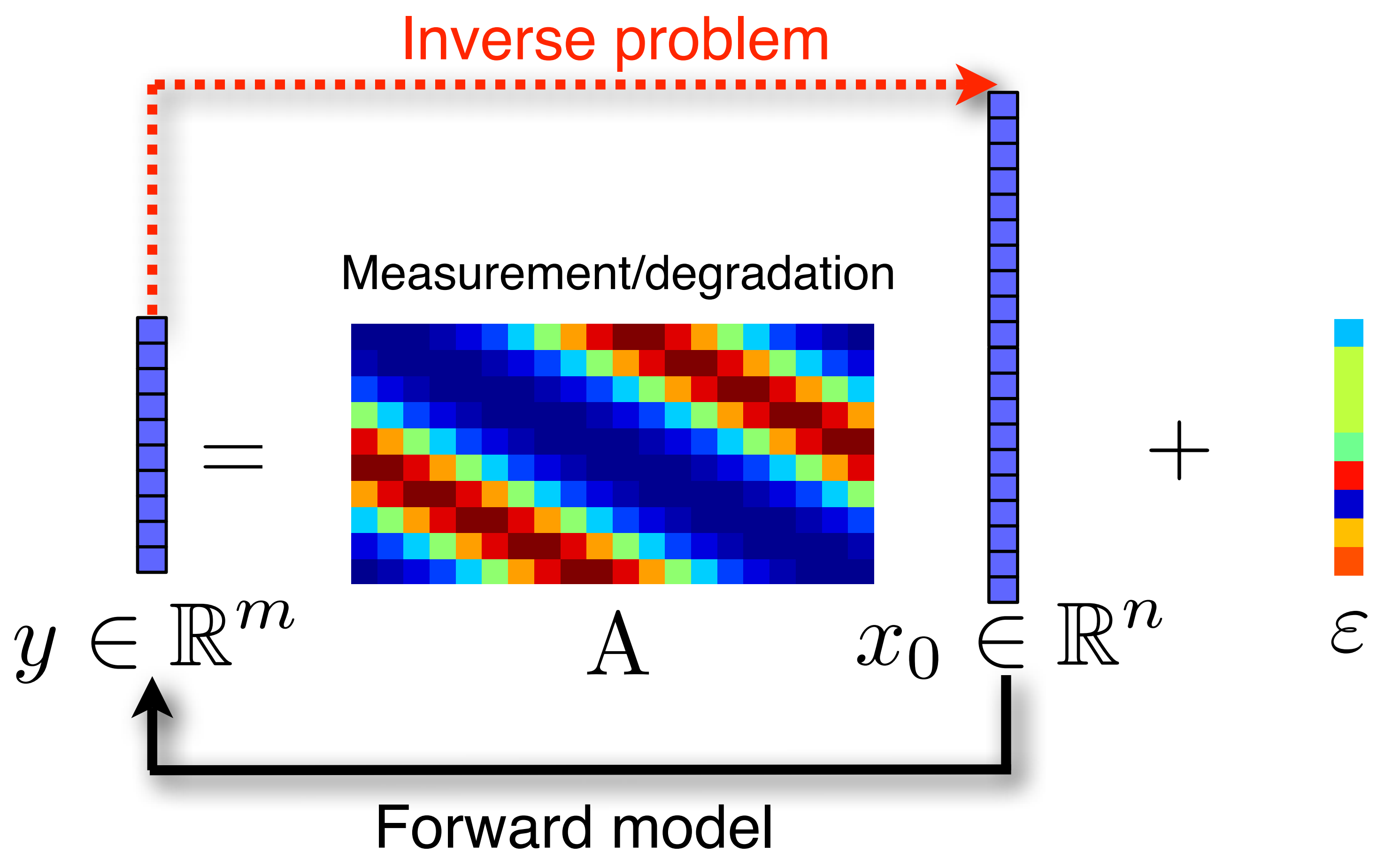
ÉCOLE PUBLIQUE D'INGÉNIEURS
CENTRE DE RECHERCHE

Linear inverse problems



- Throughout the talk : finite-dimensional real setting.
- ε is the noise.
- $\mathrm{A} : \mathbb{R}^m \to \mathbb{R}^n$ is a linear measurement/degradation operator.

Linear inverse problems

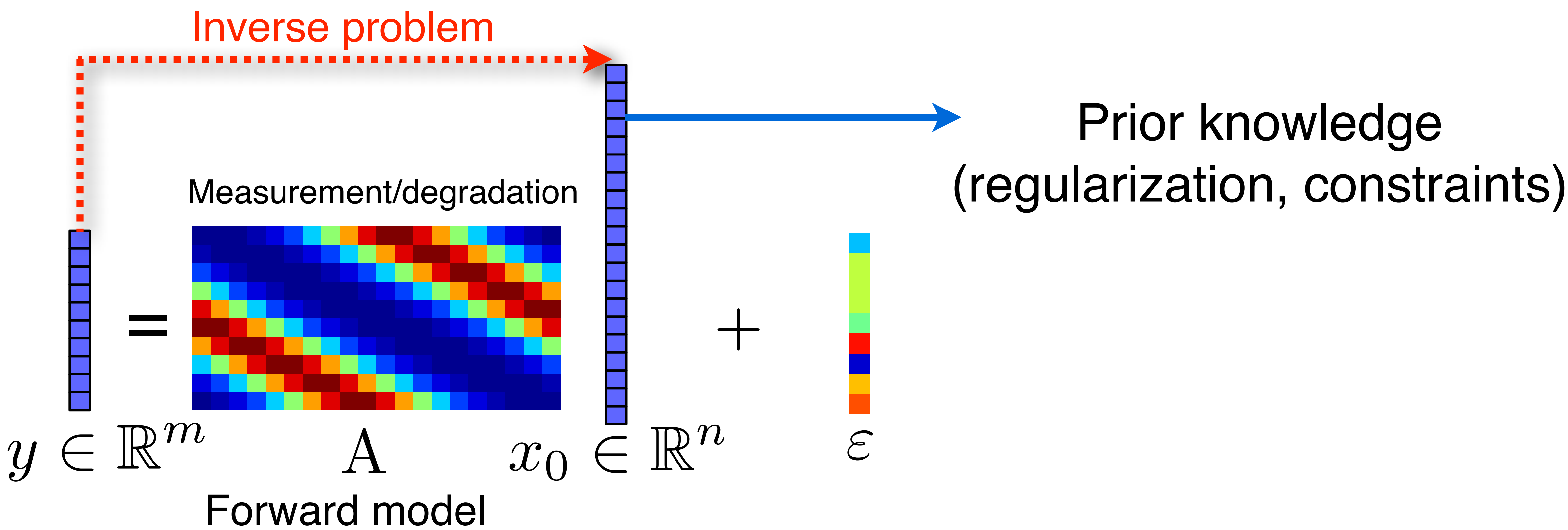


- Throughout the talk : finite-dimensional real setting.
- ε is the noise.
- $\mathrm{A} : \mathbb{R}^m \to \mathbb{R}^n$ is a linear measurement/degradation operator.

Objective

Recover x_0 from y is an ill-posed inverse problem.

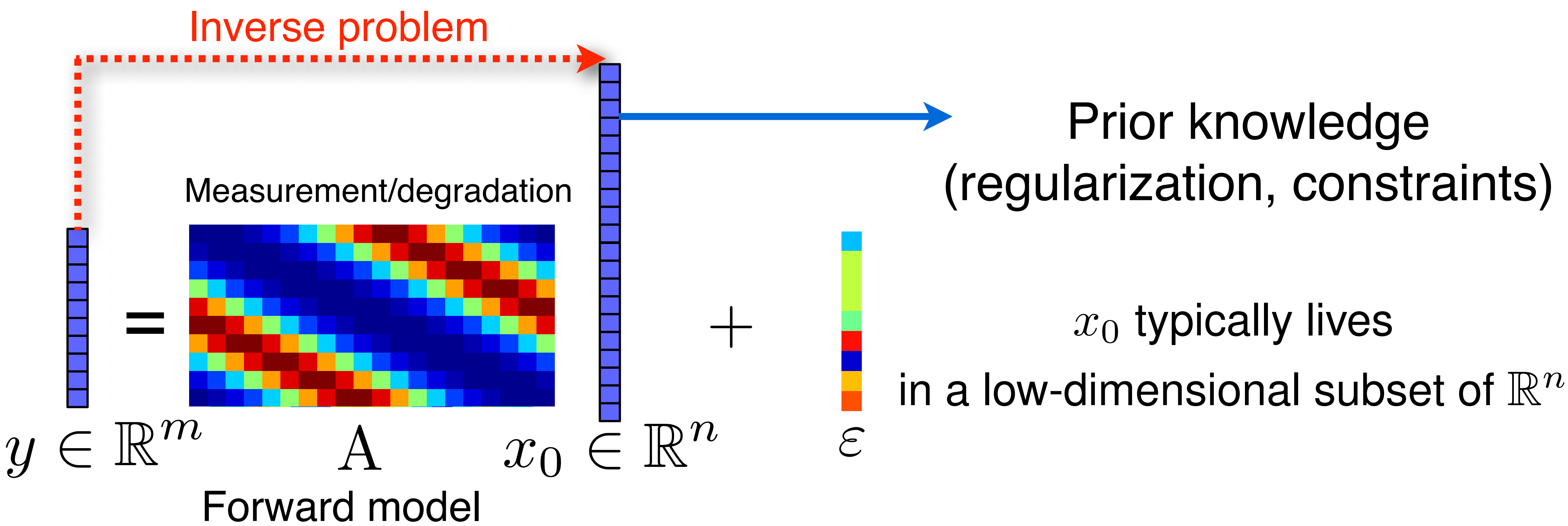
Regularization



x_0 typically lives

in a low-dimensional subset of $\mathbb{R}^n$

Regularization



- Many applications in data sciences: signal/image processing, machine learning, statistics, etc..
- Solve an inverse problem through regularization :

$$x^\star(y,\lambda) \in \operatorname*{Argmin}_{x\in\mathbb{R}^n} \underbrace{F(\mathrm{A}x, y)}_{\text{Data fidelity}} + \lambda \underbrace{R(x)}_{\text{Regularization, constraints}}$$

$F(\cdot, y)$ and $R \in \Gamma_0(\mathbb{R}^n)$

- R promotes objects living in the same manifold as x_0.

Low-complexity regularization

$$\min_{x \in \mathbb{R}^n} F(\mathrm{A}x, y) + \lambda R(x) \qquad F(\cdot, y) \text{ and } R \in \Gamma_0(\mathbb{R}^n)$$

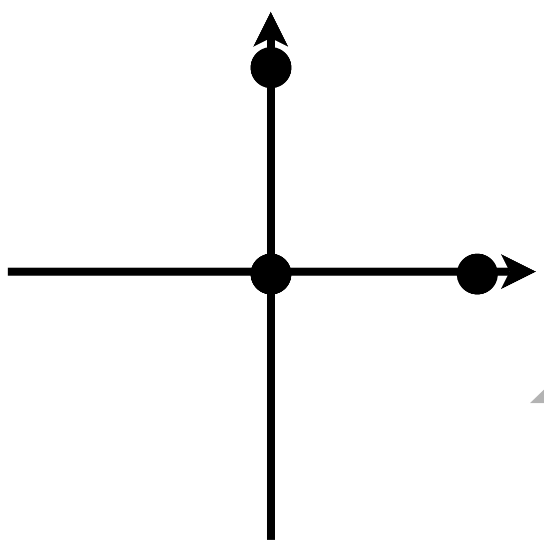
Low-complexity of x_0 $\iff$ Low-dimensional subset M_{x_0}

Low-complexity regularization

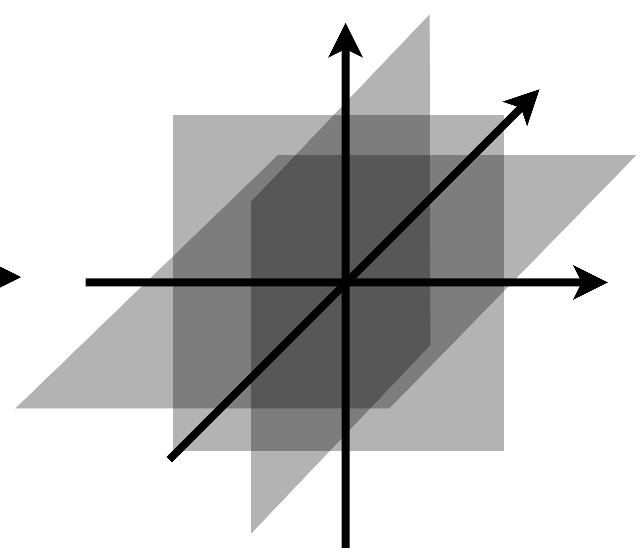
$$\min_{x\in\mathbb{R}^n} F(\mathrm{A}x, y) + \lambda R(x) \qquad F(\cdot, y) \text{ and } R \in \Gamma_0(\mathbb{R}^n)$$

Low-complexity of x_0 $\Longleftrightarrow$ Low-dimensional subset M_{x_0}

Sparse vectors



$\mathbb{R}^2$



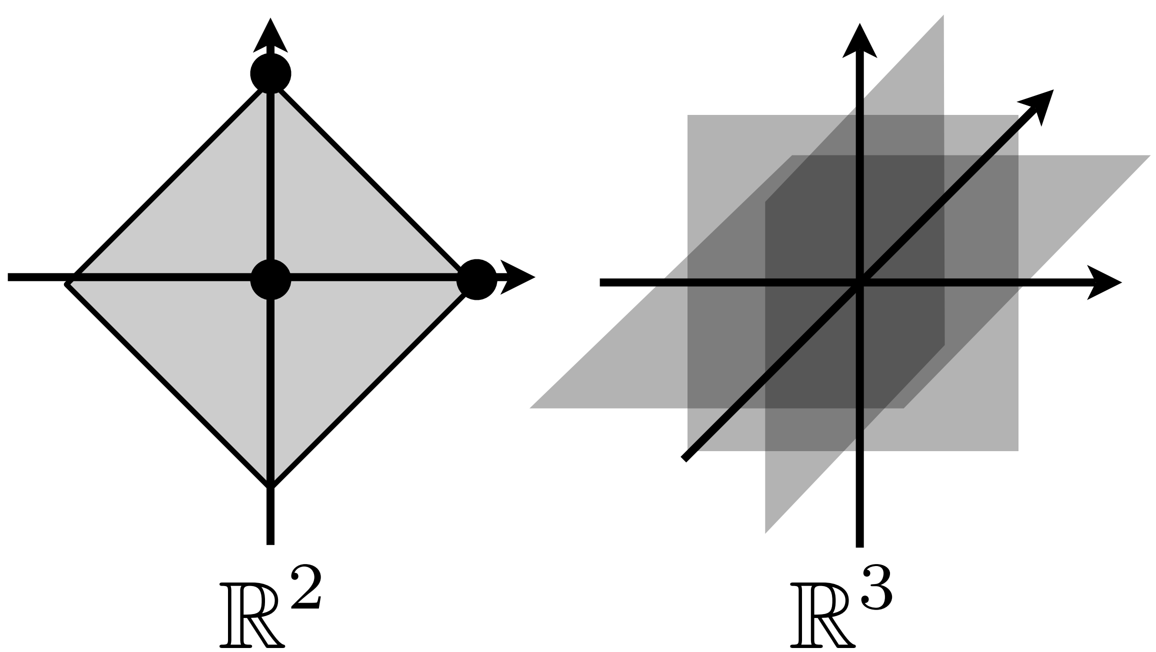
$\mathbb{R}^3$

Low-complexity regularization

$$\min_{x \in \mathbb{R}^n} F(\mathrm{A}x, y) + \lambda R(x) \qquad F(\cdot, y) \text{ and } R \in \Gamma_0(\mathbb{R}^n)$$

Low-complexity of x_0 $\Longleftrightarrow$ Low-dimensional subset M_{x_0}

Sparse vectors



$$R(x) = \|x\|_1$$

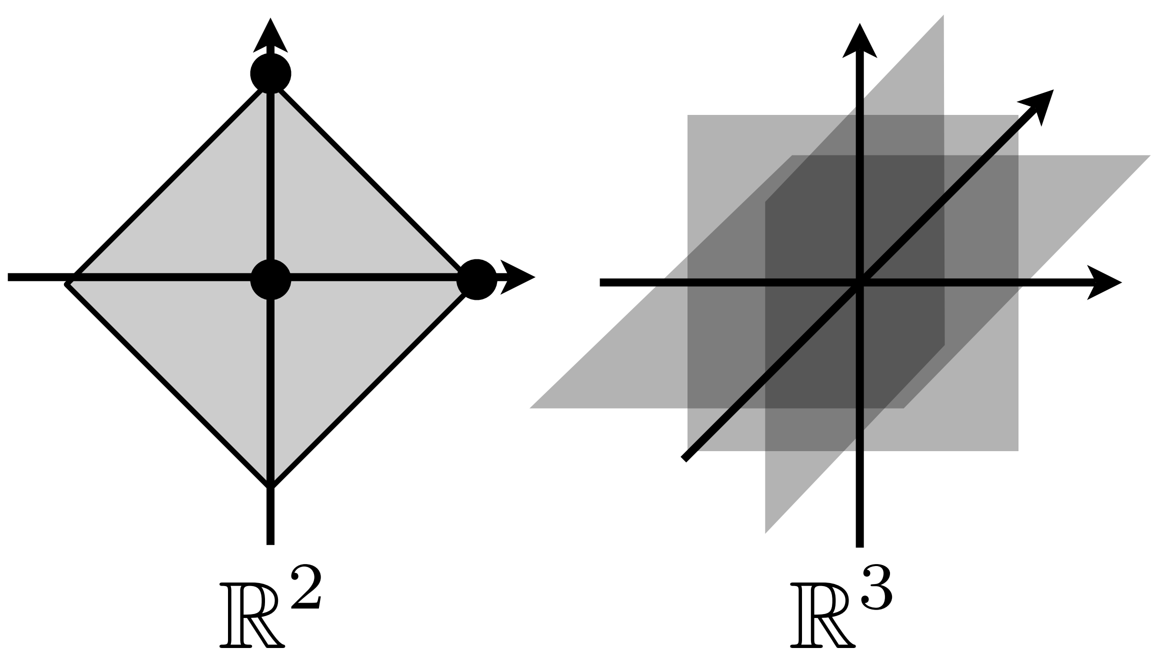
(tightest convex relaxation of ℓ_0)

Low-complexity regularization

$$\min_{x \in \mathbb{R}^n} F(\mathrm{A}x, y) + \lambda R(x) \qquad F(\cdot, y) \text{ and } R \in \Gamma_0(\mathbb{R}^n)$$

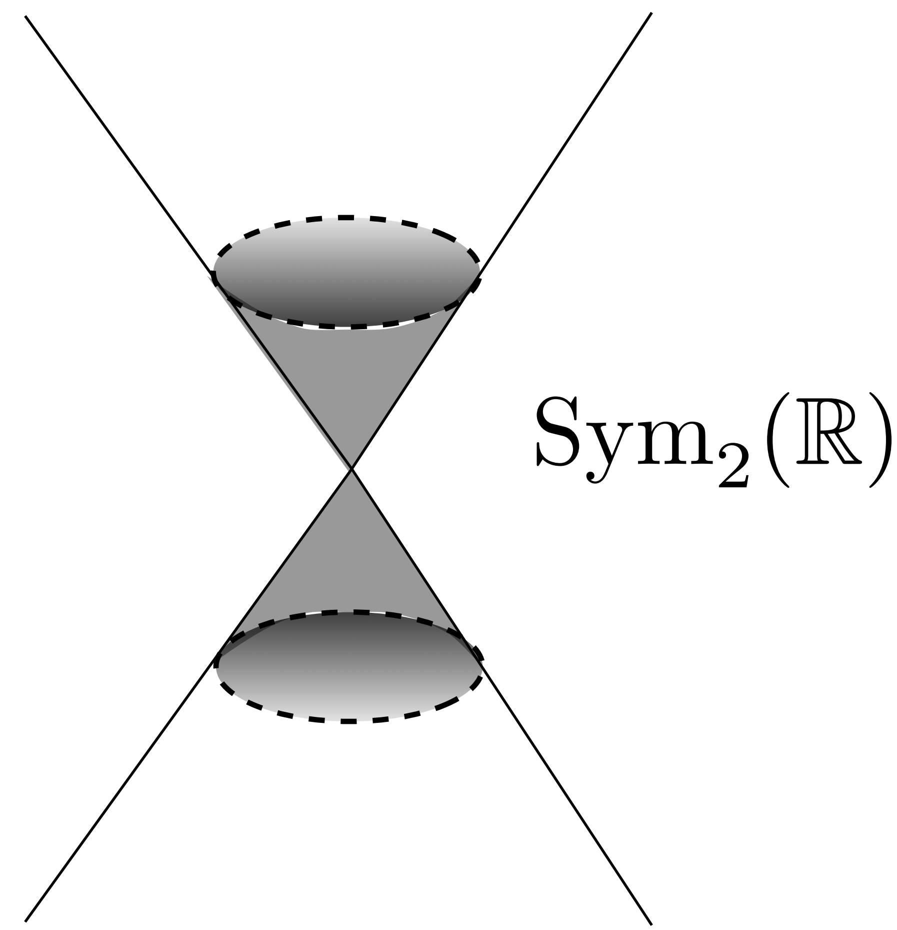
Low-complexity of x_0 $\Longleftrightarrow$ Low-dimensional subset M_{x_0}

