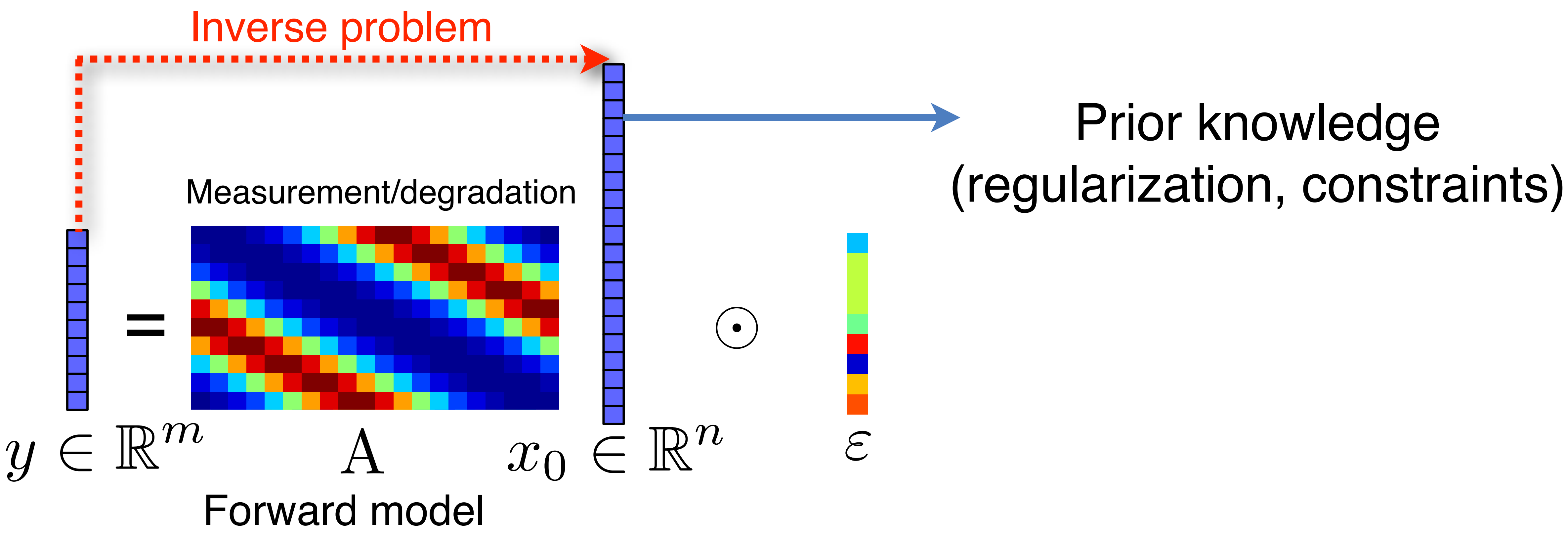
Finite Identification and Local Linear Convergence of Proximal Splitting Algorithms

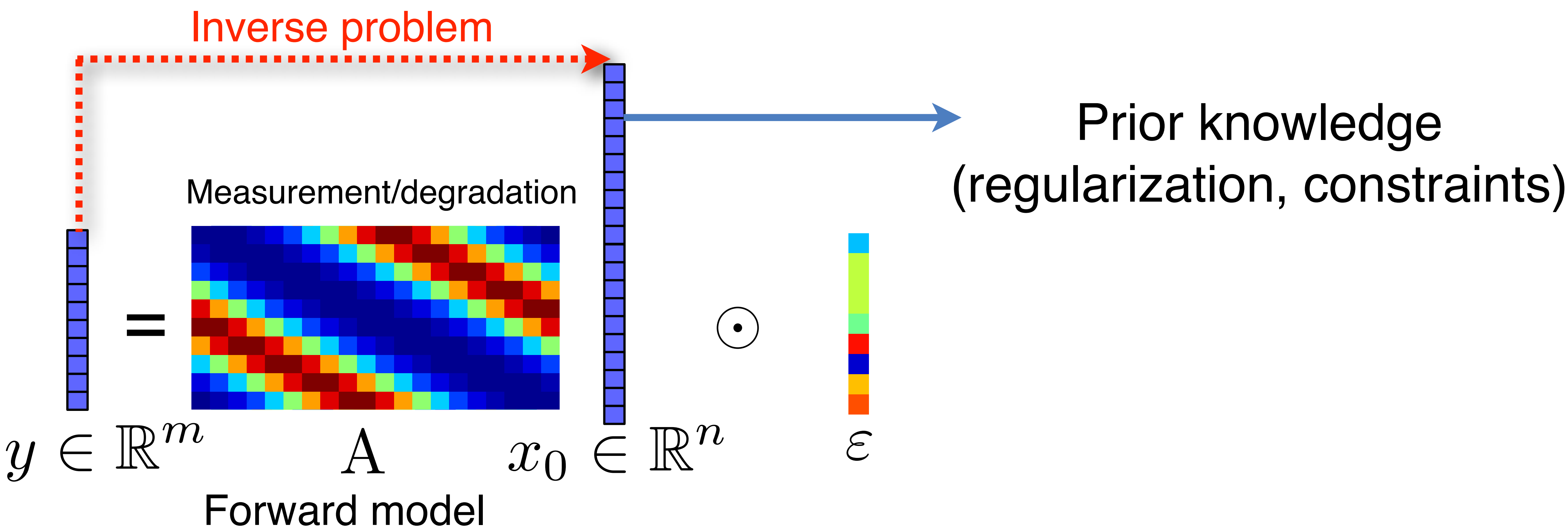
Jalal Fadili
Normandie Université-ENSICAEN, GREYC CNRS UMR 6072
Joint work with
Jingwei Liang, Gabriel Peyré and Russell Luke