Sparse vectors



$R(x) = \|x\|_1$
(tightest convex relaxation of ℓ_0)

Low-rank matrices

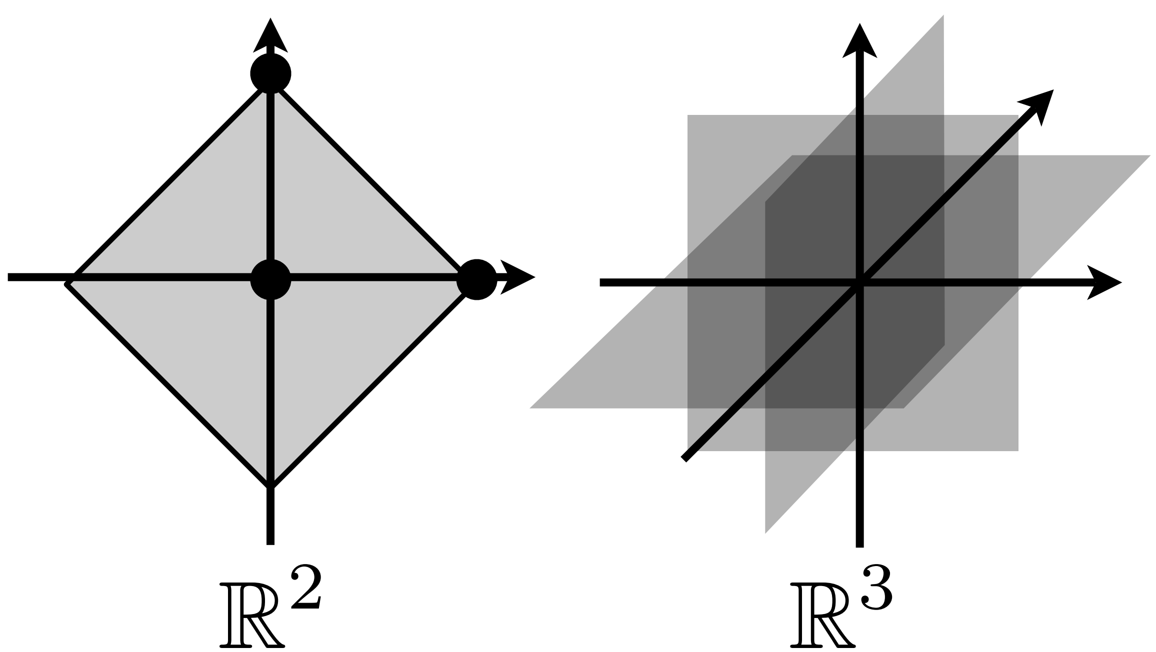


Low-complexity regularization

$$\min_{x \in \mathbb{R}^n} F(\mathrm{A}x, y) + \lambda R(x) \qquad F(\cdot, y) \text{ and } R \in \Gamma_0(\mathbb{R}^n)$$

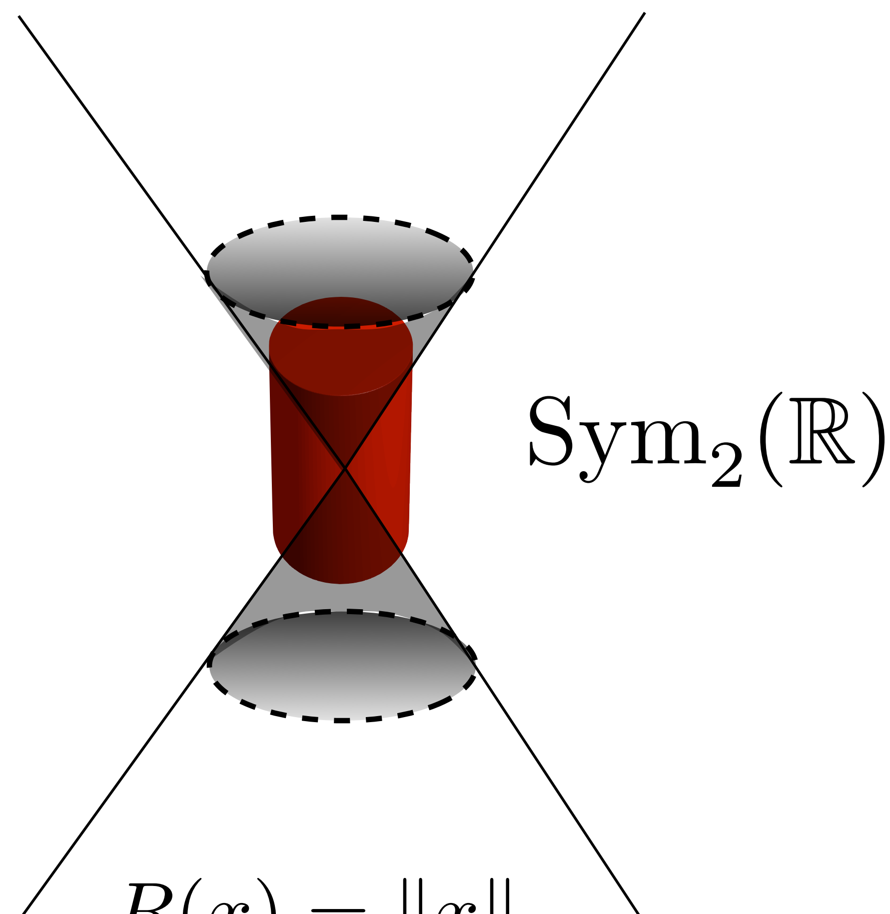
Low-complexity of x_0 $\Longleftrightarrow$ Low-dimensional subset M_{x_0}

Sparse vectors



$R(x) = \|x\|_1$
(tightest convex relaxation of ℓ_0)

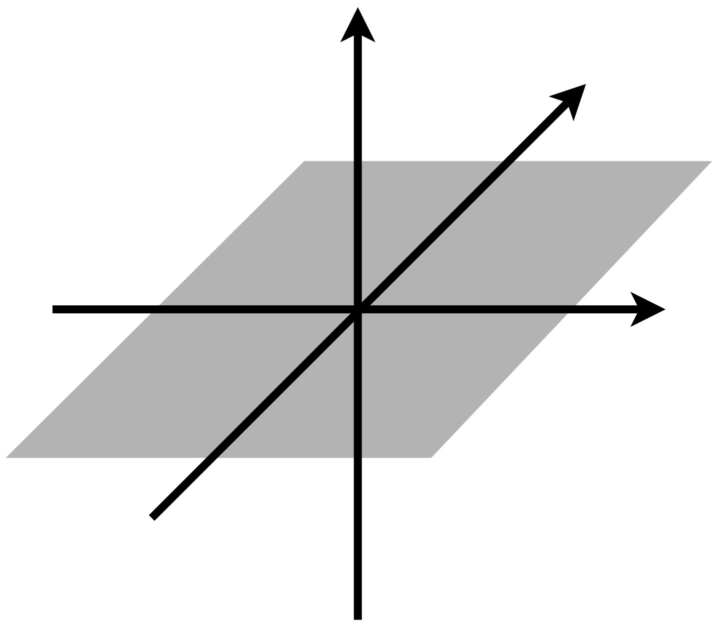
Low-rank matrices

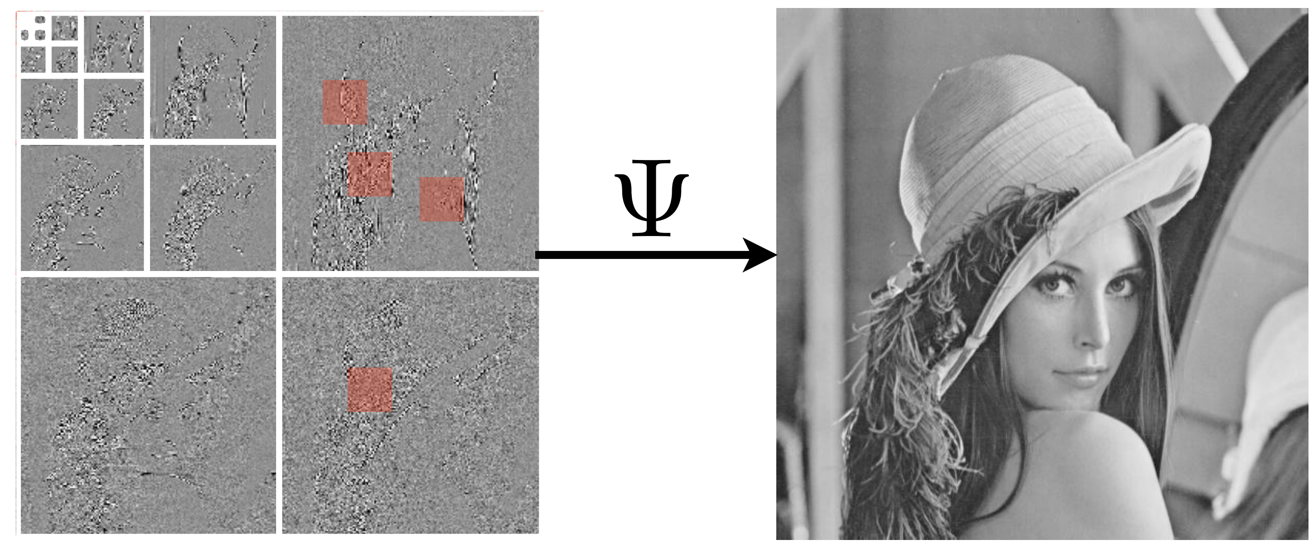


$R(x) = \|x\|_*$
(tightest convex relaxation of rank)

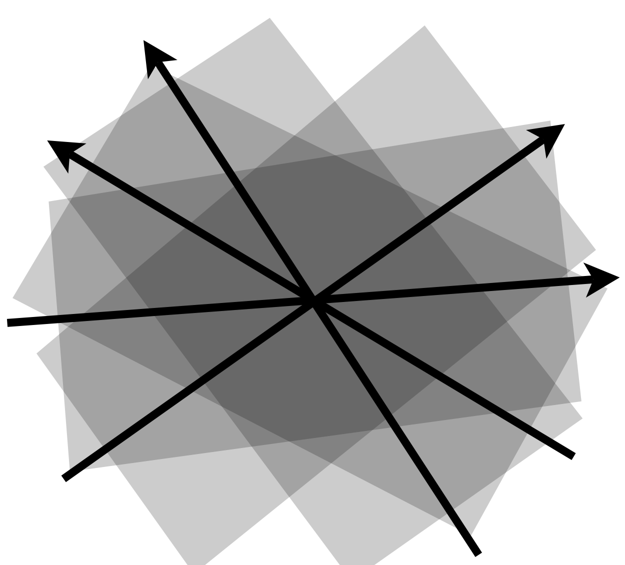
Low-complexity regularization

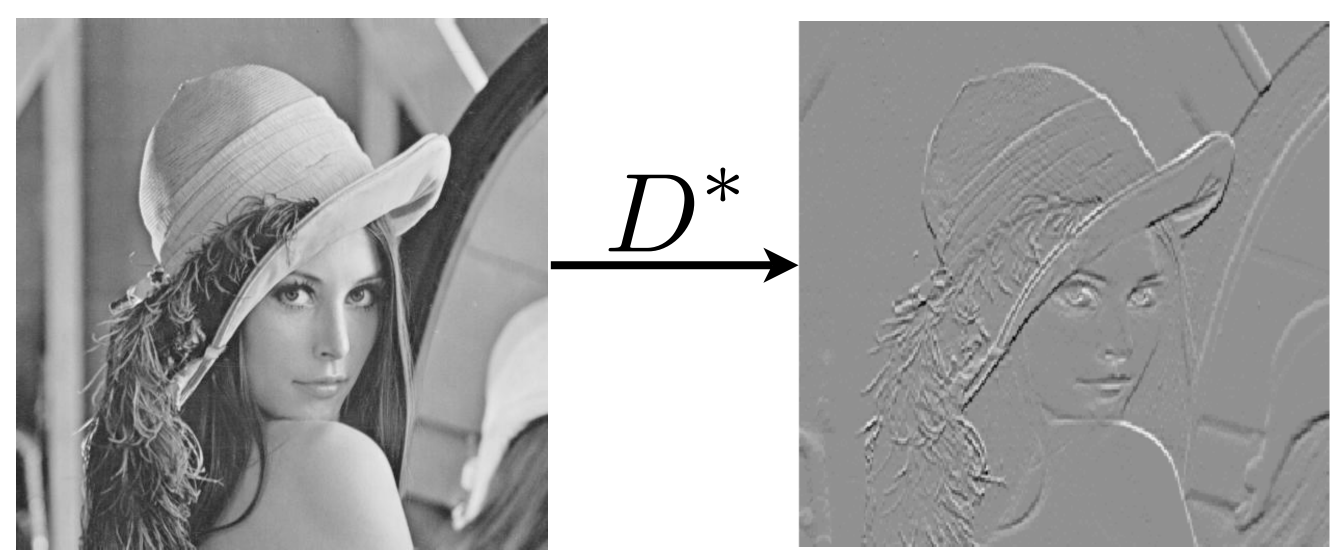
Group sparsity





Analysis sparsity





Regularization guarantees

$$y = \mathrm{A}x_0 + \varepsilon$$

$$x^\star(y,\lambda) \in \operatorname*{Argmin}_{x \in \mathbb{R}^n} F(\mathrm{A}x, y) + \lambda R(x)$$

Regularization guarantees

$$y = \mathrm{A}x_0 + \varepsilon$$

$$x^\star(y,\lambda) \in \operatorname*{Argmin}_{x\in\mathbb{R}^n} F(\mathrm{A}x, y) + \lambda R(x)$$

- Which conditions to ensure :
  - Exact recovery in the noiseless case : $x^\star(0^+, \mathrm{A}x_0) = x_0$.
  - Stable recovery : $\varphi(x^\star(y,\lambda), x_0) = O(\|\varepsilon\|)$.
  - $x^\star(y,\lambda) \in M_{x_0}$ and/or $M_{x^\star(y,\lambda)} \subseteq M_{x_0}$.

Regularization guarantees

$$y = \mathrm{A}x_0 + \varepsilon$$

$$x^\star(y,\lambda) \in \underset{x \in \mathbb{R}^n}{\mathrm{Argmin}}\ F(\mathrm{A}x, y) + \lambda R(x)$$

- Which conditions to ensure :
  - Exact recovery in the noiseless case : $x^\star(0^+, \mathrm{A}x_0) = x_0$.
  - Stable recovery : $\varphi(x^\star(y,\lambda), x_0) = O(\|\varepsilon\|)$.
  - $x^\star(y,\lambda) \in M_{x_0}$ and/or $M_{x^\star(y,\lambda)} \subseteq M_{x_0}$.
- Algorithmic implications :
  - Run an algorithm $\mathcal{A}$ and generate iterates $x_k \to x^\star(y,\lambda)$.
  - Which conditions guarantee finite activity identification

$$x_k \in M_{x^\star(y,\lambda)} \quad \text{and/or} \quad M_{x_k} \subseteq M_{x^\star(y,\lambda)}.$$

Notations

- ∂R : subdifferential of R.
- N_C : normal cone of C.
- $\mathrm{Aff}(C)$: affine hull of C.
- $\mathrm{par}(C)$: subspace parallel to C.
- $\mathrm{ri}(C)$: relative interior of C.
- $S_x = \mathrm{par}(\partial R(x)), T_x = S_x^{\perp}$.

Known results: stable recovery

$$y = \mathrm{A}x_0 + \varepsilon$$

$$x^\star(y, \lambda) \in \operatorname*{Argmin}_{x \in \mathbb{R}^n} \frac{1}{2} \left\| y - \mathrm{A}x \right\|^2 + \lambda R(x)$$

Known results: stable recovery

$$y = \mathrm{A}x_0 + \varepsilon$$

$$x^\star(y,\lambda) \in \underset{x\in\mathbb{R}^n}{\mathrm{Argmin}}\ \frac{1}{2}\|y - \mathrm{A}x\|^2 + \lambda R(x)$$

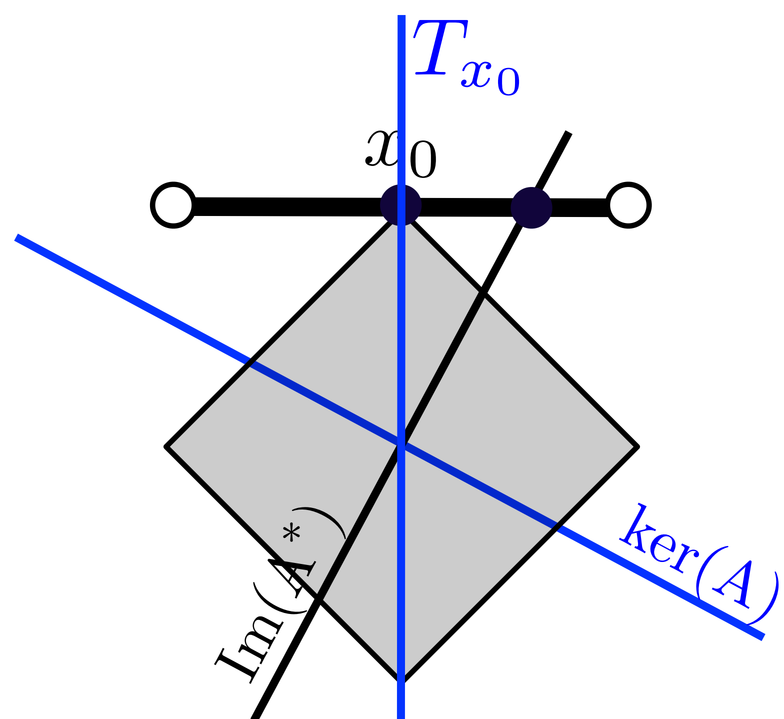
Theorem *Assume that*

Non-degeneracy condition

$T_{x_0} = \mathrm{par}(\partial R(x_0))^\perp$ $\quad \boxed{\mathrm{Im}(\mathrm{A}^*) \cap \mathrm{ri}\,(\partial R(x_0))}$ *and* $\ker(\mathrm{A}) \cap T_{x_0} = \{0\}$.

Then, choosing $\lambda = c\,\|\varepsilon\|$, $c > 0$, *any minimizer* $x^\star(y,\lambda)$ *obeys*

$$\|x^\star(y,\lambda) - x_0\| = \|x^\star(y_0+\varepsilon,\lambda) - x^\star_{y_0,0^+}\| = O(\|\varepsilon\|)\ .$$



Known results: stable model recovery

$$y = \mathrm{A}x_0 + \varepsilon$$

$$x^\star(y,\lambda) \in \underset{x\in\mathbb{R}^n}{\mathrm{Argmin}}\ \frac{1}{2}\|y - \mathrm{A}x\|^2 + \lambda R(x)$$

Minimal ℓ_2-norm certificate $q_0 = \mathrm{A}_{T_{x_0}}^{+,*} e_{x_0}$.

$T_{x_0} = \mathrm{par}(\partial R(x_0))^\perp$
$e_{x_0} = \mathrm{P}_{\mathrm{Aff}(\partial R(x_0))}(0)$

Known results: stable model recovery

$$y = \mathrm{A}x_0 + \varepsilon$$

$$x^\star(y,\lambda) \in \underset{x\in\mathbb{R}^n}{\mathrm{Argmin}}\ \frac{1}{2}\|y - \mathrm{A}x\|^2 + \lambda R(x)$$

Minimal ℓ_2-norm certificate $q_0 = \mathrm{A}_{T_{x_0}}^{+,*} e_{x_0}$.

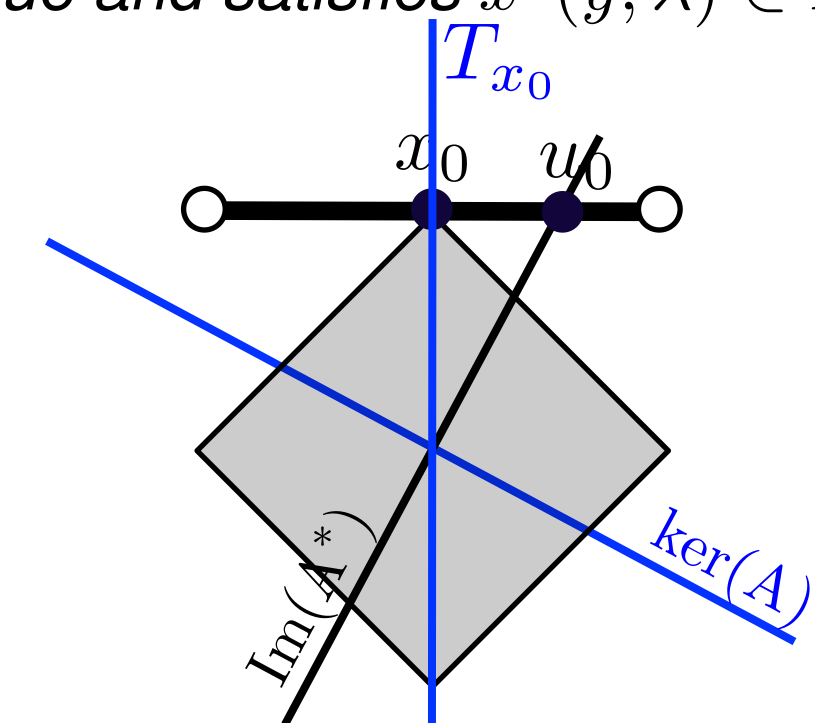
$T_{x_0} = \mathrm{par}(\partial R(x_0))^\perp$
$e_{x_0} = \mathrm{P}_{\mathrm{Aff}(\partial R(x_0))}(0)$

Theorem *Let M_{x_0} be the "model" subset of x_0. Assume that R has a nice structure on a nice subset M_{x_0} nearby x_0, and that*

Non-degeneracy condition

$$\boxed{u_0 = \mathrm{A}^* q_0 \in \mathrm{ri}\,(\partial R(x_0))} \quad \textit{and} \quad \ker(\mathrm{A}) \cap T_{x_0} = \{0\}.$$

Then $\exists C_0, C_1 > 0$ depending only on x_0 such that for all (ε,λ) s.t. $C_0\|\varepsilon\| \leq \lambda \leq C_1$, the solution $x^\star(\lambda, y)$ is unique and satisfies $x^\star(y,\lambda) \in M_{x_0}$.



Known results: finite activity identification

$$\min_{x \in \mathbb{R}^n} \frac{1}{2} \|y - \mathrm{A}x\|^2 + \lambda R(x)$$

$$\gamma_k \in \left[\epsilon, 2/|||\mathrm{A}|||^2 - \epsilon\right] \quad x_{k+1} = \mathrm{prox}_{\lambda \gamma_k R}\left(x_k + \gamma_k \mathrm{A}^*(y - \mathrm{A}x_k)\right)$$

Known results: finite activity identification

$$\min_{x\in\mathbb{R}^n} \frac{1}{2}\|y - \mathrm{A}x\|^2 + \lambda R(x)$$

$$\gamma_k \in \left[\epsilon, 2/|||\mathrm{A}|||^2 - \epsilon\right] \quad x_{k+1} = \mathrm{prox}_{\lambda\gamma_k R}\left(x_k + \gamma_k \mathrm{A}^*(y - \mathrm{A}x_k)\right)$$

Theorem *Let the FB be used to create a sequence x_k which converges to $x^\star$. Assume that R has a nice structure on a nice subset $M_{x^\star}$ nearby $x^\star$, and*

Non-degeneracy condition

$$\boxed{\mathrm{A}^*(y - Ax^\star) \in \lambda \mathrm{ri}\ (\partial R(x^\star))\,.}$$

Then $x_k \in M_{x^\star}$ for k large enough.

Known results: finite activity identification

$$\min_{x \in \mathbb{R}^n} \frac{1}{2} \|y - \mathrm{A}x\|^2 + \lambda R(x)$$

$$\gamma_k \in \left[\epsilon, 2/|||\mathrm{A}|||^2 - \epsilon\right] \quad x_{k+1} = \mathrm{prox}_{\lambda\gamma_k R}\left(x_k + \gamma_k \mathrm{A}^*(y - \mathrm{A}x_k)\right)$$

Theorem *Let the FB be used to create a sequence x_k which converges to $x^\star$. Assume that R has a nice structure on a nice subset $M_{x^\star}$ nearby $x^\star$, and*

Non-degeneracy condition

$$\mathrm{A}^*(y - Ax^\star) \in \lambda \mathrm{ri}\ (\partial R(x^\star))\,.$$

Then $x_k \in M_{x^\star}$ for k large enough.

If R has a nice structure on a nice subset M_{x_0} nearby x_0, and

Non-degeneracy condition

$$u_0 = \mathrm{A}^* q_0 \in \mathrm{ri}\ (\partial R(x_0)) \quad \textit{and} \quad \ker(\mathrm{A}) \cap T_{x_0} = \{0\}.$$

Then $\exists C_0, C_1 > 0$ depending only on x_0 such that for all (ε, λ) s.t. $C_0 \|\varepsilon\| \leq \lambda \leq C_1$, $x_k \in M_{x_0}$ for k large enough.

Regularization guarantees: summary

Regularization guarantees: summary

- Non-degeneracy is ubiquitous.

Regularization guarantees: summary

- Non-degeneracy is ubiquitous.
- Structure of the regularizer is important.

Regularization guarantees: summary

- Non-degeneracy is ubiquitous.
- Structure of the regularizer is important.
- Does non-degeneracy hold in practical problems:

Regularization guarantees: summary

- Non-degeneracy is ubiquitous.
- Structure of the regularizer is important.
- Does non-degeneracy hold in practical problems:
  - Yes in e.g. compressed sensing (CS):

Regularization guarantees: summary

- Non-degeneracy is ubiquitous.
- Structure of the regularizer is important.
- Does non-degeneracy hold in practical problems:
  - Yes in e.g. compressed sensing (CS):

Proposition $R = \|\cdot\|_1$, $\mathrm{A}_{i,j} \sim_{iid} \mathcal{N}(0, 1)$, *and* $s = \|x_0\|_0$.

(i) If $m > 2s\log(n/s) + 7/5s$*, then there exists a non-degenerate dual certificate.*

(ii) If $m > 2ts\log(n) + s$*,* $t > 1$*, then* q_0 *is a non-degenerate dual certificate.*

Regularization guarantees: summary

- Non-degeneracy is ubiquitous.
- Structure of the regularizer is important.
- Does non-degeneracy hold in practical problems:
  - Yes in e.g. compressed sensing (CS):

Proposition $R = \|\cdot\|_1$, $\mathrm{A}_{i,j} \sim_{iid} \mathcal{N}(0,1)$, *and* $s = \|x_0\|_0$.

(i) If $m > 2s\log(n/s) + 7/5s$*, then there exists a non-degenerate dual certificate.*

(ii) If $m > 2ts\log(n) + s$*,* $t > 1$*, then* q_0 *is a non-degenerate dual certificate.*

  - Mostly unlikely for challenging inverse problems (e.g. deconvolution).

Regularization guarantees: summary

- Non-degeneracy is ubiquitous.
- Structure of the regularizer is important.
- Does non-degeneracy hold in practical problems:
  - Yes in e.g. compressed sensing (CS):

Proposition $R = \|\cdot\|_1$, $\mathrm{A}_{i,j} \sim_{iid} \mathcal{N}(0,1)$, *and* $s = \|x_0\|_0$.

(i) If $m > 2s\log(n/s) + 7/5s$*, then there exists a non-degenerate dual certificate.*

(ii) If $m > 2ts\log(n) + s$*,* $t > 1$*, then* q_0 *is a non-degenerate dual certificate.*

  - Mostly unlikely for challenging inverse problems (e.g. deconvolution).
- What about these intricate situations ?

Regularization guarantees: summary

- Non-degeneracy is ubiquitous.
- Structure of the regularizer is important.
- Does non-degeneracy hold in practical problems:
  - Yes in e.g. compressed sensing (CS):

Proposition $R = \|\cdot\|_1$, $\mathrm{A}_{i,j} \sim_{iid} \mathcal{N}(0,1)$, *and* $s = \|x_0\|_0$.

(i) If $m > 2s\log(n/s) + 7/5s$, *then there exists a non-degenerate dual certificate.*

(ii) If $m > 2ts\log(n) + s$, $t > 1$, *then* q_0 *is a non-degenerate dual certificate.*

  - Mostly unlikely for challenging inverse problems (e.g. deconvolution).
- What about these intricate situations ?
- What about even CS with insufficient number of measurements ?

Regularization guarantees: summary

- Non-degeneracy is ubiquitous.
- Structure of the regularizer is important.
- Does non-degeneracy hold in practical problems:
  - Yes in e.g. compressed sensing (CS):

Proposition $R = \|\cdot\|_1$, $\mathrm{A}_{i,j} \sim_{iid} \mathcal{N}(0,1)$, *and* $s = \|x_0\|_0$.

(i) If $m > 2s\log(n/s) + 7/5s$, *then there exists a non-degenerate dual certificate.*

(ii) If $m > 2ts\log(n) + s$, $t > 1$, *then* q_0 *is a non-degenerate dual certificate.*

  - Mostly unlikely for challenging inverse problems (e.g. deconvolution).
- What about these intricate situations ?
- What about even CS with insufficient number of measurements ?
- Can we say anything at all under degeneracy ?

Regularization guarantees: recap

The answer is YES if we have some structure

Regularization guarantees: recap

The answer is YES if we have some structure

- Known results:
  - Assume non-degeneracy and some structure.
  - Exact model recovery.

Regularization guarantees: recap

The answer is YES if we have some structure

- Known results:
  - Assume non-degeneracy and some structure.
  - Exact model recovery.
- Our results in this talk:
  - Degeneracy and some structure.
  - Some model recovery: model localization.

Outline

- Mirror-stratifiable functions.
- Sensitivity analysis.
- Regularized inverse problems.
- Algorithmic implications.
- Numerical results.
- Conclusion.

Outline

- Mirror-stratifiable functions.
- Sensitivity analysis.
- Regularized inverse problems.
- Algorithmic implications.
- Numerical results.
- Conclusion.

Stratification

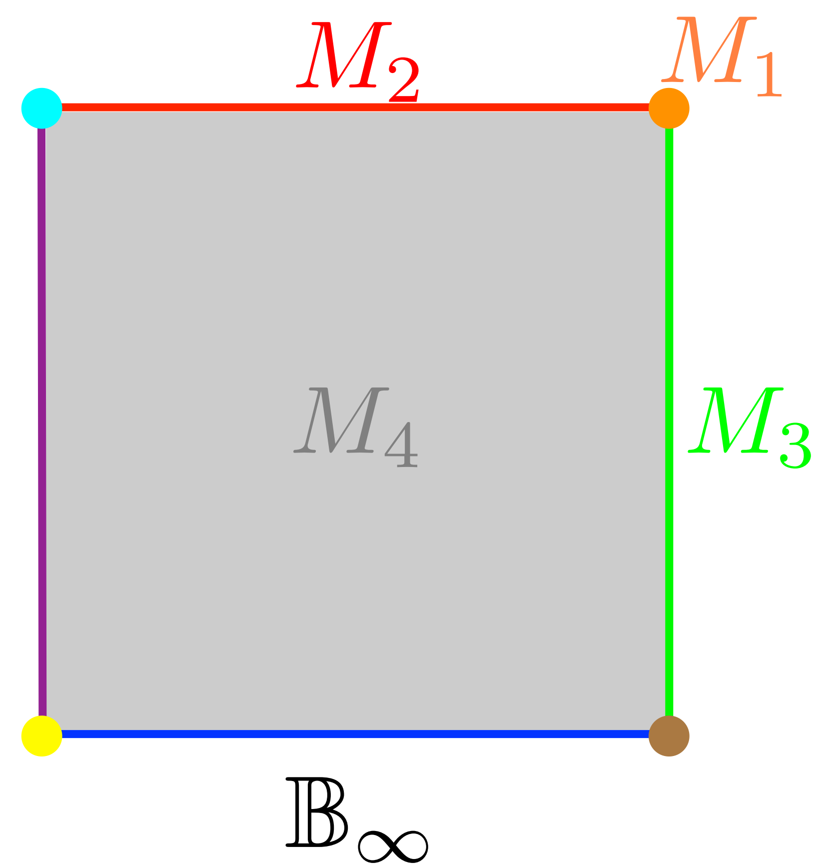
Definition *A stratification of a set $D \subset \mathbb{R}^n$ is a finite disjoint partition $\mathcal{M} = \{M_i\}_{i \in I}$ such that the partitioning sets (strata) fit nicely, i.e.* $\forall (M, M') \in \mathcal{M}^2$ *we have*

$$M \cap \mathrm{cl}(M') \neq \emptyset \implies M \subset \mathrm{cl}(M').$$

Stratification

Definition *A stratification of a set* $D \subset \mathbb{R}^n$ *is a finite disjoint partition* $\mathcal{M} = \{M_i\}_{i \in I}$ *such that the partitioning sets (strata) fit nicely, i.e.* $\forall (M, M') \in \mathcal{M}^2$ *we have*

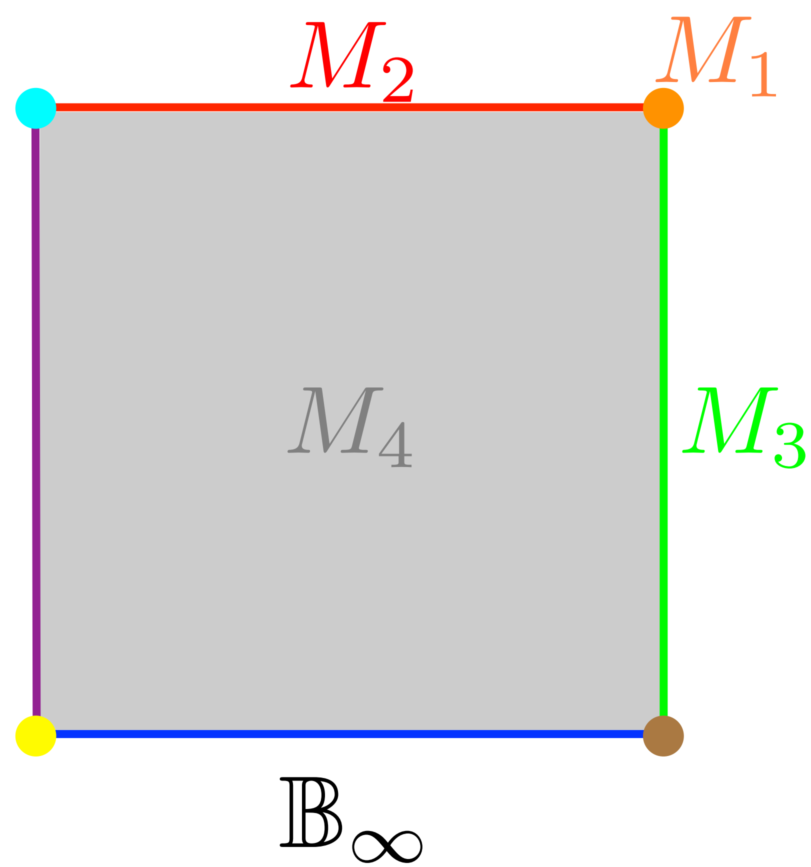
$$M \cap \mathrm{cl}(M') \neq \emptyset \implies M \subset \mathrm{cl}(M').$$



Stratification

Definition *A stratification of a set* $D \subset \mathbb{R}^n$ *is a finite disjoint partition* $\mathcal{M} = \{M_i\}_{i \in I}$ *such that the partitioning sets (strata) fit nicely, i.e.* $\forall (M, M') \in \mathcal{M}^2$ *we have*

$$M \cap \mathrm{cl}(M') \neq \emptyset \implies M \subset \mathrm{cl}(M').$$



- A stratification is naturally endowed with the partial ordering $\leq$:

$$M \leq M' \iff M \subset \mathrm{cl}(M') \iff M \cap \mathrm{cl}(M') \neq \emptyset.$$

$M_1 \leq M_2 \leq M_4$ $\qquad$ $M_1 \leq M_3 \leq M_4$

Mirror-stratifiable functions

Definition (Correspondence operator [Daniilidis-Drusvyatskiy-Lewis 13]**)** $R\colon \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ *is a proper lsc convex function. The associated correspondence operator* $\mathcal{J}_R : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ *is*

$$\mathcal{J}_R(S) \overset{\text{def}}{=} \bigcup_{x \in S} \mathrm{ri}(\partial R(x)),$$

Mirror-stratifiable functions

Definition (Correspondence operator [Daniilidis-Drusvyatskiy-Lewis 13]) $R\colon \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ *is a proper lsc convex function. The associated correspondence operator* $\mathcal{J}_R : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ *is*

$$\mathcal{J}_R(S) \stackrel{\text{def}}{=} \bigcup_{x \in S} \mathrm{ri}(\partial R(x)),$$

- $\mathcal{J}_R$ is increasing for the inclusion.
- To be interesting for sensitivity analysis, it should be decreasing for the partial ordering $\leq$.
- This is captured in mirror stratification.