TerryFest, Limoges 2015

Class of problems: motivations



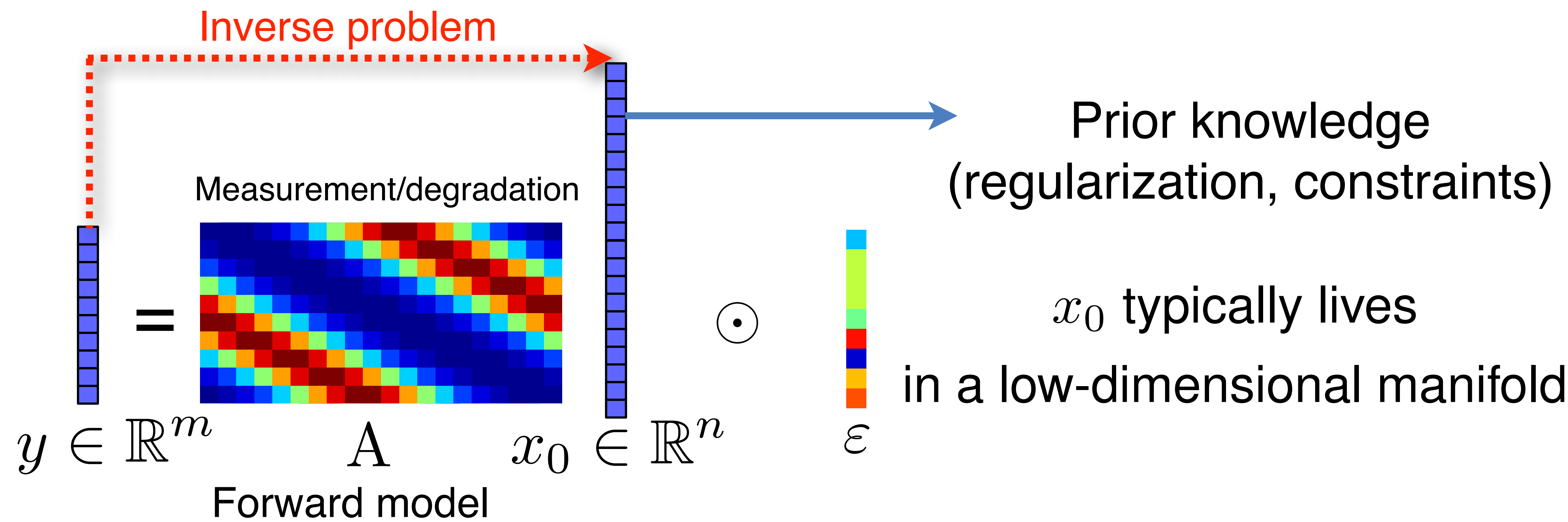
Class of problems: motivations



x_0 typically lives

in a low-dimensional manifold

Class of problems: motivations



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

F and $G \in \Gamma_0(\mathbb{R}^n)$

$$\min_{x \in \mathbb{R}^n} \underbrace{F(x)}_{\text{Data fidelity}} + \underbrace{G(x)}_{\text{Regularization, constraints}}$$

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

Low-complexity regularization

$$\min_{x \in \mathbb{R}^n} F(x) + G(x) \qquad F \text{ and } G \in \Gamma_0(\mathbb{R}^n)$$

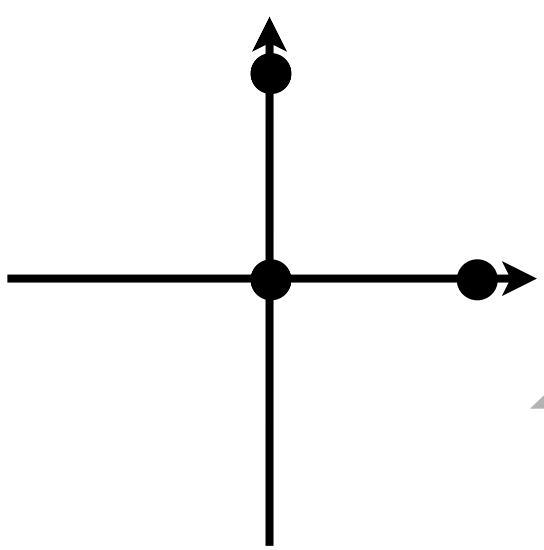
Low-complexity $\Longleftrightarrow$ Low-dimensional manifold

Low-complexity regularization

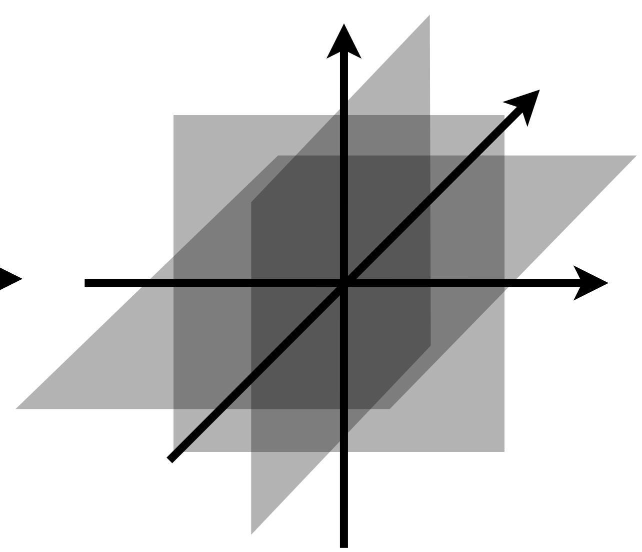
$$\min_{x \in \mathbb{R}^n} F(x) + G(x) \qquad F \text{ and } G \in \Gamma_0(\mathbb{R}^n)$$

Low-complexity $\Longleftrightarrow$ Low-dimensional manifold

Sparse vectors



$\mathbb{R}^2$



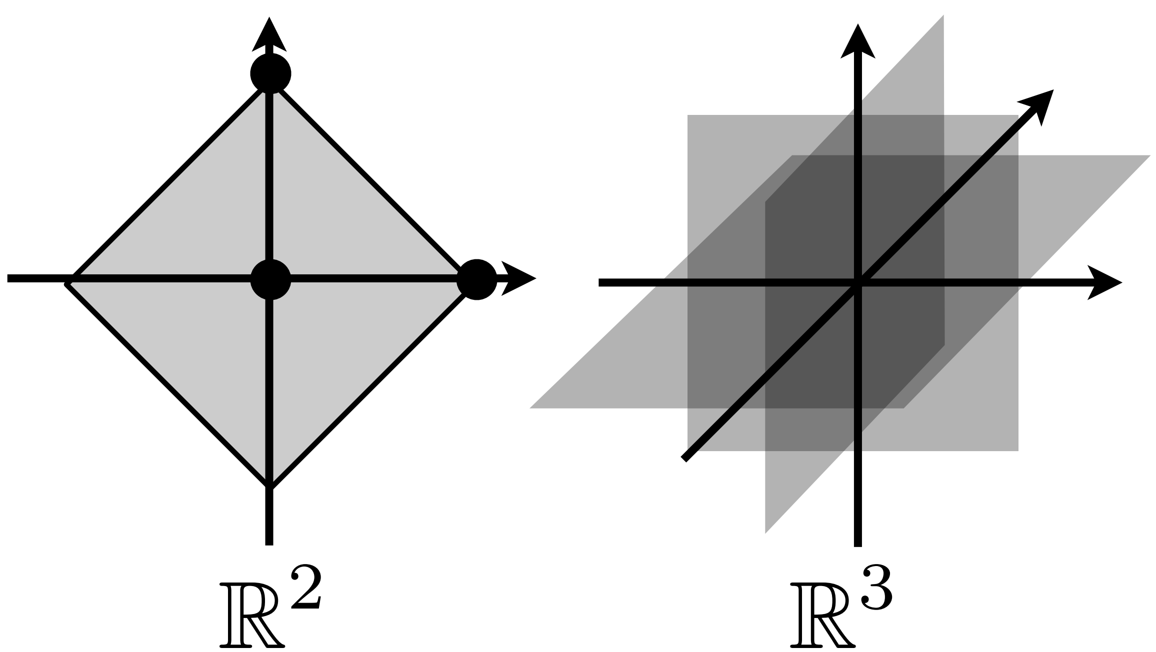
$\mathbb{R}^3$

Low-complexity regularization

$$\min_{x \in \mathbb{R}^n} F(x) + G(x) \qquad F \text{ and } G \in \Gamma_0(\mathbb{R}^n)$$

Low-complexity $\Longleftrightarrow$ Low-dimensional manifold

Sparse vectors



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

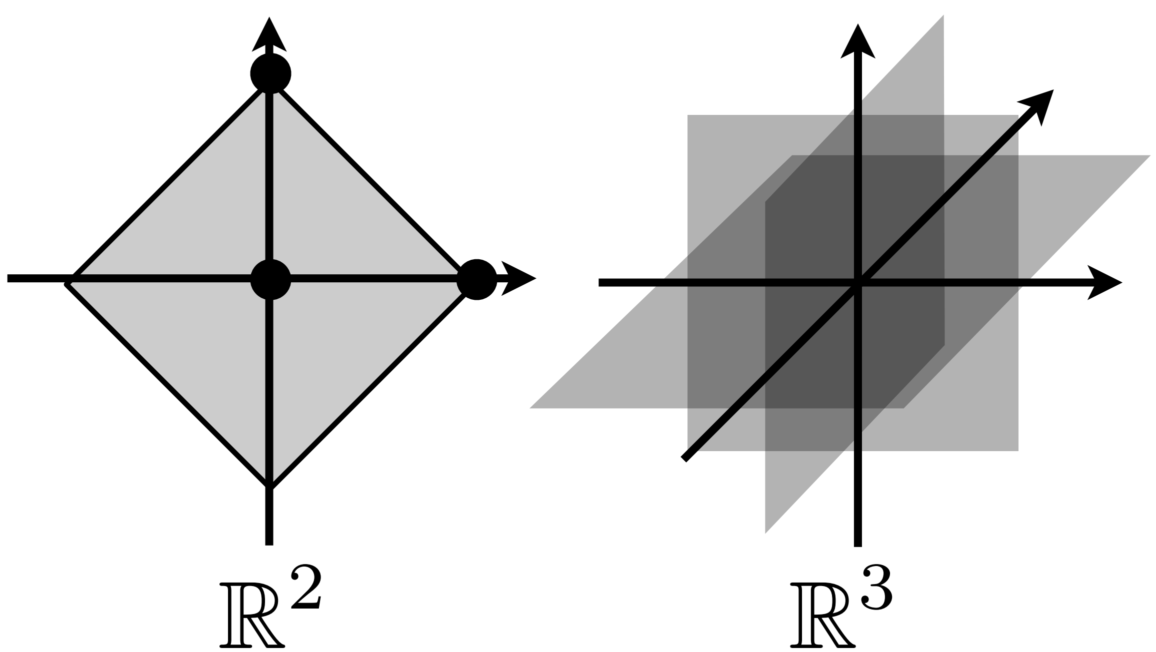
(tightest convex relaxation of ℓ_0)

Low-complexity regularization

$$\min_{x\in\mathbb{R}^n} F(x) + G(x) \qquad F \text{ and } G \in \Gamma_0(\mathbb{R}^n)$$

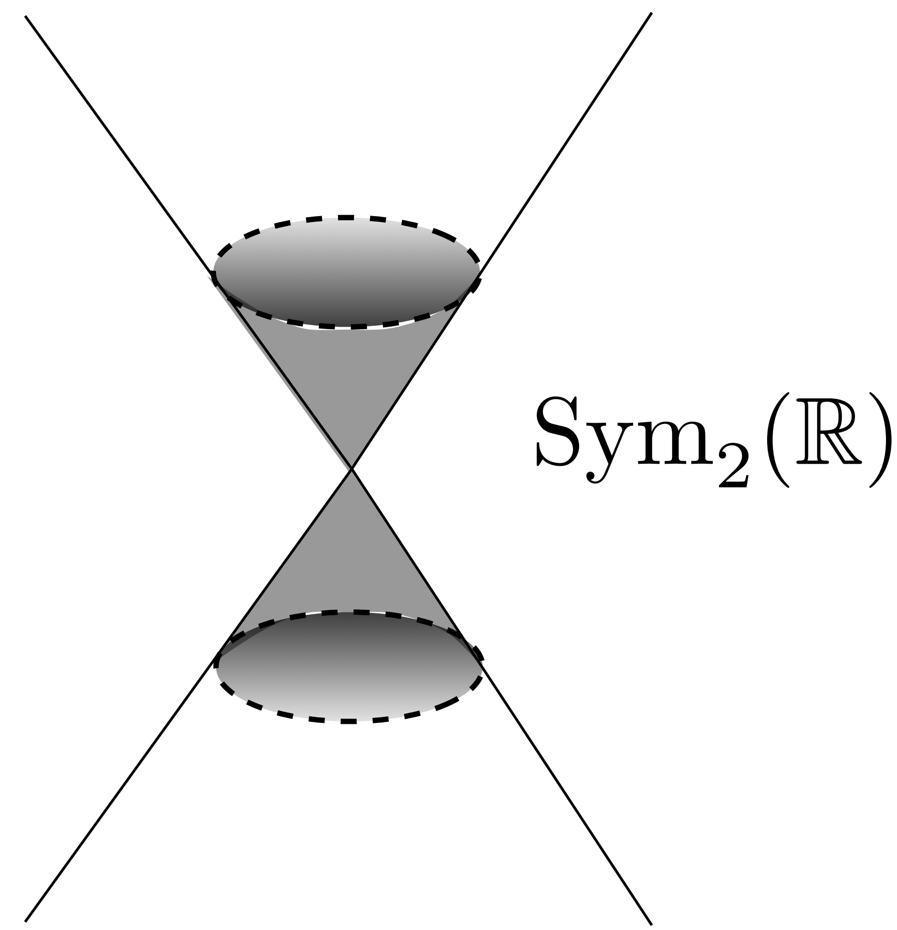
Low-complexity $\Longleftrightarrow$ Low-dimensional manifold

Sparse vectors



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

Low-rank matrices

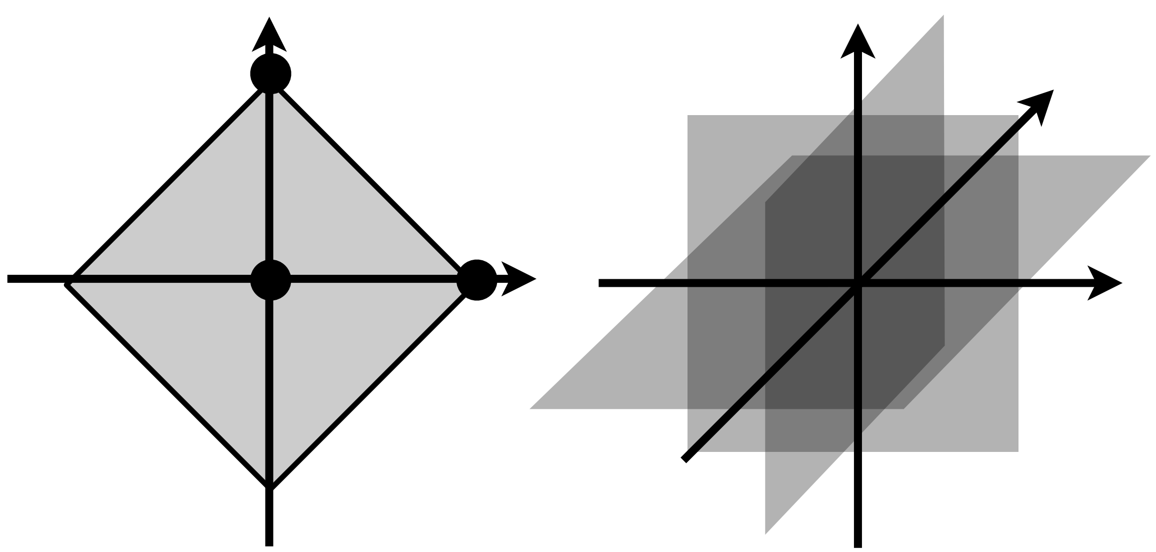


Low-complexity regularization

$$\min_{x\in\mathbb{R}^n} F(x)+G(x) \qquad F \text{ and } G \in \Gamma_0(\mathbb{R}^n)$$