Mirror-stratifiable functions

Definition (Correspondence operator [Daniilidis-Drusvyatskiy-Lewis 13]) $R \colon \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ *is a proper lsc convex function. The associated correspondence operator* $\mathcal{J}_R : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ *is*

$$\mathcal{J}_R(S) \stackrel{\text{def}}{=} \bigcup_{x \in S} \mathrm{ri}(\partial R(x)),$$

Definition (Mirror stratification) R *is mirror-stratifiable with respect to a* **primal** *stratification* $\mathcal{M} = \{M_i\}_{i \in I}$ *of* $\mathrm{dom}(\partial R)$ *and a* **dual** *stratification* $\mathcal{M}^* = \{M_i^*\}_{i \in I}$ *of* $\mathrm{dom}(\partial R^*)$ *if :*

(i) Conjugation induces a duality pairing between $\mathcal{M}$ *and* $\mathcal{M}^*$*, and* $\mathcal{J}_R : \mathcal{M} \to \mathcal{M}^*$ *is invertible with* $\mathcal{J}_R^{-1} = \mathcal{J}_{R^*}$*.*

(ii) $\mathcal{J}_R$ *is decreasing for the relation* $\leq$ *: for any* $(M, M') \in \mathcal{M}^2$

$$M \leq M' \iff \mathcal{J}_R(M) \geq \mathcal{J}_R(M').$$

Mirror-stratifiable functions

Definition **(Correspondence operator [Daniilidis-Drusvyatskiy-Lewis 13])** $R\colon \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ *is a proper lsc convex function. The associated correspondence operator* $\mathcal{J}_R : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ *is*

$$\mathcal{J}_R(S) \stackrel{\text{def}}{=} \bigcup_{x \in S} \mathrm{ri}(\partial R(x)),$$

Definition **(Mirror stratification)** R *is mirror-stratifiable with respect to a* **primal** *stratification* $\mathcal{M} = \{M_i\}_{i \in I}$ *of* $\mathrm{dom}(\partial R)$ *and a* **dual** *stratification* $\mathcal{M}^* = \{M_i^*\}_{i \in I}$ *of* $\mathrm{dom}(\partial R^*)$ *if :*

(i) Conjugation induces a duality pairing between $\mathcal{M}$ *and* $\mathcal{M}^*$*, and* $\mathcal{J}_R : \mathcal{M} \to \mathcal{M}^*$ *is invertible with* $\mathcal{J}_R^{-1} = \mathcal{J}_{R^*}$*.*

(ii) $\mathcal{J}_R$ *is decreasing for the relation* $\leq$ *: for any* $(M, M') \in \mathcal{M}^2$

$$M \leq M' \iff \mathcal{J}_R(M) \geq \mathcal{J}_R(M').$$

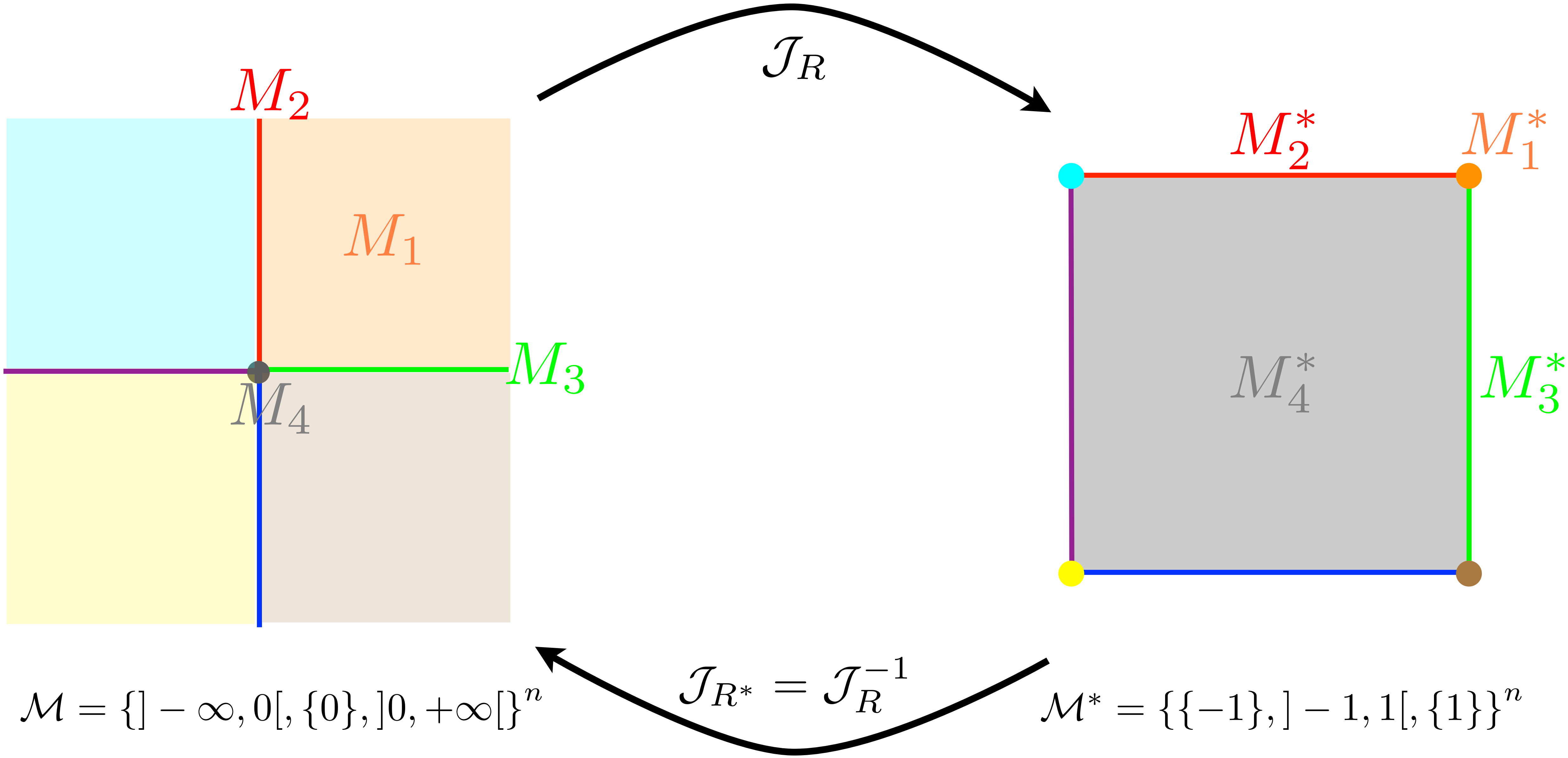
Remark *The theory can be generalized to set-valued maximal monotone mappings just as well.*

Mirror-stratifiable functions

$$R = \|\cdot\|_1 \qquad\qquad R^* = \iota_{\|v\|_\infty \leq 1}$$

$$\mathrm{dom}(\partial R) = \mathbb{R}^n \qquad\qquad \mathrm{dom}(N_{\mathbb{B}_\infty}) = \mathbb{B}_\infty$$

$$\mathcal{J}_{\|\cdot\|_1}(M_i) = \bigcup_{x \in M_i} \mathrm{ri}(\partial \|x\|_1) = M_i^*$$

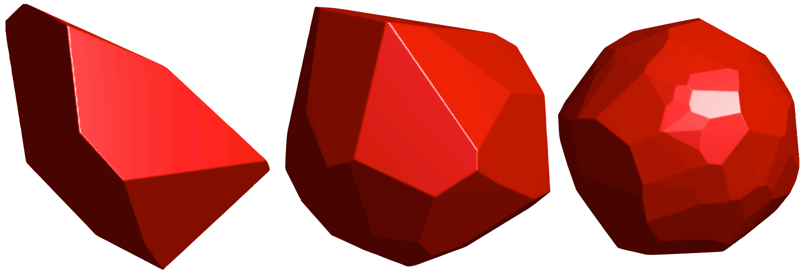


$\mathcal{M} = \{]-\infty, 0[, \{0\},]0, +\infty[\}^n$

$\mathcal{M}^* = \{\{-1\},]-1, 1[, \{1\}\}^n$

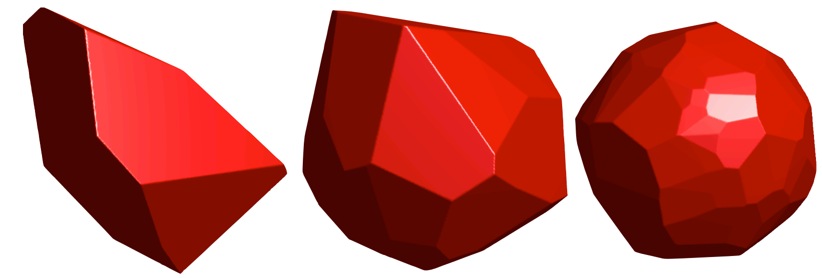
Examples

Example (Polyhedral functions) *A polyhedral function is mirror-stratifiable.*

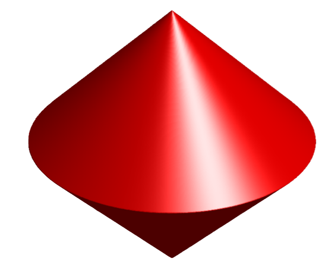


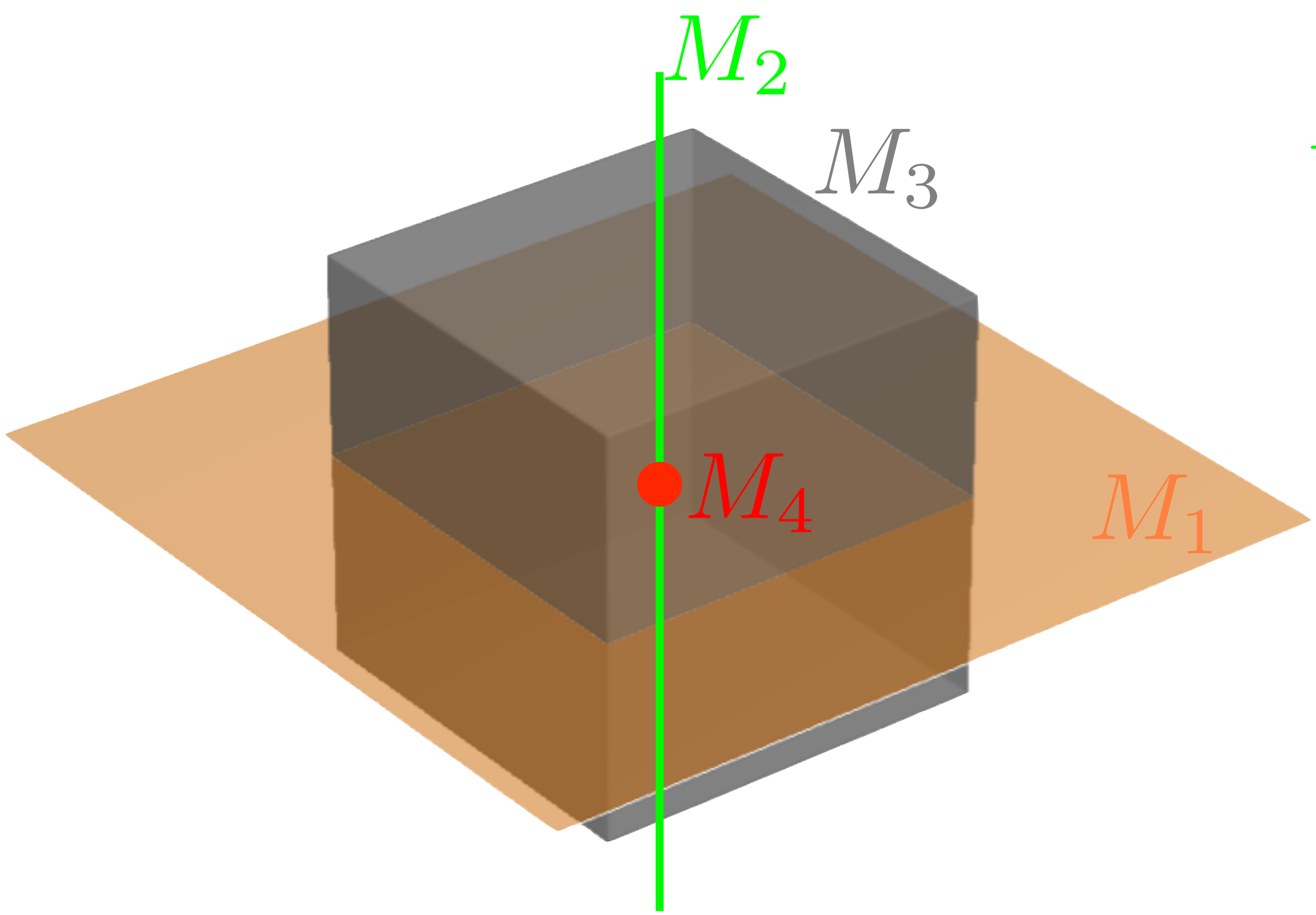
Examples

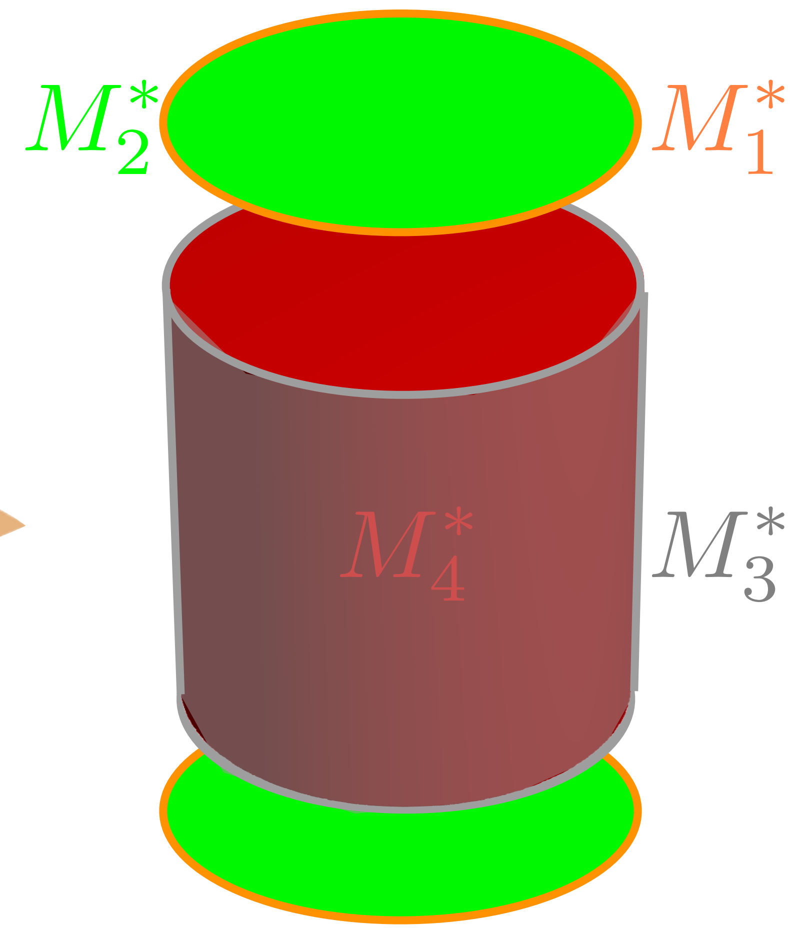
Example (Polyhedral functions) *A polyhedral function is mirror-stratifiable.*



Example ($\ell_{1,2}$ norm) $R(x) = \|x\|_{\mathcal{B}} = \sum_{B\in\mathcal{B}} \|x_B\|_2$, *is mirror-stratifiable with respect to* $\mathcal{M} = \{\{0\}, \mathbb{R}^{|B|} \setminus \{0\}\}^K$ *and* $\mathcal{M}^* = \{\mathbb{S}^{|B|-1}, \mathrm{int}(\mathbb{B}_{\ell_2^{|B|}})\}^K$.

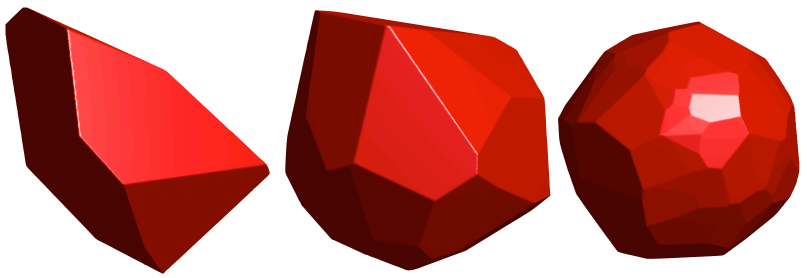




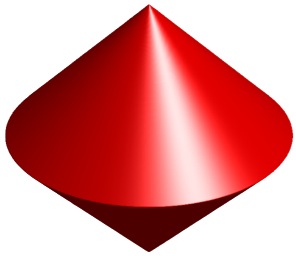


Examples

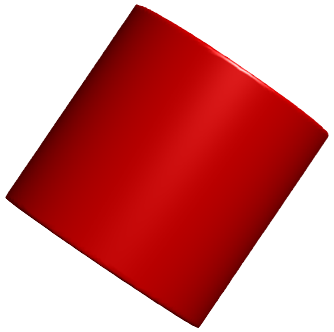
Example (Polyhedral functions) *A polyhedral function is mirror-stratifiable.*



Example ($\ell_{1,2}$ norm) $R(x) = \|x\|_{\mathcal{B}} = \sum_{B \in \mathcal{B}} \|x_B\|_2$, *is mirror-stratifiable with respect to* $\mathcal{M} = \{\{0\}, \mathbb{R}^{|B|} \setminus \{0\}\}^K$ *and* $\mathcal{M}^* = \{\mathbb{S}^{|B|-1}, \mathrm{int}(\mathbb{B}_{\ell_2^{|B|}})\}^K$.



Example (Nuclear norm) $R(x) = \|x\|_* = \sum_{i=1}^r \sigma_i(x)$, $x \in \mathbb{R}^{n_1 \times n_2}$ *is mirror-stratifiable with respect to*



$$
\begin{aligned}
M_i &= \left\{X \in \mathbb{R}^{n \times n} \ : \ \mathrm{rang}(X) = i\right\} \\
M_i^* &= \{U \in \mathbb{R}^n \ : \ \sigma_1(U) = \cdots = \sigma_i(U) = 1, \forall j > i, \ |\sigma_j(U)| < 1\}.
\end{aligned}
$$

Calculus

Proposition (Separability) *If* $R_l : \mathbb{R}^{n_l} \to \mathbb{R} \cup \{+\infty\}$ *is mirror-stratifiable,* $l = , \cdots , L$*, w.r.t.* $\mathcal{M}_l$ *and* $\mathcal{M}_l^*$*, then* $\sum_{l=1}^{L} R_l(x_l)$ *is mirror-stratifiable w.r.t.* $\mathcal{M}_1 \times \cdots \times \mathcal{M}_L$ *and* $\mathcal{M}_1^* \times \cdots \times \mathcal{M}_L^*$*.*

Calculus

Proposition (Separability) *If* $R_l : \mathbb{R}^{n_l} \to \mathbb{R} \cup \{+\infty\}$ *is mirror-stratifiable,* $l = , \cdots, L$*, w.r.t.* $\mathcal{M}_l$ *and* $\mathcal{M}_l^*$*, then* $\sum_{l=1}^{L} R_l(x_l)$ *is mirror-stratifiable w.r.t.* $\mathcal{M}_1 \times \cdots \times \mathcal{M}_L$ *and* $\mathcal{M}_1^* \times \cdots \times \mathcal{M}_L^*$*.*

Proposition (Spectral Lifting of Polyhedral Functions) *Let* R^{sym} *be a convex polyhedral and (signed) permutation-invariant function. Then the spectrally lifted function* $R = R^{\mathrm{sym}} \circ \sigma$ *is mirror-stratifiable w.r.t. the smooth stratification* $\{\sigma^{-1}(M^{\mathrm{sym}})\}$ *and its image by* $\mathcal{J}_R$*.*

Calculus

Proposition (Separability) *If $R_l : \mathbb{R}^{n_l} \to \mathbb{R} \cup \{+\infty\}$ is mirror-stratifiable, $l = , \cdots , L$, w.r.t. $\mathcal{M}_l$ and $\mathcal{M}_l^*$, then $\sum_{l=1}^{L} R_l(x_l)$ is mirror-stratifiable w.r.t. $\mathcal{M}_1 \times \cdots \times \mathcal{M}_L$ and $\mathcal{M}_1^* \times \cdots \times \mathcal{M}_L^*$.*

Proposition (Spectral Lifting of Polyhedral Functions) *Let R^{sym} be a convex polyhedral and (signed) permutation-invariant function. Then the spectrally lifted function $R = R^{\mathrm{sym}} \circ \sigma$ is mirror-stratifiable w.r.t. the smooth stratification $\{\sigma^{-1}(M^{\mathrm{sym}})\}$ and its image by $\mathcal{J}_R$.*

Remark (Instability under the sum) *The family of mirror-stratifiable functions is not stable under the sum :*

- *Consider the pair $R(x) = |x| + x^2/2$ and $R^*(u) = (\max\{|u| - 1, 0\})^2/2$.*
- *$\mathcal{J}_R(\mathbb{R}) = \mathbb{R} \backslash \{-1, 1\} \Rightarrow \mathcal{J}_R$ cannot be a pairing between any two stratifications of $\mathbb{R}$.*
- *The reason : convex conjugacy is a duality between strict/strong convexity and smoothness.*

Ubiquity of mirror-stratification

Ubiquity of mirror-stratification

- Assumptions seem strong, but virtually all regularizers routinely used in signal/image processing, machine learning, statistics are mirror stratifiable.

Ubiquity of mirror-stratification

- Assumptions seem strong, but virtually all regularizers routinely used in signal/image processing, machine learning, statistics are mirror stratifiable.
- Strata are generally not necessarily manifolds.

Ubiquity of mirror-stratification

- Assumptions seem strong, but virtually all regularizers routinely used in signal/image processing, machine learning, statistics are mirror stratifiable.
- Strata are generally not necessarily manifolds.
- Polyhedral, semialgebraic, tame functions, where stratifications are manifolds.

Outline

- Mirror-stratifiable functions.
- Sensitivity analysis.
- Regularized inverse problems.
- Algorithmic implications.
- Numerical results.
- Conclusion.