Low-complexity $\Longleftrightarrow$ Low-dimensional manifold

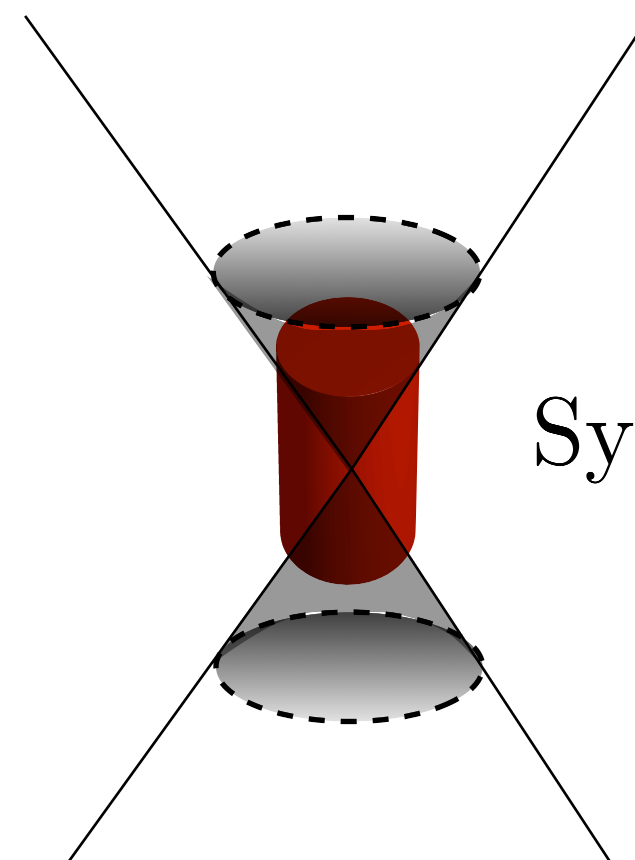
Sparse vectors



$\mathbb{R}^2$ $\mathbb{R}^3$

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

Low-rank matrices



$\mathrm{Sym}_2(\mathbb{R})$

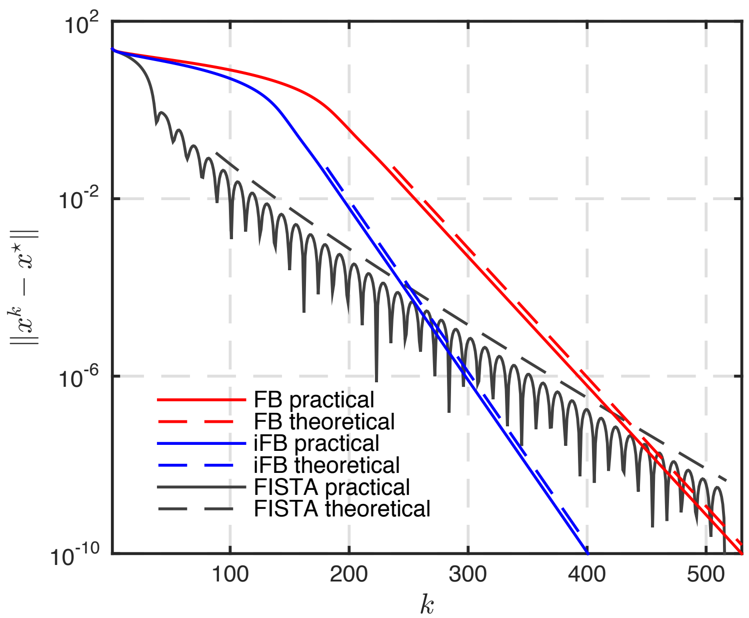
$G(x)=\|x\|_*$
(tightest convex relaxation of rank)

Proximal splitting and local linear convergence

(Inertial) Forward-Backward

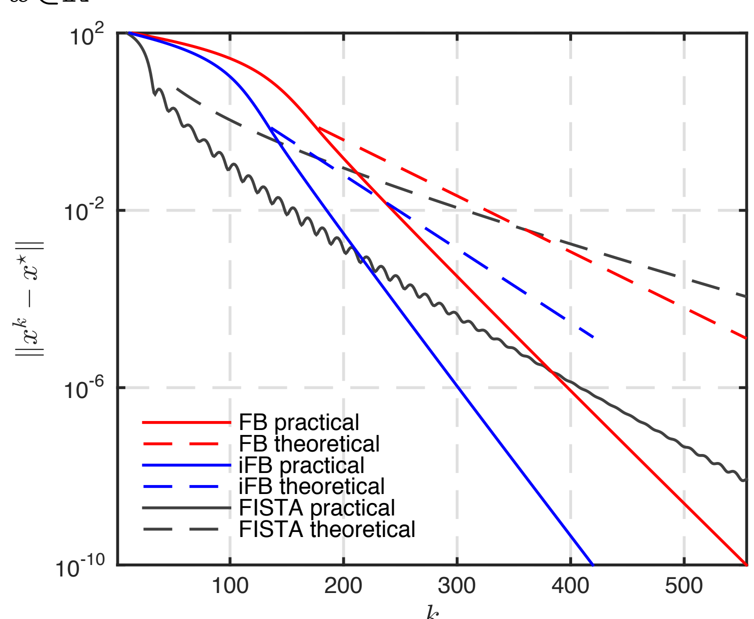
LASSO

$$\min_{x \in \mathbb{R}^n} \frac{1}{2} \|y - \mathrm{A}x\|_2^2 + \lambda \|x\|_1$$



Trace LASSO

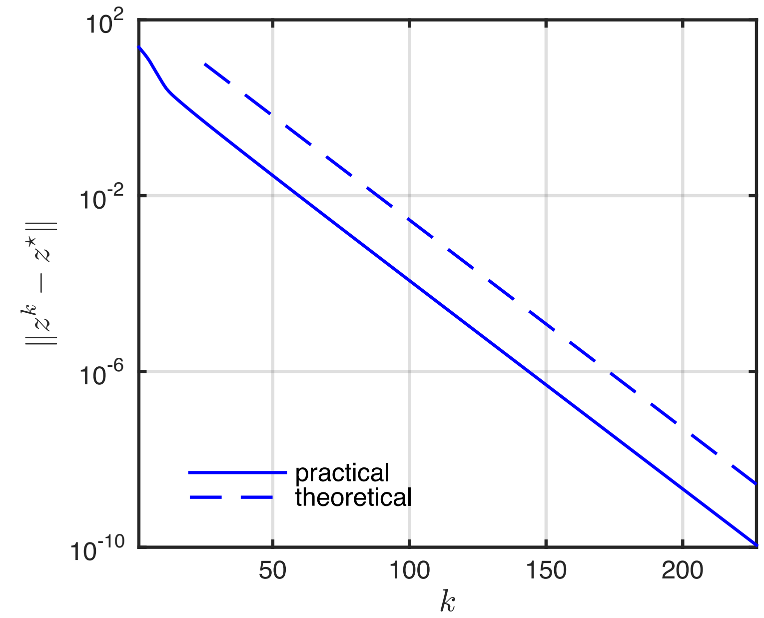
$$\min_{x \in \mathbb{R}^{n \times n}} \frac{1}{2} \|y - \mathcal{A}(x)\|_2^2 + \lambda \|x\|_*$$