Parametric composite problems

- Parametric convex composite (smooth+nonsmooth) optimization problem :

$$x^\star(p) \in \operatorname*{Argmin}_{x\in\mathbb{R}^n} E(x,p) \overset{\text{def}}{=} F(x,p) + R(x),$$

  - $p \in \Pi$, Π an open subset of a finite dimensional linear space.
  - $F(\cdot, p)$ is convex and smooth.
  - R is proper lsc and convex.
- Sensitivity analysis is about the properties of $x^\star(p)$ to perturbations of $p \in \Pi$ around some reference point p_0.
- Existing sensitivity results require non-degeneracy :

$$-\nabla F(x^\star(p_0), p_0) \in \mathrm{ri}(\partial R(x^\star(p_0))).$$

Sensitivity result

$$x^\star(p) \in \operatorname*{Argmin}_{x\in\mathbb{R}^n} E(x,p) \stackrel{\text{def}}{=} F(x,p) + R(x),$$

Theorem *Assume that :*

(i) **Well-posedness** *: $E(\cdot, p_0)$ has a unique minimizer $x^\star(p_0)$;*

(ii) **Continuity** *: E is lsc on $\mathbb{R}^n \times \Pi$, $E(x^\star(p_0), \cdot)$ is continuous at p_0, and ∇F is continuous at $(x^\star(p_0), p_0)$;*

(iii) **Compactness** *: E is level-bounded in x uniformly in p locally around p_0.*

If R is mirror-stratifiable w.r.t. $(\mathcal{M},\mathcal{M}^*)$, then $\forall p \sim p_0$, any minimizer $x^\star(p)$ obeys

$$M_{x^\star(p_0)} \leq M_{x^\star(p)} \leq \mathcal{J}_{R^*}(M^*_{u^\star(p_0)}) \qquad u^\star(p_0) \stackrel{\text{def}}{=} -\nabla F(x^\star(p_0), p_0).$$

Sensitivity result

$$x^\star(p) \in \operatorname*{Argmin}_{x \in \mathbb{R}^n} E(x,p) \stackrel{\text{def}}{=} F(x,p) + R(x),$$

Theorem *Assume that :*

(i) **Well-posedness** *: $E(\cdot, p_0)$ has a unique minimizer $x^\star(p_0)$;*

(ii) **Continuity** *: E is lsc on $\mathbb{R}^n \times \Pi$, $E(x^\star(p_0), \cdot)$ is continuous at p_0, and ∇F is continuous at $(x^\star(p_0), p_0)$;*

(iii) **Compactness** *: E is level-bounded in x uniformly in p locally around p_0.*

If R is mirror-stratifiable w.r.t. $(\mathcal{M},\mathcal{M}^*)$, then $\forall p \sim p_0$, any minimizer $x^\star(p)$ obeys

$$M_{x^\star(p_0)} \leq M_{x^\star(p)} \leq \mathcal{J}_{R^*}(M^*_{u^\star(p_0)}) \qquad u^\star(p_0) \stackrel{\text{def}}{=} -\nabla F(x^\star(p_0), p_0).$$

- If R has a nice structure (*partial smoothness*) around $x^\star(p_0)$ and

$$u^\star(p_0) \in \mathrm{ri}(\partial R(x^\star(p_0))),$$

then $\forall p \sim p_0$ [Lewis 2006]

$$M_{x^\star(p)} = M_{x^\star(p_0)} = \mathcal{J}_{R^*}(M^*_{u^\star(p_0)}).$$

Example

$$\min_{x \in \mathbb{B}_\infty} \frac{1}{2} \|x - p\|^2 \quad \text{(P)}$$

Example

$$\min_{x \in \mathbb{B}_\infty} \frac{1}{2} \|x - p\|^2 \quad \text{(P)} \qquad\qquad \min_{u \in \mathbb{R}^n} \frac{1}{2} \|u - p\|^2 + \|u\|_1 \quad \text{(D)}$$

Example

$$\min_{x \in \mathbb{B}_\infty} \frac{1}{2} \|x - p\|^2 \quad \text{(P)} \qquad\qquad \min_{u \in \mathbb{R}^n} \frac{1}{2} \|u - p\|^2 + \|u\|_1 \quad \text{(D)}$$

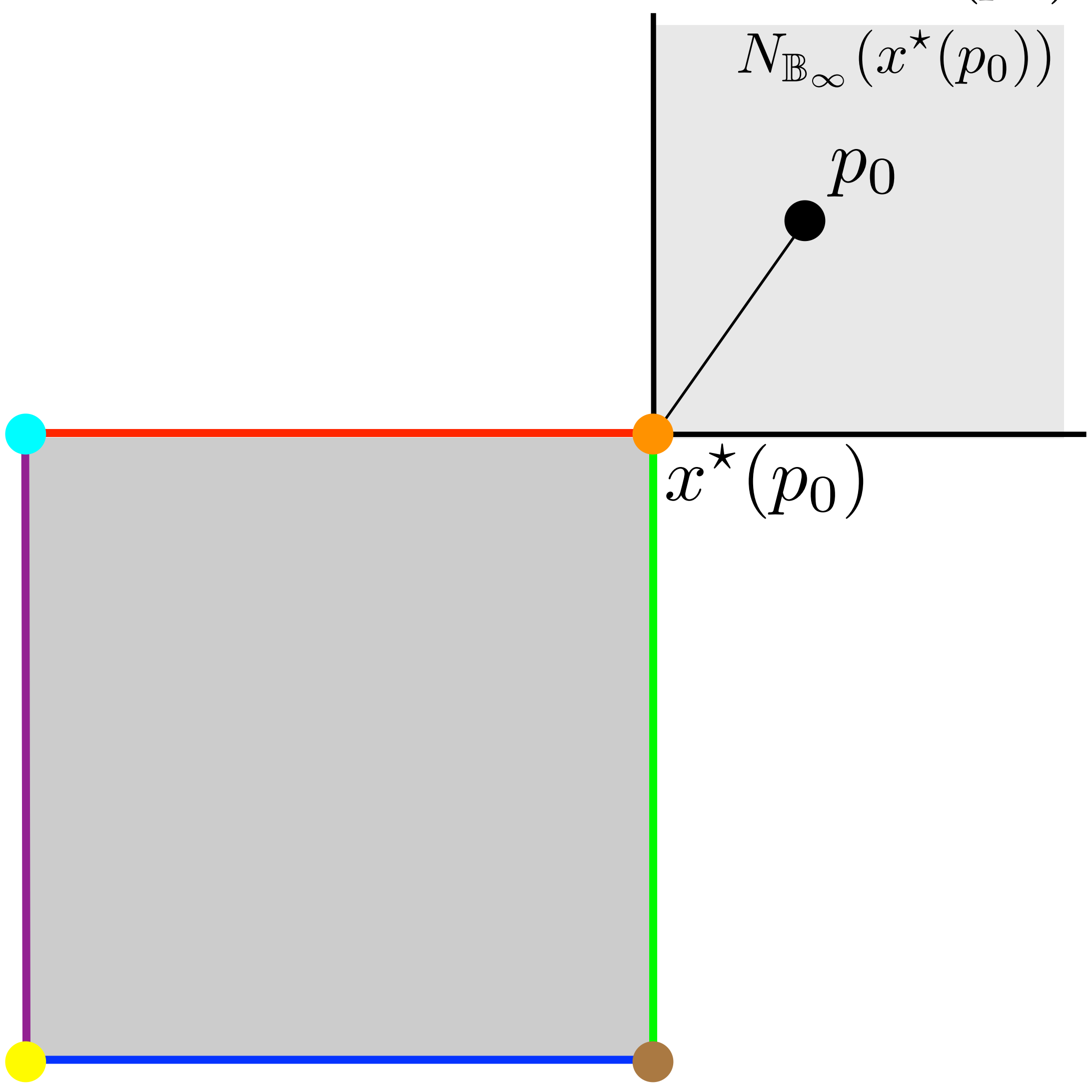
$$\iff \; u^\star(p) = p - x^\star(p) \in N_{\mathbb{B}_\infty}(x^\star(p))$$

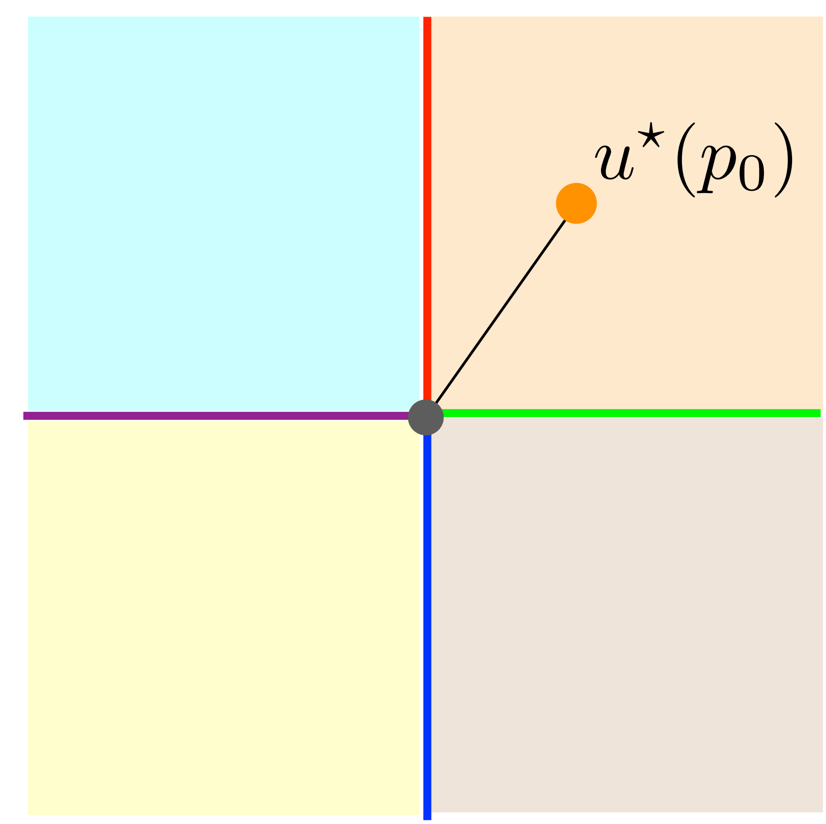
Example

$$\min_{x \in \mathbb{B}_\infty} \frac{1}{2} \|x - p\|^2 \quad \text{(P)} \qquad\qquad \min_{u \in \mathbb{R}^n} \frac{1}{2} \|u - p\|^2 + \|u\|_1 \quad \text{(D)}$$

$\iff \quad u^\star(p) = p - x^\star(p) \in N_{\mathbb{B}_\infty}(x^\star(p))$

$u^\star(p_0) \in \mathrm{ri}(N_{\mathbb{B}_\infty}(x^\star(p_0)))$





$\mathcal{M}^* = \{\{-1\},]-1, 1[, \{1\}\}^n$

$\mathcal{M} = \{]-\infty, 0[, \{0\},]0, +\infty[\}^n$

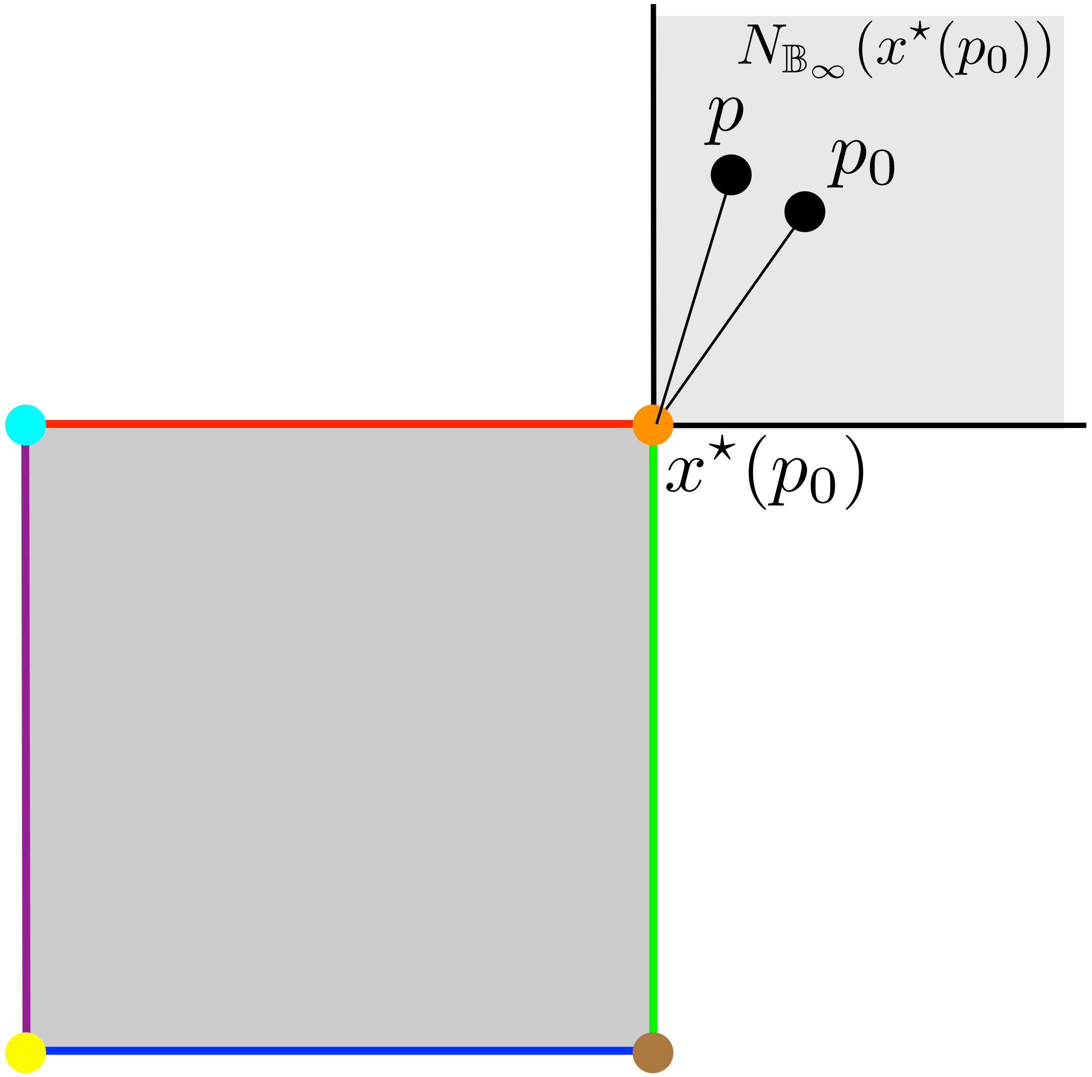
Example

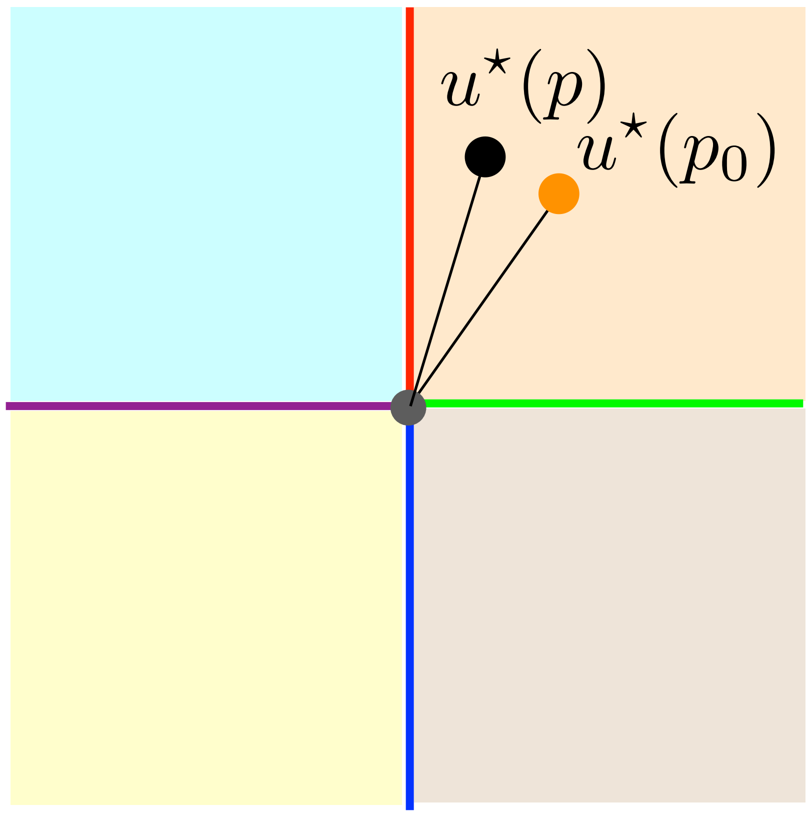
$$\min_{x\in\mathbb{B}_\infty} \frac{1}{2}\|x-p\|^2 \quad \text{(P)}$$

$$\min_{u\in\mathbb{R}^n} \frac{1}{2}\|u-p\|^2 + \|u\|_1 \quad \text{(D)}$$

$\iff \; u^\star(p) = p - x^\star(p) \in N_{\mathbb{B}_\infty}(x^\star(p))$

$u^\star(p_0) \in \mathrm{ri}(N_{\mathbb{B}_\infty}(x^\star(p_0)))$





$\mathcal{M}^* = \{\{-1\},]-1,1[, \{1\}\}^n$

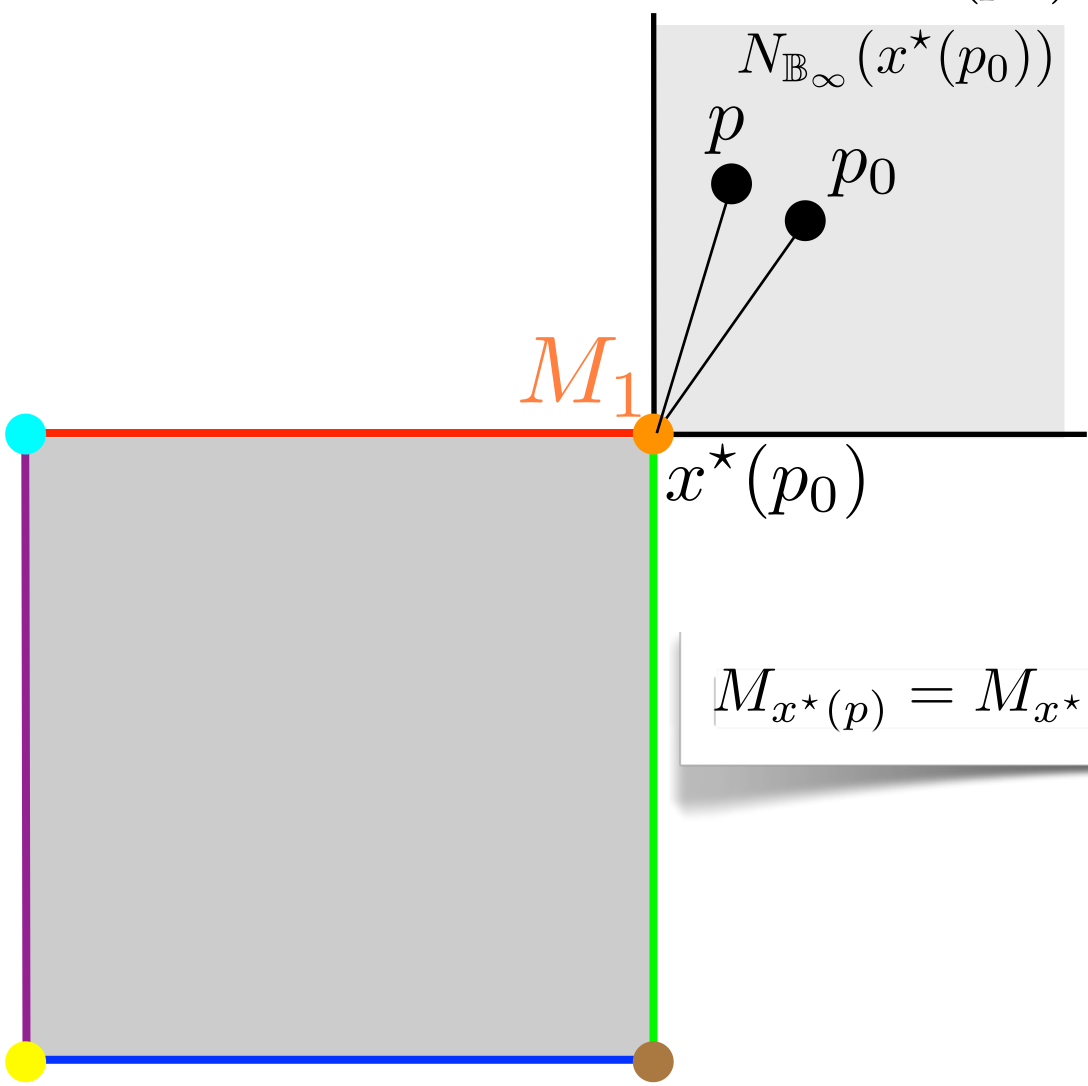
$\mathcal{M} = \{]-\infty,0[, \{0\},]0,+\infty[\}^n$