Douglas-Rachford

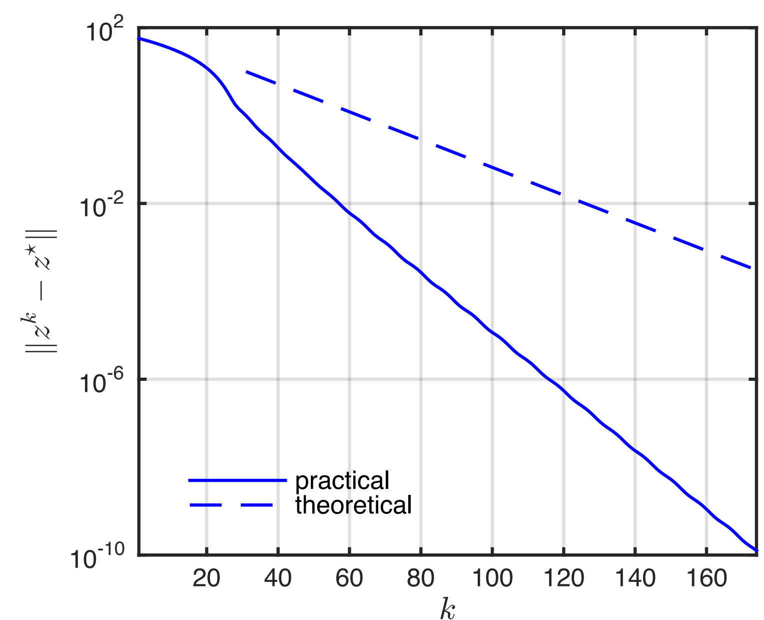
Basis Pursuit

$$\min_{x \in \mathbb{R}^n} \|x\|_1 \ \text{ s.t. } y = \mathrm{A}x$$



Trace Basis Pursuit

$$\min_{x \in \mathbb{R}^{n \times n}} \|x\|_* \ \text{ s.t. } y = \mathcal{A}(x)$$



Proximal splitting and local linear convergence

- In all examples, G (and possibly F) enjoy rich structure : **partial smoothness** (TBD shortly).
- The rationale behind observed behaviour :
  - Finite activity identification.
  - Linearization of the implicit steps.
  - Matrix recurrence and rates through sharp spectral analysis.

Outline

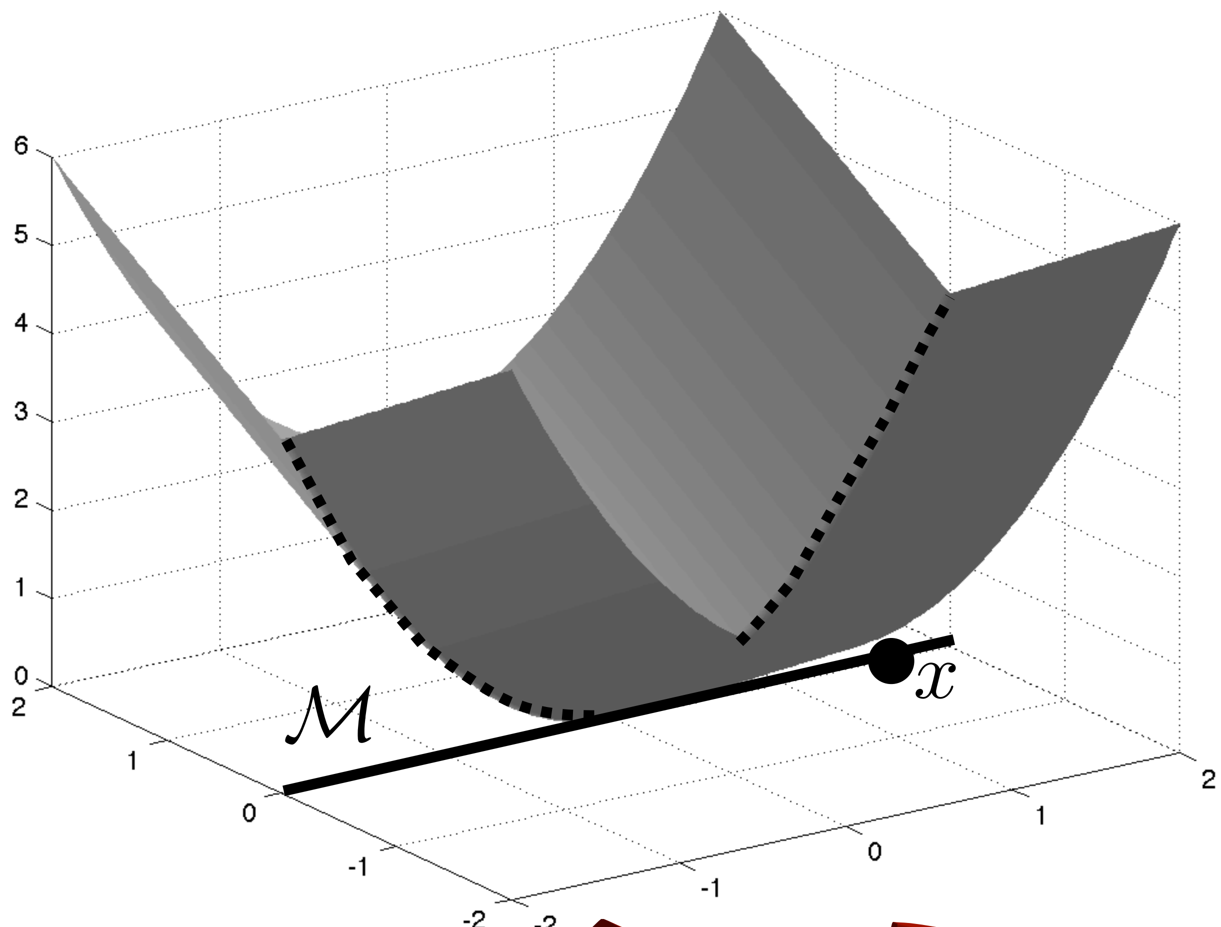
- Partial smoothness.
- Inertial Forward-Backward.
- Douglas-Rachford.
- Conclusion and future work.