Example

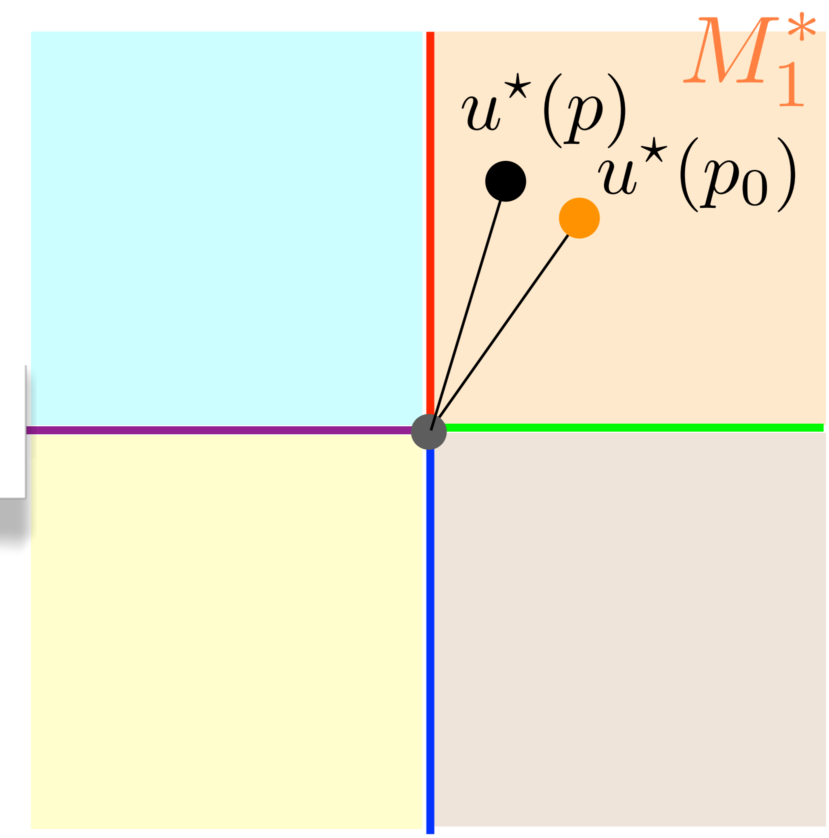
$$\min_{x\in\mathbb{B}_\infty} \frac{1}{2}\|x-p\|^2 \quad \text{(P)} \qquad\qquad \min_{u\in\mathbb{R}^n} \frac{1}{2}\|u-p\|^2 + \|u\|_1 \quad \text{(D)}$$

$$\iff \; u^\star(p) = p - x^\star(p) \in N_{\mathbb{B}_\infty}(x^\star(p))$$

$$u^\star(p_0) \in \mathrm{ri}(N_{\mathbb{B}_\infty}(x^\star(p_0)))$$



$$M_{x^\star(p)} = M_{x^\star(p_0)} = M_1 = \mathcal{J}_{\|\cdot\|_1}(M_1^*).$$



$$\mathcal{M}^* = \{\{-1\},]-1,1[, \{1\}\}^n \qquad\qquad \mathcal{M} = \{]-\infty,0[, \{0\},]0,+\infty[\}^n$$

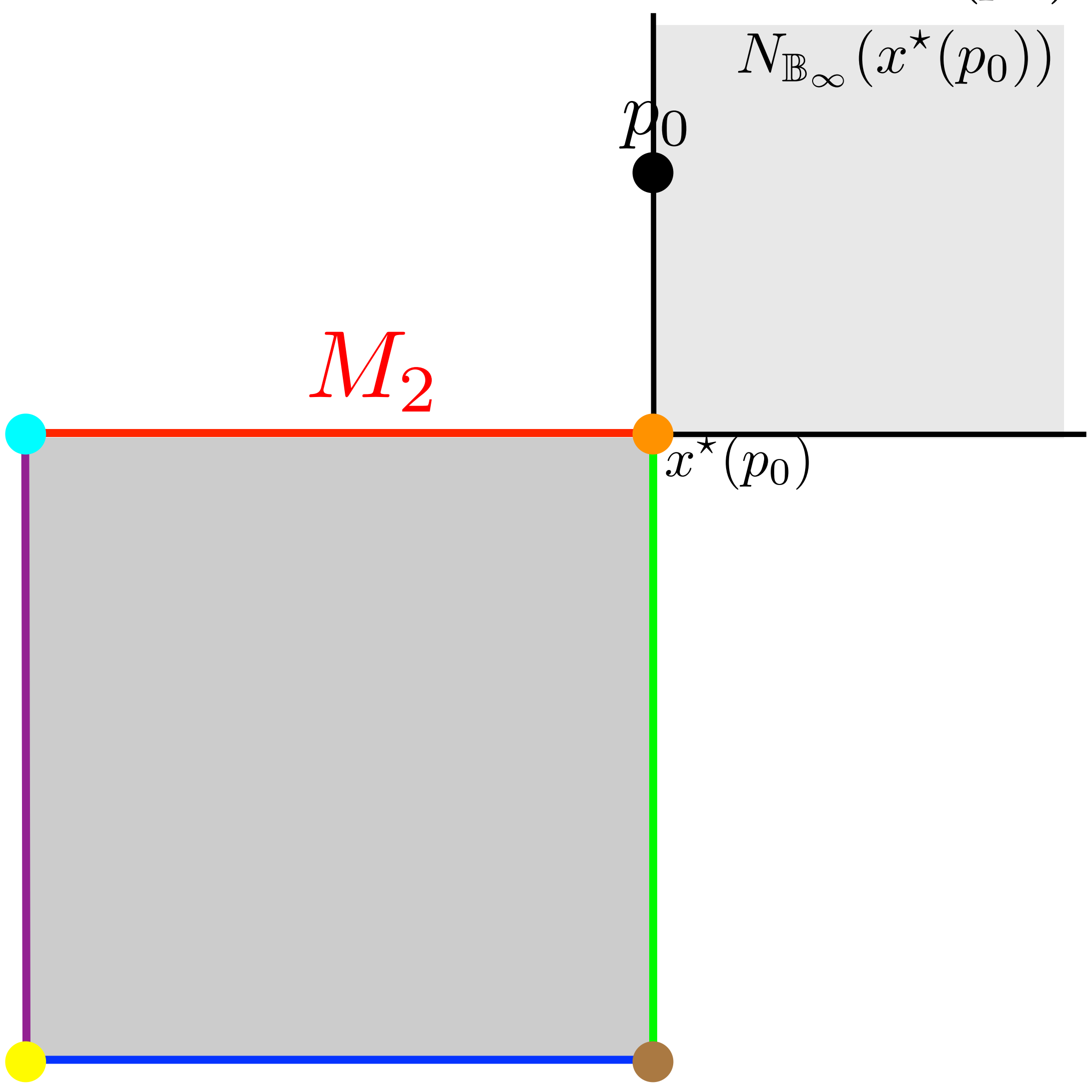
Example

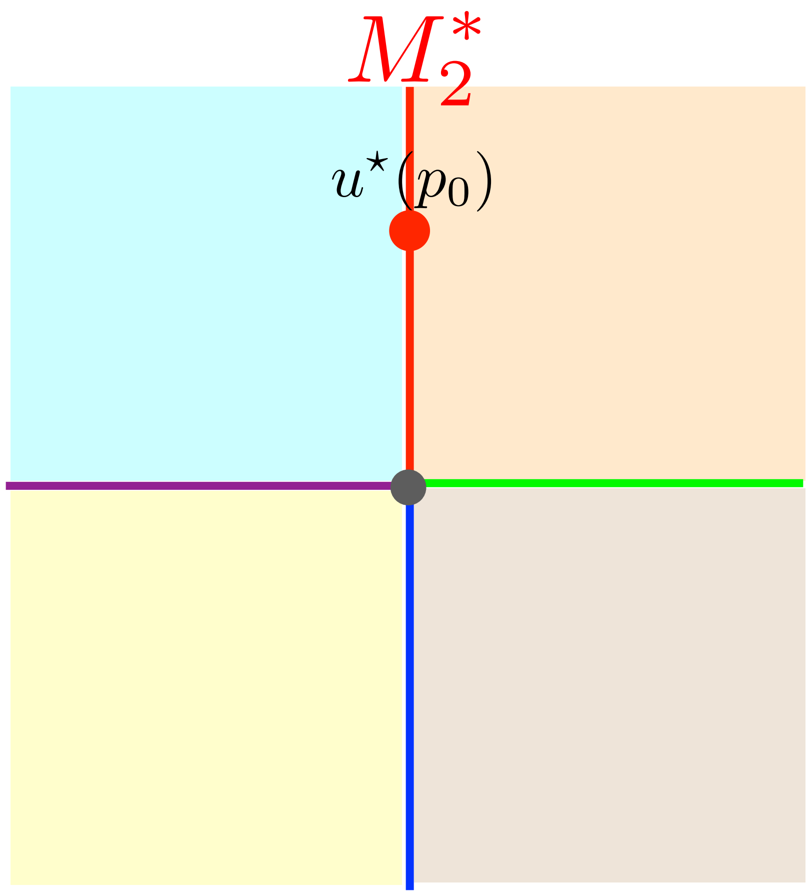
$$\min_{x \in \mathbb{B}_\infty} \frac{1}{2} \|x - p\|^2 \quad \text{(P)}$$

$$\min_{u \in \mathbb{R}^n} \frac{1}{2} \|u - p\|^2 + \|u\|_1 \quad \text{(D)}$$

$$\iff \quad u^\star(p) = p - x^\star(p) \in N_{\mathbb{B}_\infty}(x^\star(p))$$

$$u^\star(p_0) \in \mathrm{rbd}(N_{\mathbb{B}_\infty}(x^\star(p_0)))$$





$$\mathcal{M}^* = \{\{-1\},]-1, 1[, \{1\}\}^n$$

$$\mathcal{M} = \{]-\infty, 0[, \{0\},]0, +\infty[\}^n$$

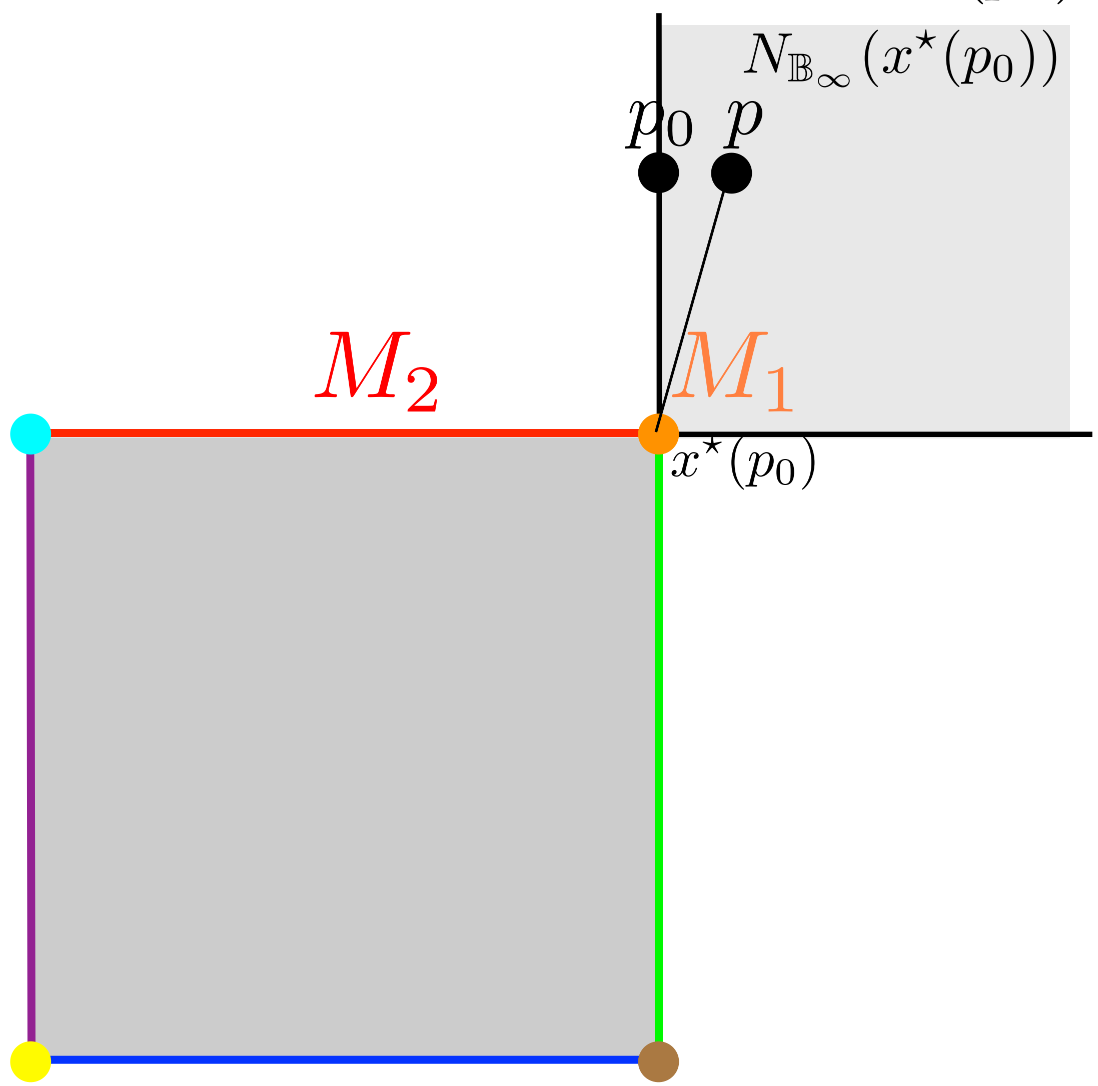
Example

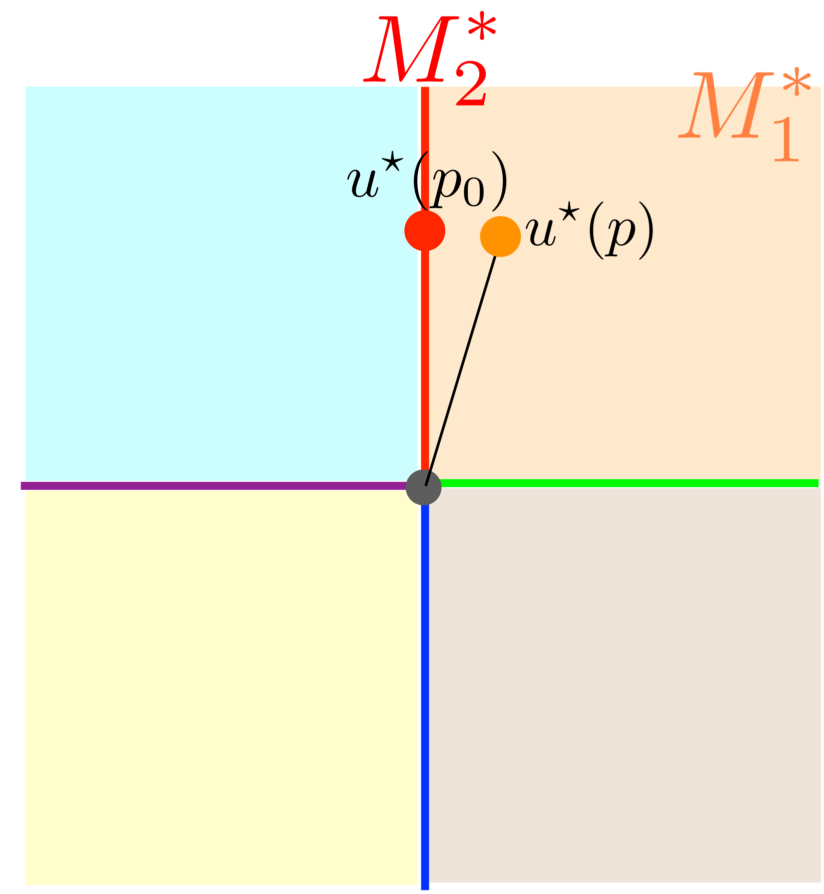
$$\min_{x \in \mathbb{B}_\infty} \frac{1}{2}\|x-p\|^2 \quad \text{(P)}$$

$$\min_{u \in \mathbb{R}^n} \frac{1}{2}\|u-p\|^2 + \|u\|_1 \quad \text{(D)}$$

$$\iff \; u^\star(p) = p - x^\star(p) \in N_{\mathbb{B}_\infty}(x^\star(p))$$

$$u^\star(p_0) \in \mathrm{rbd}(N_{\mathbb{B}_\infty}(x^\star(p_0)))$$





$$\mathcal{M}^* = \{\{-1\},]-1,1[, \{1\}\}^n$$

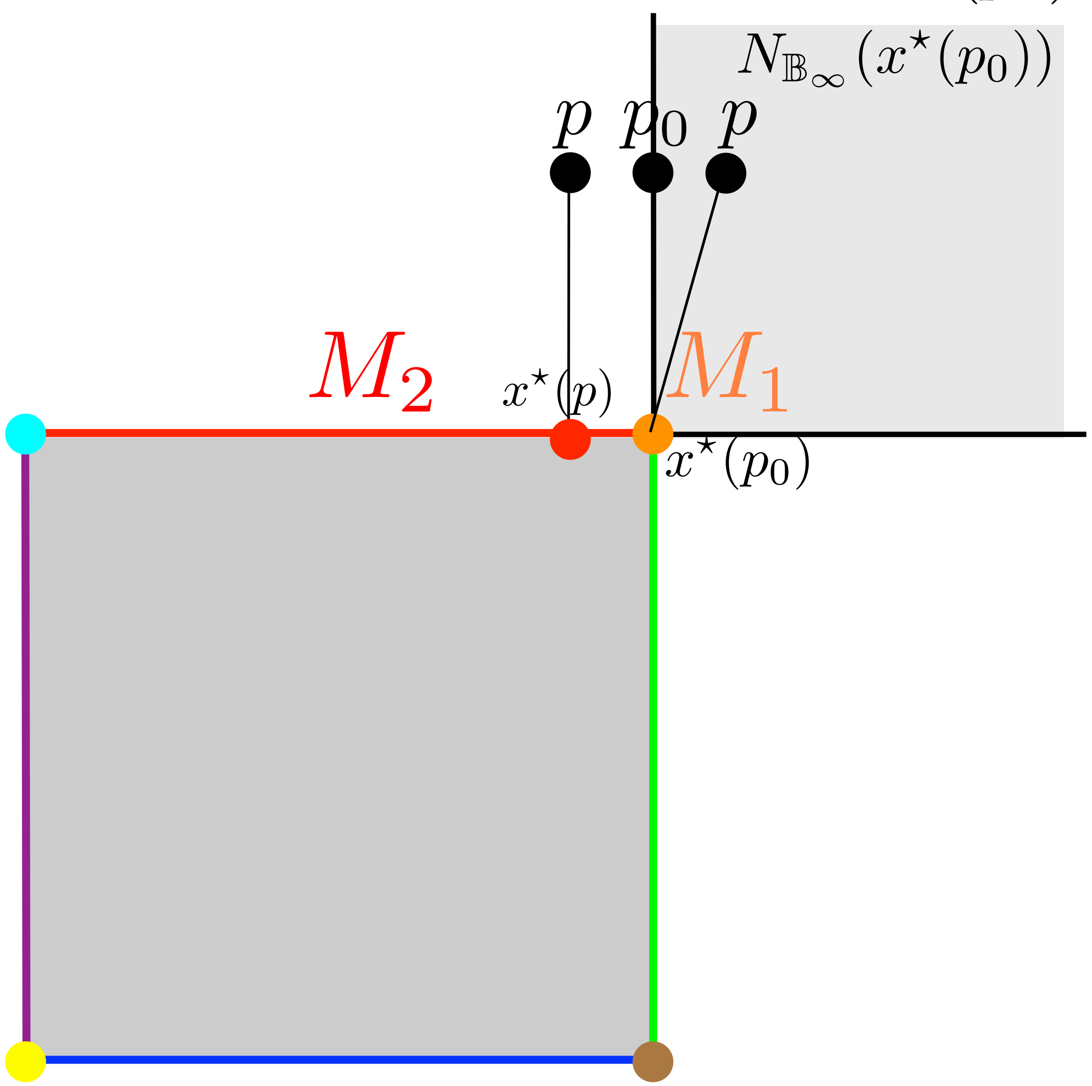
$$\mathcal{M} = \{]-\infty,0[, \{0\},]0,+\infty[\}^n$$