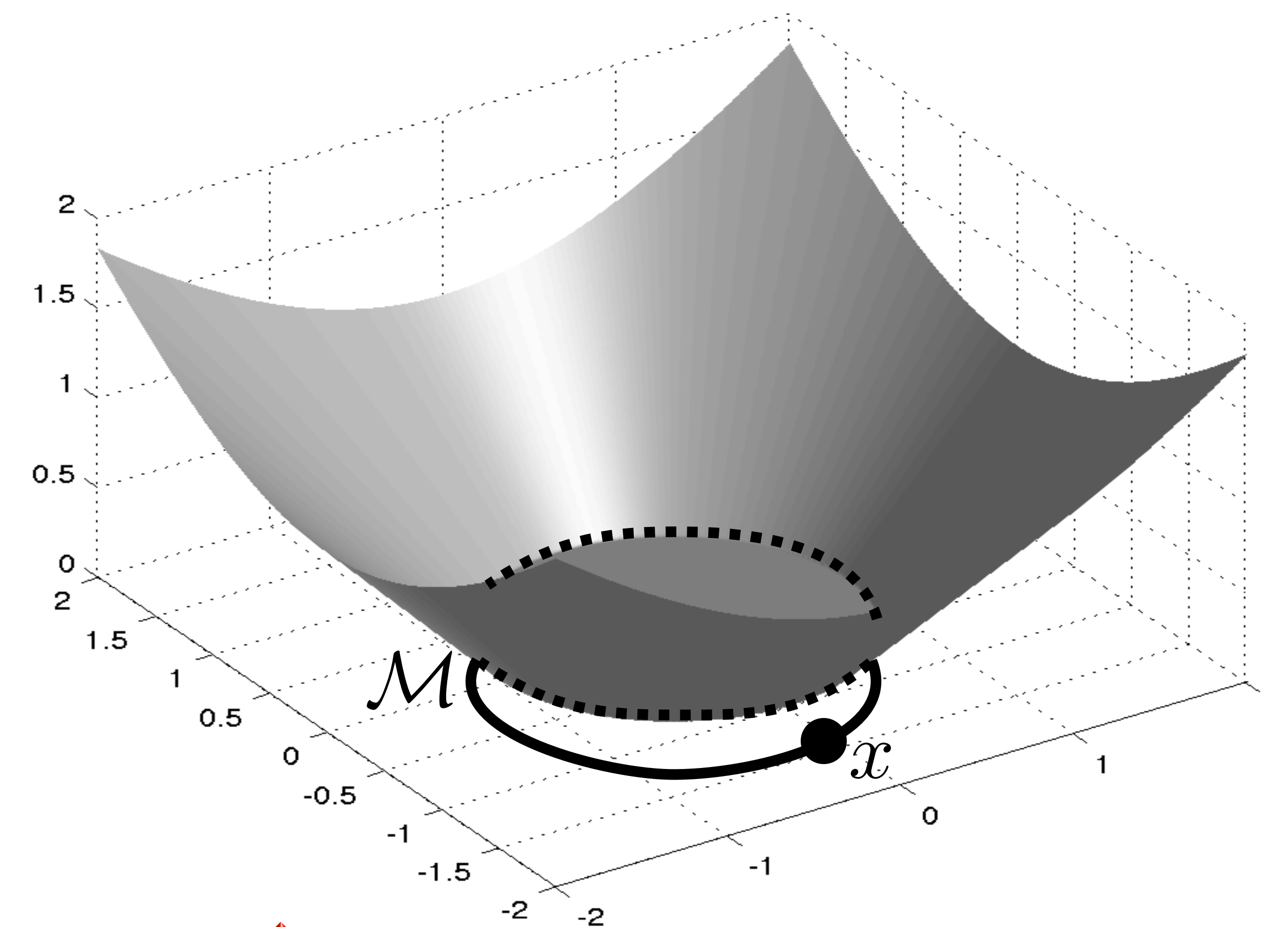
Outline

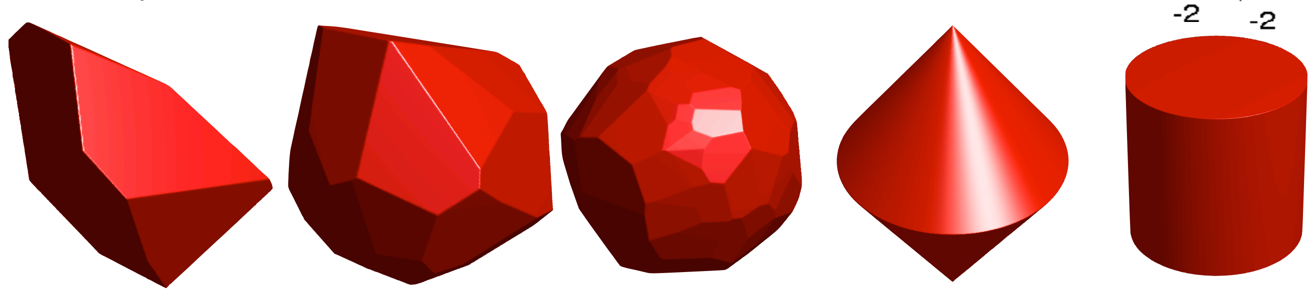
- Partial smoothness.
- Inertial Forward-Backward.
- Douglas-Rachford.
- Conclusion and future work.

Partly smooth functions

- A partly smooth function [Lewis 2002] behaves smoothly along a manifold $\mathcal{M}$, and sharply normal to it.
- The behaviour of the function and its minimizers (or critical points) depend essentially on its restriction to $\mathcal{M}$.
- Offering a powerful framework for sensitivity analysis and activity identification.







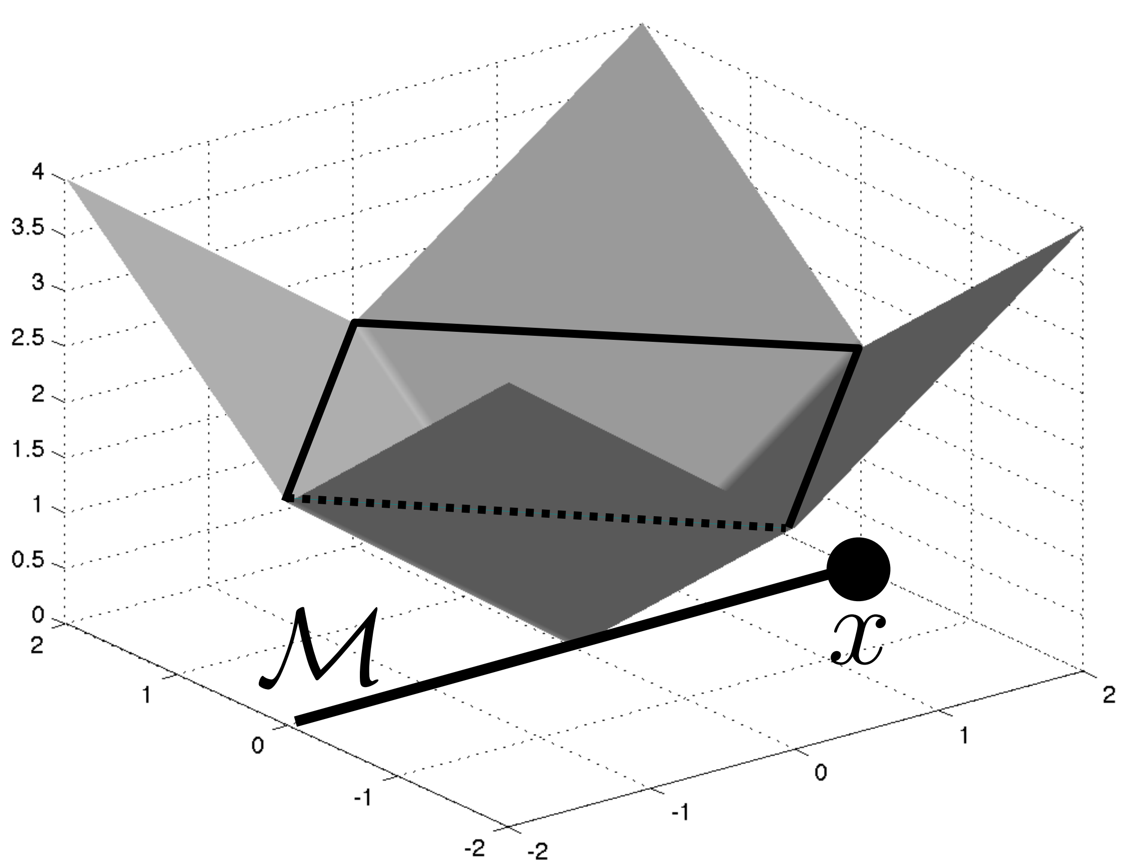
Partly smooth functions

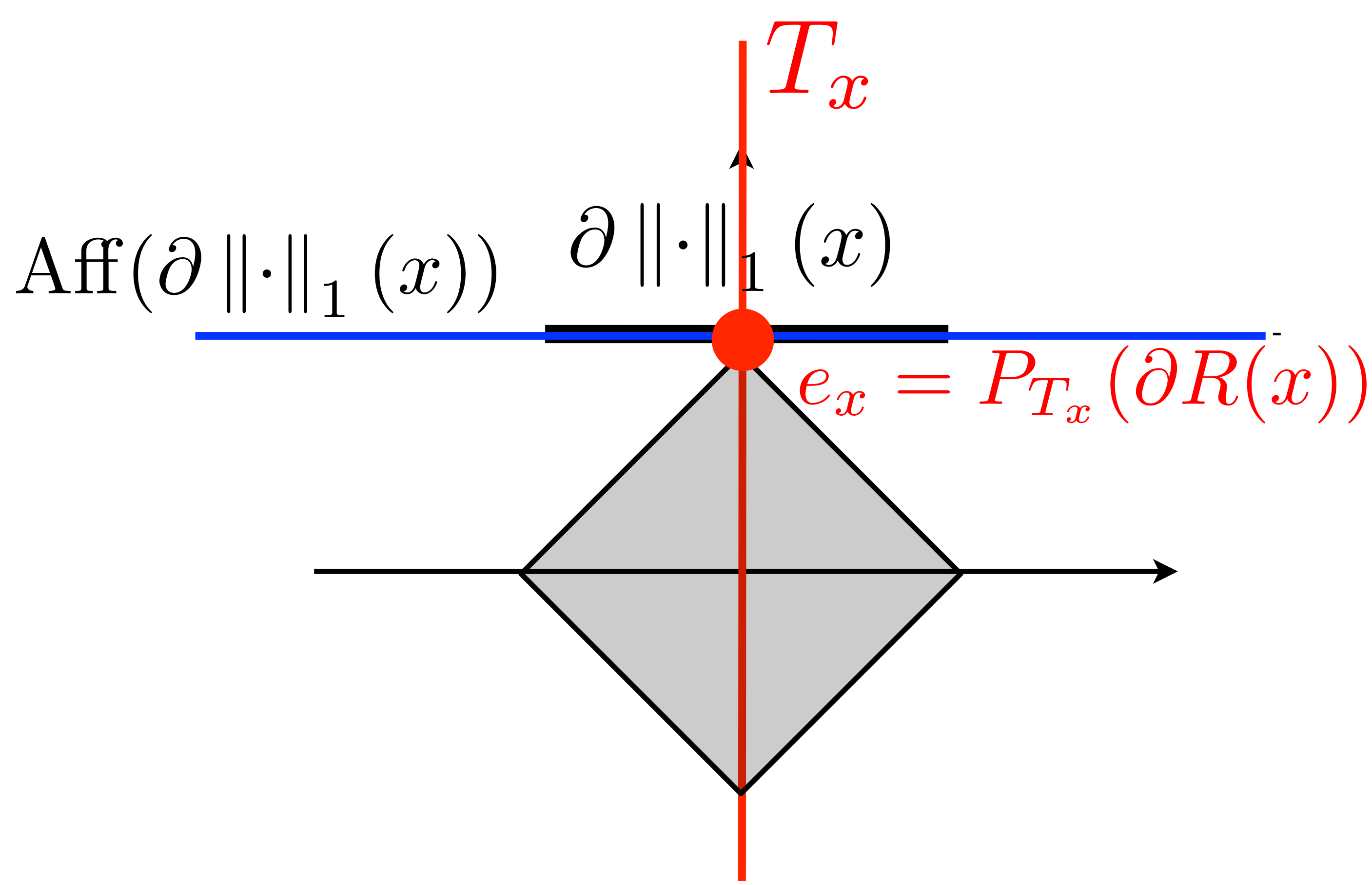
Definition *Let* $R \in \Gamma_0(\mathbb{R}^n)$. R *is* partly smooth at x relative to a set $\mathcal{M} \ni x$, *i.e.* $R \in \mathrm{PS}_x(\mathcal{M})$, *if*

(i) (Smoothness) $\mathcal{M}$ *is a* C^2*-manifold around* x *and* R *restricted to* $\mathcal{M}$ *is* C^2 *around* x.

(ii) (Sharpness) The tangent space $\mathcal{T}_x(\mathcal{M}) = T_x := \mathrm{par}(\partial R(x))^\perp$.

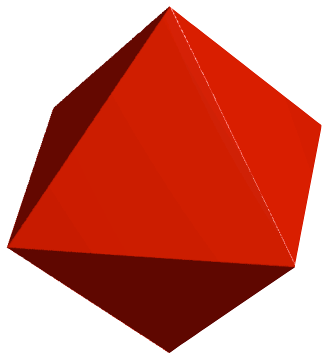
(iii) (Continuity) The set-valued mapping ∂R *is continuous at* x *relative to* $\mathcal{M}$.