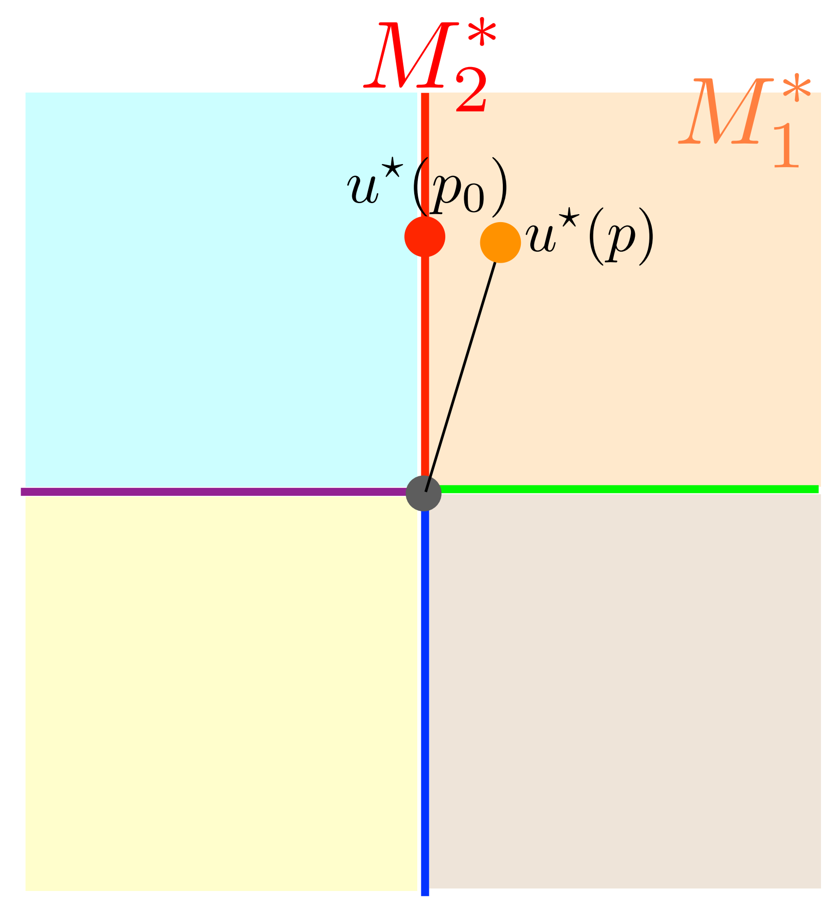
Example

$$\min_{x\in\mathbb{B}_\infty} \frac{1}{2}\|x-p\|^2 \quad \text{(P)} \qquad\qquad \min_{u\in\mathbb{R}^n} \frac{1}{2}\|u-p\|^2 + \|u\|_1 \quad \text{(D)}$$

$$\Longleftrightarrow \quad u^\star(p) = p - x^\star(p) \in N_{\mathbb{B}_\infty}(x^\star(p))$$

$$u^\star(p_0) \in \mathrm{rbd}(N_{\mathbb{B}_\infty}(x^\star(p_0)))$$





$\mathcal{M}^* = \{\{-1\},]-1,1[, \{1\}\}^n$

$\mathcal{M} = \{]-\infty,0[, \{0\},]0,+\infty[\}^n$

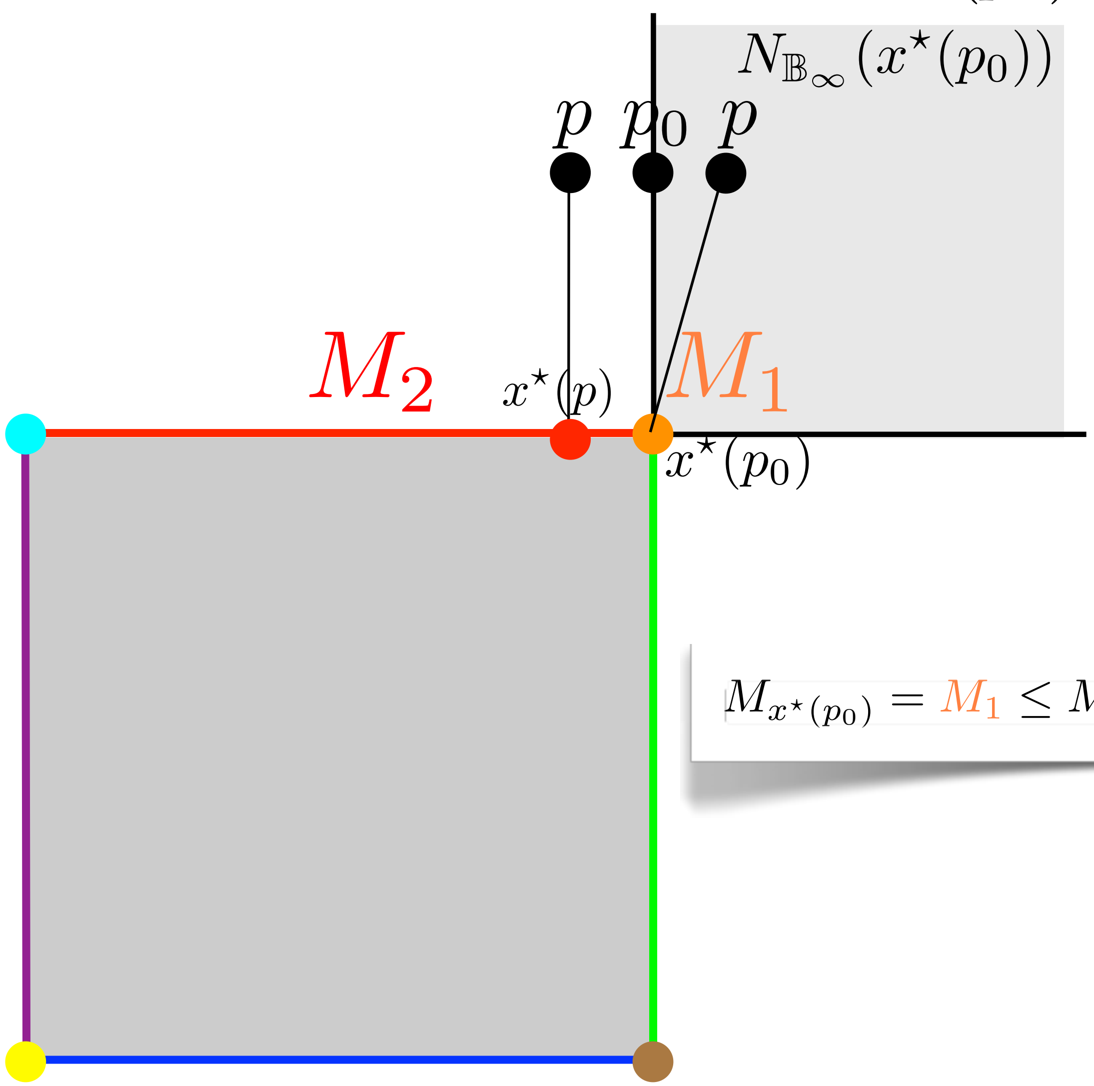
Example

$$\min_{x \in \mathbb{B}_\infty} \frac{1}{2} \|x - p\|^2 \quad \text{(P)}$$

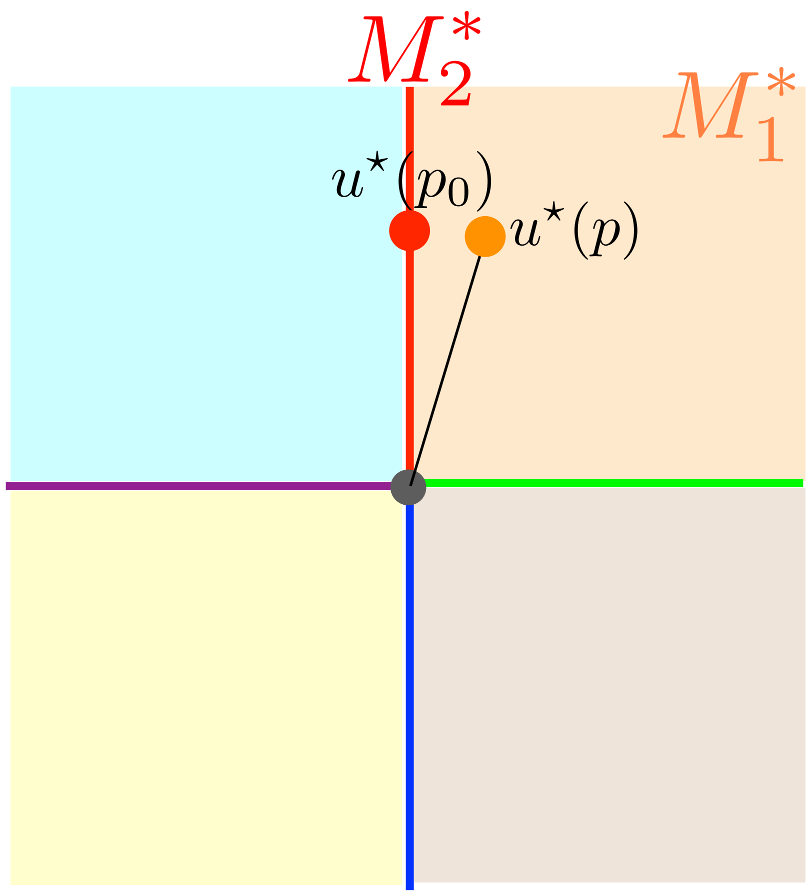
$$\min_{u \in \mathbb{R}^n} \frac{1}{2} \|u - p\|^2 + \|u\|_1 \quad \text{(D)}$$

$$\iff u^\star(p) = p - x^\star(p) \in N_{\mathbb{B}_\infty}(x^\star(p))$$

$$u^\star(p_0) \in \mathrm{rbd}(N_{\mathbb{B}_\infty}(x^\star(p_0)))$$



$$M_{x^\star(p_0)} = M_1 \leq M_{x^\star(p)} \leq \mathcal{J}_{\|\cdot\|_1}(M_{u^\star(p_0)}) = M_2.$$



$$\mathcal{M}^* = \{\{-1\},]-1, 1[, \{1\}\}^n$$

$$\mathcal{M} = \{]-\infty, 0[, \{0\},]0, +\infty[\}^n$$

Outline

- Mirror-stratifiable functions.
- Sensitivity analysis.
- Regularized inverse problems.
- Algorithmic implications.
- Numerical results.
- Conclusion.

Back to inverse problems

$$y = \mathrm{A}x_0 + \varepsilon$$

$$(\mathsf{P}(y,\lambda)) \qquad x^\star(y,\lambda) \in \underset{x\in\mathbb{R}^n}{\mathrm{Argmin}}\ \frac{1}{2}\|y - \mathrm{A}x\|^2 + \lambda R(x)$$

$$(\mathsf{P}(y,0)) \qquad x^\star(y,0) \in \underset{x\in\mathbb{R}^n}{\mathrm{Argmin}}\ R(x) \quad \text{s.t.} \quad y = \mathrm{A}x$$

- Sensitivity analysis is about the properties of $x^\star(p)$, to perturbations of $p = (y,\lambda)$ around $p_0 = (y_0, 0)$.
- Previous sensitivity result for parametric composite problems does not apply : $E(x,p)$ may not even be continuous at p_0.

Known exact model recovery

$$y = \mathrm{A}x_0 + \varepsilon$$

$$(\mathsf{P}(y,\lambda)) \qquad x^\star(y,\lambda) \in \operatorname*{Argmin}_{x\in\mathbb{R}^n} \frac{1}{2}\|y - \mathrm{A}x\|^2 + \lambda R(x)$$

Minimal ℓ_2-norm certificate $q_0 = \mathrm{A}^{+,*}_{T_{x_0}} e_{x_0}$.

$$T_{x_0} = \mathrm{par}(\partial R(x_0))^\perp$$
$$e_{x_0} = \mathrm{P}_{\mathrm{Aff}(\partial R(x_0))}(0)$$

Known exact model recovery

$$y = \mathrm{A}x_0 + \varepsilon$$

$$(\mathsf{P}(y,\lambda)) \qquad x^\star(y,\lambda) \in \operatorname*{Argmin}_{x\in\mathbb{R}^n} \frac{1}{2}\|y - \mathrm{A}x\|^2 + \lambda R(x)$$

Minimal ℓ_2-norm certificate $q_0 = \mathrm{A}_{T_{x_0}}^{+,*} e_{x_0}$.

$T_{x_0} = \mathrm{par}(\partial R(x_0))^\perp$
$e_{x_0} = \mathrm{P}_{\mathrm{Aff}(\partial R(x_0))}(0)$

Theorem *Let M_{x_0} be the "model" subset of x_0. Assume that R has a nice structure on a nice subset M_{x_0} nearby x_0, and that*

Non-degeneracy condition

$$u_0 = \mathrm{A}^* q_0 \in \mathrm{ri}\,(\partial R(x_0)) \quad \textit{and} \quad \ker(\mathrm{A}) \cap T_{x_0} = \{0\}.$$

Then $\exists C_0, C_1 > 0$ depending only on x_0 such that for all (ε,λ) s.t. $C_0\|\varepsilon\| \le \lambda \le C_1$, $(\mathsf{P}(y_0,0))$ has a unique solution $x^\star(y,\lambda)$ and satisfies

$$M_{x^\star(y,\lambda)} = M_{x_0}.$$

Enlarged model recovery

$$y = \mathrm{A}x_0 + \varepsilon$$

$$(\mathsf{P}(y,\lambda)) \qquad x^\star(y,\lambda) \in \underset{x\in\mathbb{R}^n}{\mathrm{Argmin}}\ \frac{1}{2}\left\|y - \mathrm{A}x\right\|^2 + \lambda R(x)$$

$$q_0(0,y) = \underset{q\in\mathbb{R}^m}{\mathrm{argmin}}\ \|q\| \quad \text{s.t.} \quad \mathrm{A}^* q \in \partial R(x^\star(0,y)).$$

Theorem *Suppose that x_0 is the unique solution to $(\mathsf{P}(y_0,0))$. Assume that R is mirror-stratifiable w.r.t. $(\mathcal{M},\mathcal{M}^*)$. Then $\exists C_0, C_1 > 0$ depending only on x_0 such that for all (ε,λ) s.t. $C_0\|\varepsilon\| \leq \lambda \leq C_1$, there exists a minimizer $x^\star(y,\lambda)$ of $(\mathsf{P}(y,\lambda))$ localized as*

$$M_{x_0} \leq M_{x^\star(p)} \leq \mathcal{J}_{R^*}(M^*_{u_0}) \qquad u_0 \overset{\text{def}}{=} \mathrm{A}^* q_0(0,y_0).$$

Outline

- Mirror-stratifiable functions.
- Sensitivity analysis.
- Regularized inverse problems.
- Algorithmic implications.
- Numerical results.
- Conclusion.

FB splitting: class of problems

$$\min_{x \in \mathbb{R}^n} F(x) + R(x)$$

(**A**.1) F and $R \in \Gamma_0(\mathbb{R}^n)$, $F \in C^{1,1}(\mathbb{R}^n)$ with $1/\beta$–Lipschitz gradient.

(**A**.2) Non-empty set of minimizers.

FB splitting: class of problems

$$\min_{x \in \mathbb{R}^n} F(x) + R(x)$$

(**A**.1) F and $R \in \Gamma_0(\mathbb{R}^n)$, $F \in C^{1,1}(\mathbb{R}^n)$ with $1/\beta$–Lipschitz gradient.

(**A**.2) Non-empty set of minimizers.

$\gamma_k \in [\epsilon, 2\beta - \epsilon]$ $$x_{k+1} = \mathrm{prox}_{\gamma_k R}\left(x_k - \gamma_k \nabla F(x_k)\right)$$

Exact FB activity identification

$$\min_{x \in \mathbb{R}^n} F(x) + R(x)$$

$\gamma_k \in [\epsilon, 2\beta - \epsilon]$

$$x_{k+1} = \operatorname{prox}_{\gamma_k R}(x_k - \gamma_k \nabla F(x_k))$$

Theorem *Let the FB be used to create a sequence* x_k *which converges to* $x^\star$*. Assume that* R *has a nice structure on a nice subset* $M_{x^\star}$ *nearby* $x^\star$*, and*

$$u^\star \stackrel{\textit{def}}{=} -\nabla F(x^\star) \in \mathrm{ri}\ (\partial R(x^\star)).$$

Then for k *large enough*

$$M_{x_k} = M_{x^\star}.$$

Exact FB activity identification

$$\min_{x \in \mathbb{R}^n} F(x) + R(x)$$

$\gamma_k \in [\epsilon, 2\beta - \epsilon]$ $$x_{k+1} = \operatorname{prox}_{\gamma_k R}\left(x_k - \gamma_k \nabla F(x_k)\right)$$

Theorem *Let the FB be used to create a sequence x_k which converges to $x^\star$. Assume that R has a nice structure on a nice subset $M_{x^\star}$ nearby $x^\star$, and*

Non-degeneracy condition

$$u^\star \overset{\text{def}}{=} -\nabla F(x^\star) \in \operatorname{ri}\left(\partial R(x^\star)\right).$$

Then for k large enough

$$M_{x_k} = M_{x^\star}.$$

Enlarged FB activity identification

$$\min_{x \in \mathbb{R}^n} F(x) + R(x)$$

$\gamma_k \in [\epsilon, 2\beta - \epsilon]$

$$x_{k+1} = \operatorname{prox}_{\gamma_k R}\left(x_k - \gamma_k \nabla F(x_k)\right)$$

Theorem *Let the FB be used to create a sequence x_k which converges to $x^\star$. Assume that R is mirror-stratifiable w.r.t. $(\mathcal{M}, \mathcal{M}^*)$. Then for k large enough,*

$$M_{x^\star} \leq M_{x_k} \leq \mathcal{J}_{R^*}(M^*_{u^\star}) \qquad u^\star \overset{\text{def}}{=} -\nabla F(x^\star).$$

PD splitting: class of problems

$$\min_{x \in \mathbb{R}^n} R(x) + F(x) + (J \overset{+}{\triangledown} G)(Lx)$$

Primal problem

$(J \overset{+}{\triangledown} G)(\cdot) \overset{\text{def}}{=} \inf_{v \in \mathbb{R}^m} J(\cdot) + G(\cdot - v)$

PD splitting: class of problems

$$\min_{x\in\mathbb{R}^n} R(x) + F(x) + (J \overset{+}{\triangledown} G)(Lx)$$

Primal problem

$(J \overset{+}{\triangledown} G)(\cdot) \stackrel{\text{def}}{=} \inf_{v\in\mathbb{R}^m} J(\cdot) + G(\cdot - v)$

(A.1) $R, F \in \Gamma_0(\mathbb{R}^n)$, and $\nabla F \in (1/\beta_F)$–$\mathrm{Lip}(\mathbb{R}^n)$.

(A.2) $J, G \in \Gamma_0(\mathbb{R}^m)$, and G is β_G-strongly convex.

(A.3) $L : \mathbb{R}^n \to \mathbb{R}^m$ is a linear mapping.

(A.4) $0 \in \mathrm{ran}\,(\partial R + \nabla F + L^*(\partial J \Box \partial G)L)$.

$\partial J \Box \partial G \stackrel{\text{def}}{=} (\partial J^{-1} + \partial G^{-1})^{-1}$.

PD splitting: class of problems

$$\min_{x \in \mathbb{R}^n} R(x) + F(x) + (J \overset{+}{\triangledown} G)(Lx)$$

Primal problem

$(J \overset{+}{\triangledown} G)(\cdot) \overset{\text{def}}{=} \inf_{v \in \mathbb{R}^m} J(\cdot) + G(\cdot - v)$

(A.1) $R, F \in \Gamma_0(\mathbb{R}^n)$, and $\nabla F \in (1/\beta_F)\text{–Lip}(\mathbb{R}^n)$.

(A.2) $J, G \in \Gamma_0(\mathbb{R}^m)$, and G is β_G-strongly convex.

(A.3) $L : \mathbb{R}^n \to \mathbb{R}^m$ is a linear mapping.

(A.4) $0 \in \operatorname{ran}(\partial R + \nabla F + L^*(\partial J \square \partial G)L)$.

$\partial J \square \partial G \overset{\text{def}}{=} (\partial J^{-1} + \partial G^{-1})^{-1}$.

$$\max_{v \in \mathbb{R}^m} -J^*(v) - G^*(v) - (R^* \overset{+}{\triangledown} F^*)(-L^* v)$$

Dual problem

PD splitting: class of problems

$$\min_{x\in\mathbb{R}^n} R(x)+F(x)+(J \overset{+}{\triangledown} G)(Lx)$$

Primal problem

$(J \overset{+}{\triangledown} G)(\cdot) \overset{\text{def}}{=} \inf_{v\in\mathbb{R}^m} J(\cdot)+G(\cdot - v)$

(A.1) $R, F \in \Gamma_0(\mathbb{R}^n)$, and $\nabla F \in (1/\beta_F)$–$\mathrm{Lip}(\mathbb{R}^n)$.

(A.2) $J, G \in \Gamma_0(\mathbb{R}^m)$, and G is β_G-strongly convex.

(A.3) $L : \mathbb{R}^n \to \mathbb{R}^m$ is a linear mapping.

(A.4) $0 \in \mathrm{ran}\,(\partial R + \nabla F + L^*(\partial J \Box \partial G)L)$.

$\partial J \Box \partial G \overset{\text{def}}{=} (\partial J^{-1} + \partial G^{-1})^{-1}$.

$$\max_{v\in\mathbb{R}^m} -J^*(v) - G^*(v) - (R^* \overset{+}{\triangledown} F^*)(-L^*v)$$

Dual problem

Classical Kuhn-Tucker theory : a pair $(x^\star, v^\star) \in \mathbb{R}^n \times \mathbb{R}^m$ solves both problems if

$$0 \in \begin{bmatrix} \partial R & L^* \\ -L & \partial J^* \end{bmatrix} \begin{pmatrix} x^\star \\ v^\star \end{pmatrix} + \begin{bmatrix} \nabla F & 0 \\ 0 & \nabla G^* \end{bmatrix} \begin{pmatrix} x^\star \\ v^\star \end{pmatrix},$$

Primal-dual splitting

Algorithm 1 A Primal–Dual splitting method

Initial : Choose $\gamma_R, \gamma_J > 0$ and $\theta \in [-1, +\infty[$. For $k = 0$, $x_0 \in \mathbb{R}^n$, $v_0 \in \mathbb{R}^m$ **repeat**

$$x_{k+1} = \mathrm{prox}_{\gamma_R R}\left(x_k - \gamma_R \nabla F(x_k) - \gamma_R L^* v_k\right), \quad \text{(primal GD)}$$

$$\bar{x}_{k+1} = x_{k+1} + \theta(x_{k+1} - x_k), \quad \text{(Extrapolation)}$$

$$v_{k+1} = \mathrm{prox}_{\gamma_J J^*}\left(v_k - \gamma_J \nabla G^*(v_k) + \gamma_J L\bar{x}_{k+1}\right), \quad \text{(dual GA)}$$

$k = k + 1$

until *convergence*;

- Can be view as forward-backward splitting after renorming [Vu 2011].
- Covers many existing algorithms.
- Other schemes are possible: [Briceños-Arias and Combettes, Combettes et al. 2012].

Enlarged PD activity identification

$$\min_{x\in\mathbb{R}^n} R(x) + F(x) + (J \stackrel{+}{\vee} G)(Lx)$$

$(J \stackrel{+}{\vee} G)(\cdot) \stackrel{\text{def}}{=} \inf_{v\in\mathbb{R}^m} J(\cdot) + G(\cdot - v)$

Theorem *Consider the PD Algorithm 1 with $\theta = 1$ and γ_R, γ_J such that*

$$2\min(\beta_F, \beta_G)\ \min\left(\frac{1}{\gamma_J}, \frac{1}{\gamma_R}\right)\left(1 - \sqrt{\gamma_J\gamma_R\|L\|^2}\right) > 1.$$

Then $(x_k, v_k) \to (x^\star, v^\star)$, a Kuhn-Tucker pair. If moreover R (resp. J^*) is mirror-stratifiable wrt primal and dual stratifications $\{(M_i^R, M_i^{R^*})\}_i$ (resp. $\{(M_i^{J^*}, M_i^J)\}_i$), then for all k large enough, we have

$$M_{x^\star}^R \le M_{x_k}^R \le \mathcal{J}_{R^*}\left(M^{R^*}_{-Lv^\star - \nabla F(x^\star)}\right),$$
$$M_{v^\star}^{J^*} \le M_{v_k}^{J^*} \le \mathcal{J}_J\left(M^J_{Lx^\star - \nabla G^*(v^\star)}\right).$$

$\mathcal{M} \le \mathcal{M}' \iff \mathcal{M} \subset \mathrm{cl}(\mathcal{M}')$

Outline

- Mirror-stratifiable functions.
- Sensitivity analysis.
- Regularized inverse problems.
- Algorithmic implications.
- Numerical results.
- Conclusion.

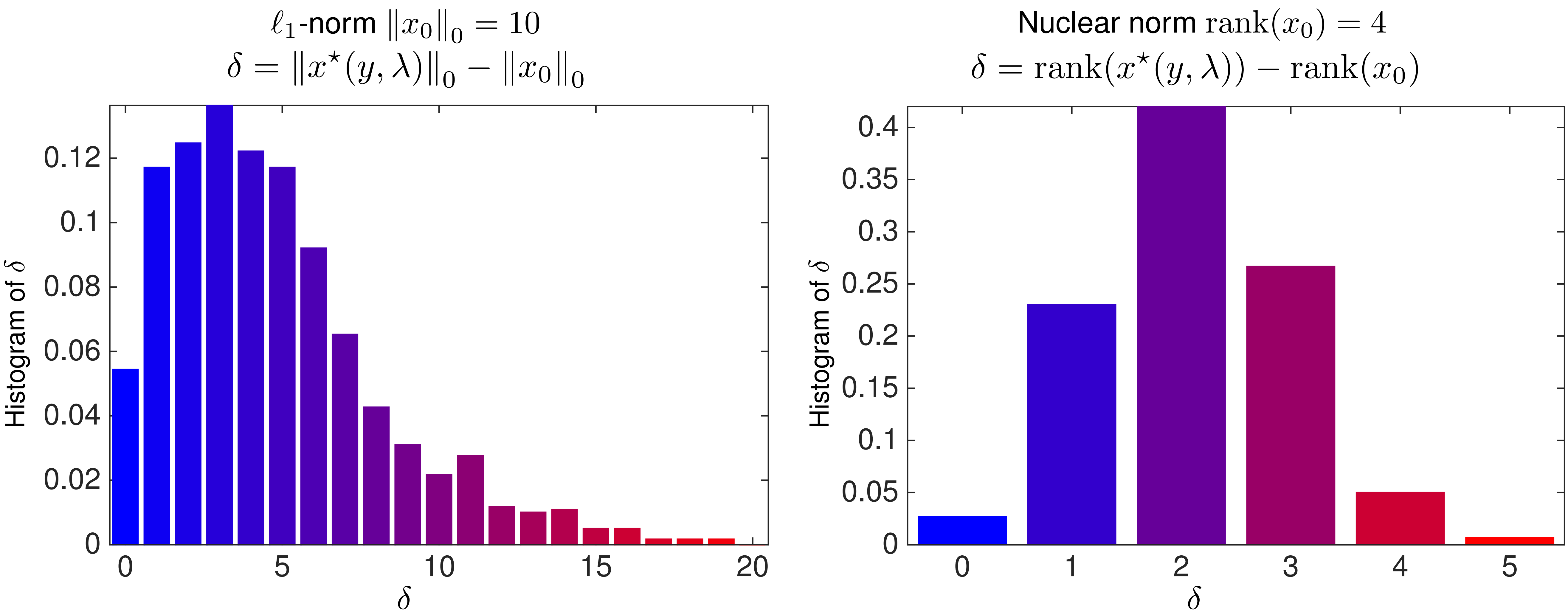
CS scenario

- 1000 realizations at random of $(x_0, \mathrm{A}, \varepsilon)$: A Gaussian ensemble, $x_0 \in \{0,1\}^n$.
- $R = \|\cdot\|_1$: $(n, m) = (100, 50)$, $\|x_0\|_0 \in \{1, \cdots, 30\}$.
- $R = \|\cdot\|_*$: $(20 \times 20, 300) = (100, 50)$, $\mathrm{rank}(x_0) \in \{1, \cdots, 9\}$.
- Complexity of x_0 too large for CS theory to apply.
- Solve :

$$x^\star(y, \lambda) \in \operatorname*{Argmin}_{x \in \mathbb{R}^n} \frac{1}{2} \|y - \mathrm{A}x\|^2 + \lambda R(x)$$

$$u_0 = \mathrm{A}^* \operatorname*{argmin}_{q \in \mathbb{R}^m} \|q\| \quad \text{s.t.} \quad \mathrm{A}^* q \in \partial R(x_0).$$

Stable exact model recovery ?



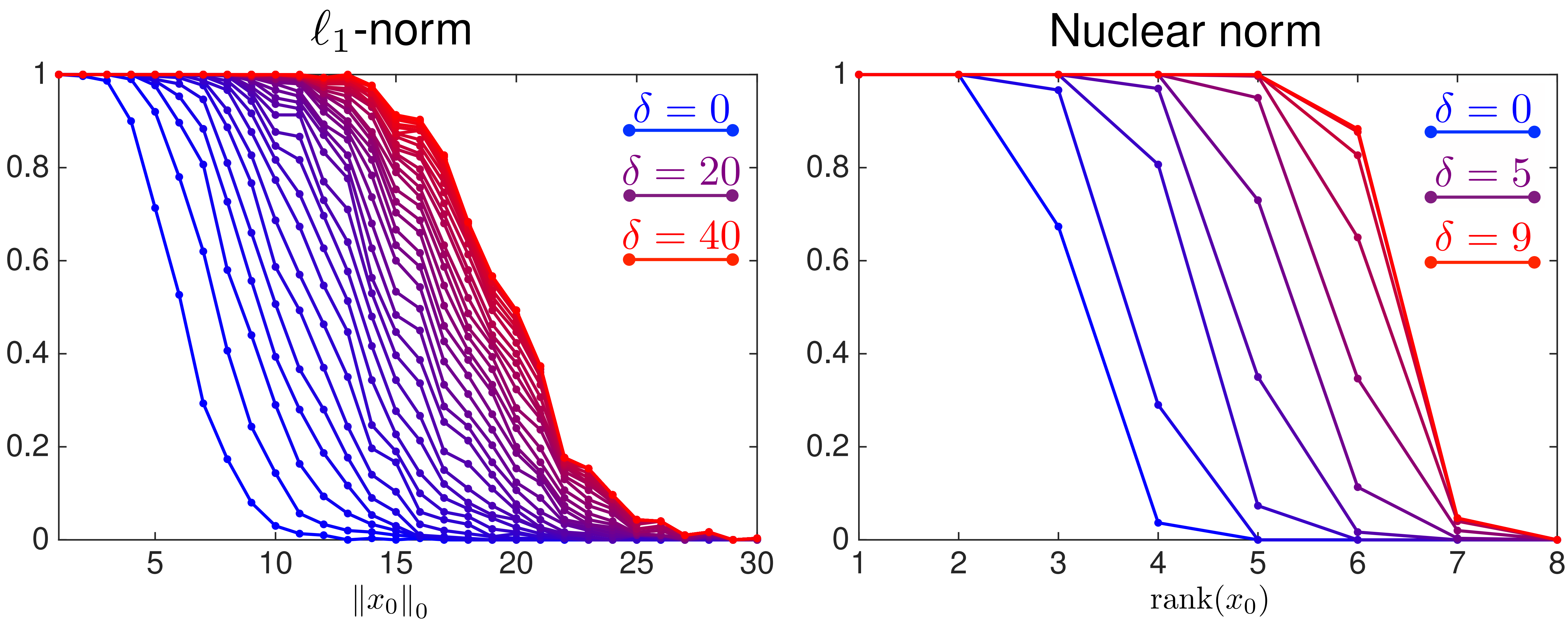
blue : exact recovery $\longrightarrow$ red : enlarged recovery

No exact recovery in general

Enlarged model recovery

Proportion of x_0 such that it is is the unique solution of $(\mathrm{P}(y_0, 0)))$ and $\delta(x_0) \leq \delta$.

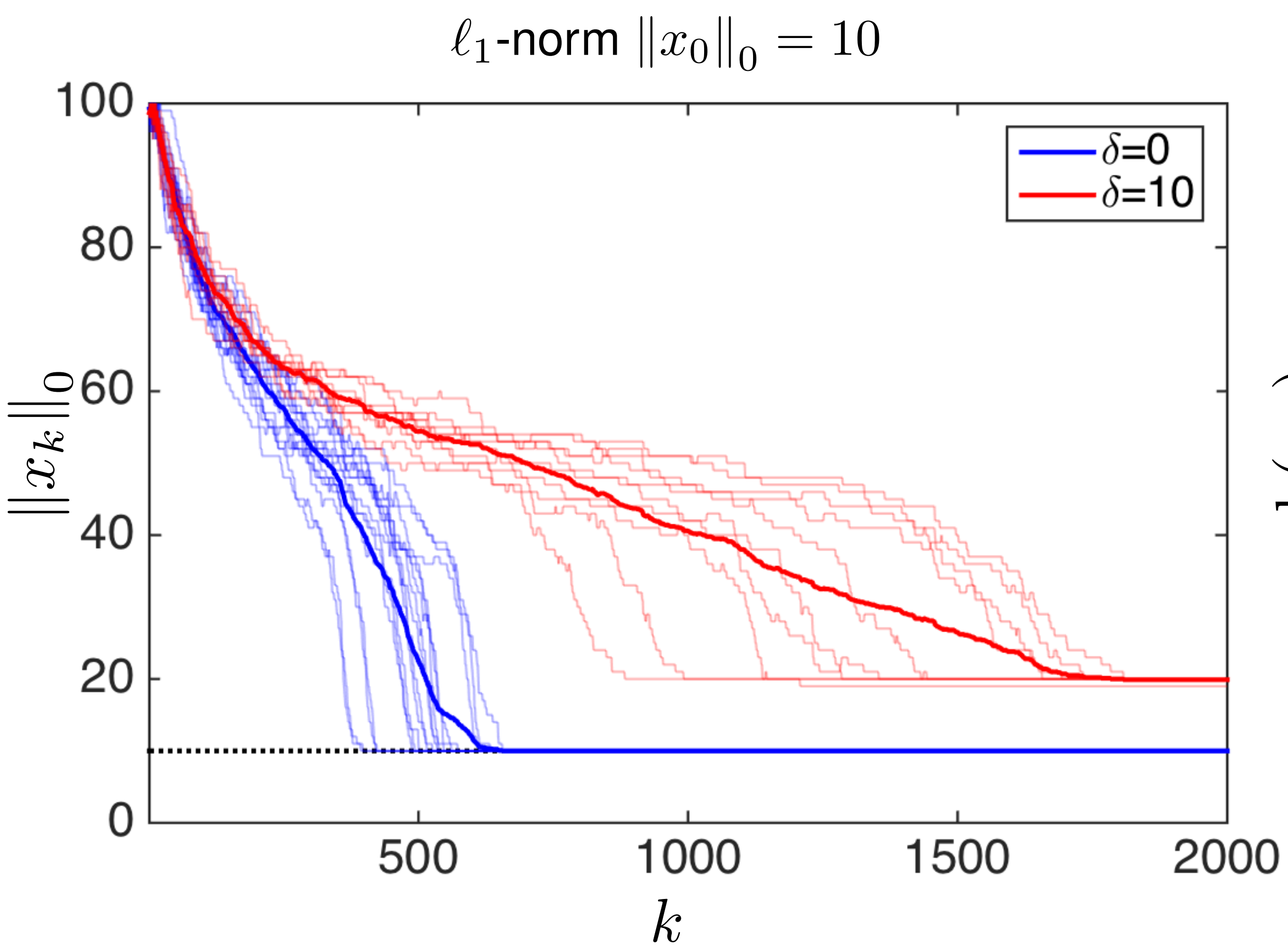
$$\delta(x_0) = \dim(\mathcal{J}_{R^*}(M^*_{u_0})) - \dim(M_{x_0})$$

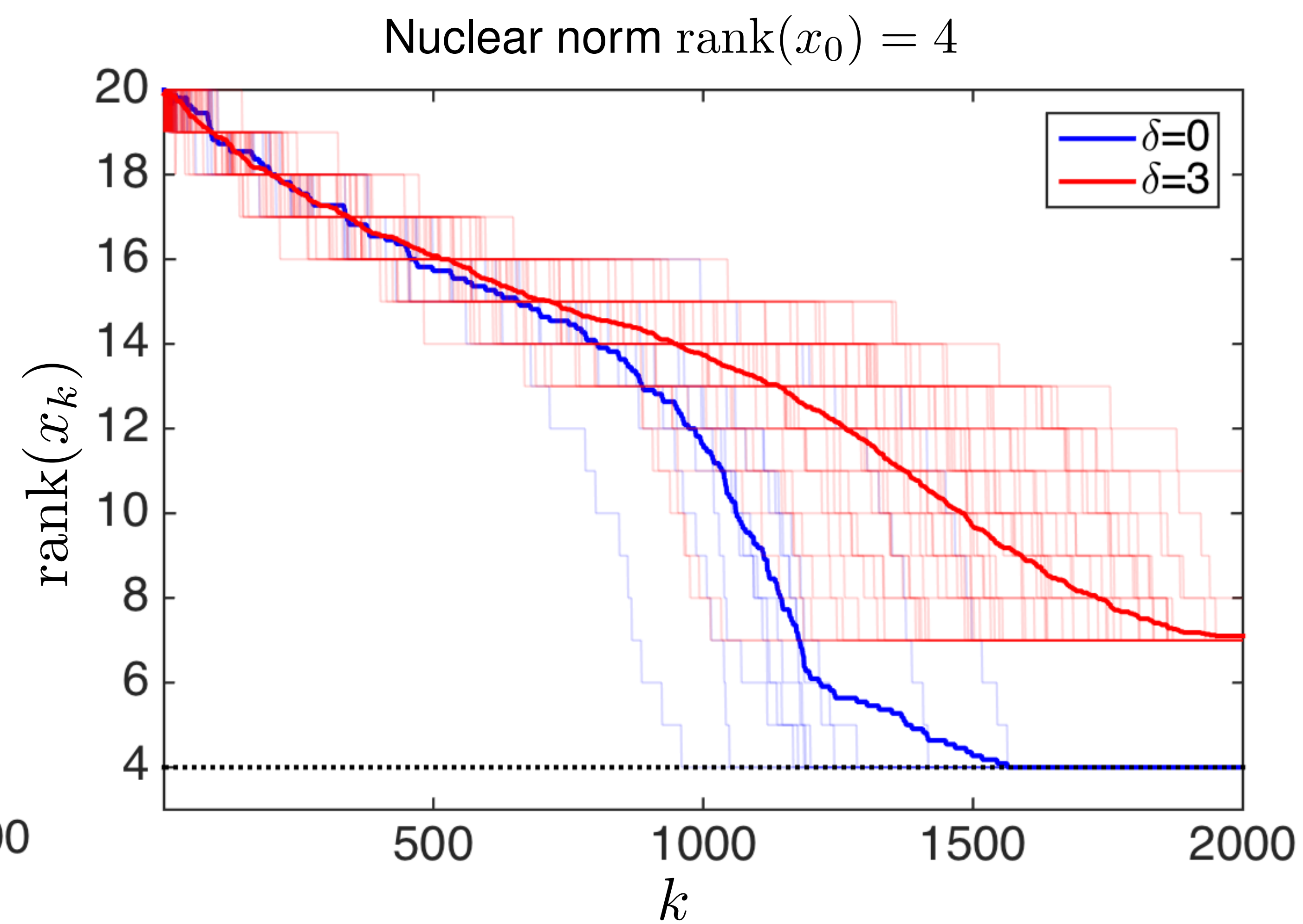


blue : stable exact recovery $\longrightarrow$ red : unstable enlarged recovery

The intermediate curves: typical tradeoff between complexity of input vectors and stable model stratum recovery in presence of even small noise.

Finite enlarged activity identification





Blue : trajectory for each realization x_0 such that $\delta(x_0) = 0 \Rightarrow$ finite identification of M_{x_0}.

Red : trajectory for each realization x_0 such that $\delta(x_0) > 0 \Rightarrow$ finite identification of

$$M_{x_0} \leq M \leq \mathcal{J}_{R^*}(M^*_{u_0}).$$

Take away messages

- A unified analysis of model recovery under degeneracy.
- Mirror-stratification is a key:
  - Enlarged model recovery for inverse problems.
  - Enlarged activity identification for proximal splitting algorithms.

- Can be generalized to maximal monotone operators.
- Why the largest stratum seems to be identified ?
- Beyond convexity.
- Infinite dimension.

Preprints on arxiv and papers on

https://fadili.users.greyc.fr/

Thanks
Any questions ?