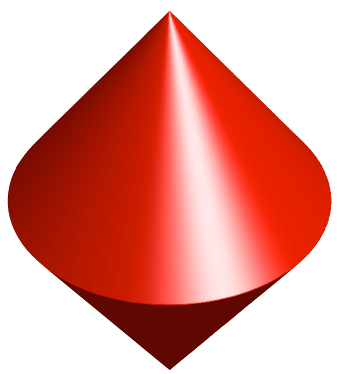
Examples

$\|x\|_1$



$$\mathcal{M} = T_x = \left\{u \in \mathbb{R}^N \; ; \; \mathrm{supp}(u) \subseteq \mathrm{supp}(x)\right\}.$$

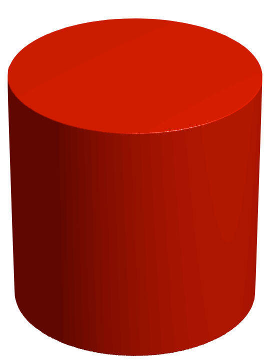
$\sum_b \|x_b\|_2$



$$\mathcal{M} = T_x = \{u \; ; \; \mathrm{supp}_{\mathcal{B}}(u) \subseteq \mathrm{supp}_{\mathcal{B}}(x)\}$$

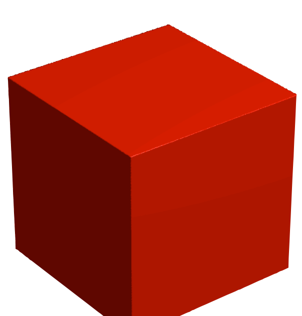
$\mathrm{supp}_{\mathcal{B}}(x) = \bigcup \{b \; ; \; x_b \neq 0\}.$

$\|x\|_*$



$$\mathcal{M} = \{u \; ; \; \mathrm{rank}(u) = \mathrm{rank}(x)\}.$$

$\|x\|_\infty$



$$\mathcal{M} = T_x = \left\{u \in \mathbb{R}^N \; ; \; u_I \propto \mathrm{sign}(x_I)\right\}$$

$I = \{i \; ; \; |x_i| = \|x\|_\infty\}.$

Calculus rules

Proposition ([Lewis 2002, Daniilidis et al. 2014])

- ***Sum and pre-composition*** *: The set of continuous convex partly smooth functions is closed under addition and pre-composition by a linear operator (under a mild transversality condition).*

Calculus rules

Proposition ([Lewis 2002, Daniilidis et al. 2014])

- ***Sum and pre-composition*** *: The set of continuous convex partly smooth functions is closed under addition and pre-composition by a linear operator (under a mild transversality condition).*
- **Smooth perturbation** : If a function is partly smooth function, then so is its smooth perturbation.

Calculus rules

Proposition ([Lewis 2002, Daniiilidis et al. 2014])

- ***Sum and pre-composition*** *: The set of continuous convex partly smooth functions is closed under addition and pre-composition by a linear operator (under a mild transversality condition).*
- **Smooth perturbation** : If a function is partly smooth function, then so is its smooth perturbation.
- **Spectral lift** : Absolutely permutation-invariant convex and partly smooth functions of the singular values of a real matrix are convex and partly smooth.

Outline

- Partial smoothness.
- Inertial Forward-Backward.
- Douglas-Rachford.
- Conclusion and future work.

Inertial Forward-Backward

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

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

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

$$\gamma_k \in [\epsilon, 2\beta - \epsilon] \quad \begin{cases} y_k^a &= x_k + a_k(x_k - x_{k-1}), \quad a_k \in [0,1] \\ y_k^b &= x_k + b_k(x_k - x_{k-1}), \quad b_k \in [0,1] \\ x_{k+1} &= \operatorname{prox}_{\gamma_k G}\left(y_k^a - \gamma_k \nabla F(y_k^b)\right) \end{cases}$$

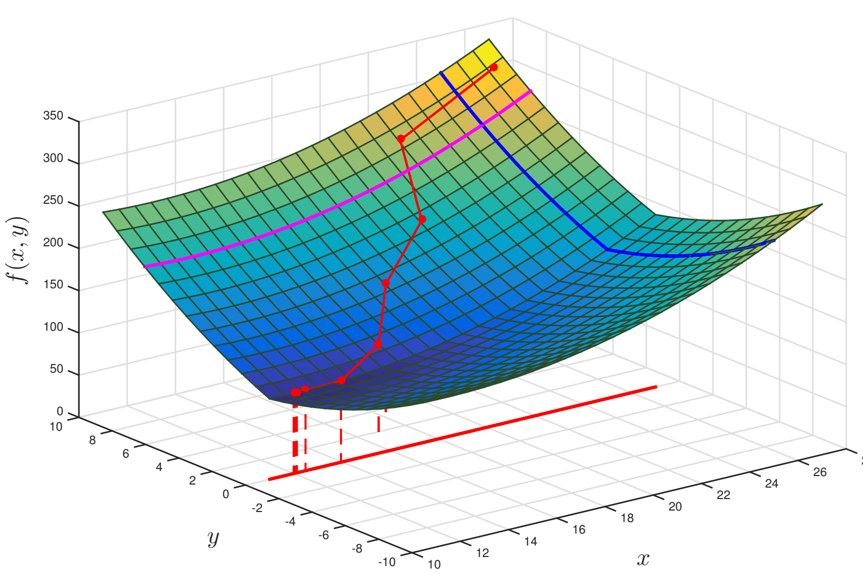
Inertial Forward-Backward

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

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

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

$$\gamma_k \in [\epsilon, 2\beta - \epsilon] \quad \begin{cases} y_k^a & = x_k + a_k(x_k - x_{k-1}), \quad a_k \in [0,1] \\ y_k^b & = x_k + b_k(x_k - x_{k-1}), \quad b_k \in [0,1] \\ x_{k+1} & = \mathrm{prox}_{\gamma_k G}\left(y_k^a - \gamma_k \nabla F(y_k^b)\right) \end{cases}$$



Inertial Forward-Backward

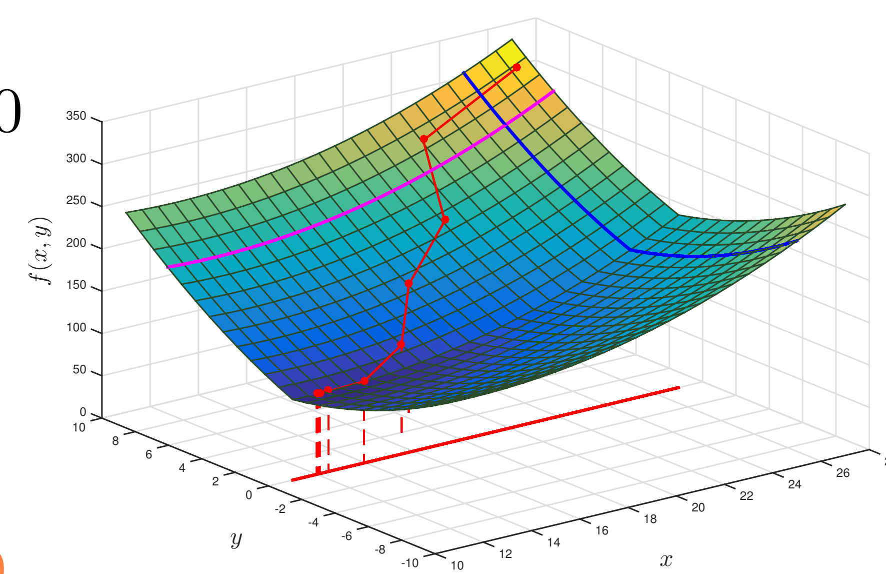
$$\min_{x\in\mathbb{R}^n} F(x)+G(x)$$

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

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

$$\gamma_k \in [\epsilon, 2\beta-\epsilon] \quad \begin{cases} y_k^a & = x_k + a_k(x_k - x_{k-1}), \quad a_k \in [0,1] \\ y_k^b & = x_k + b_k(x_k - x_{k-1}), \quad b_k \in [0,1] \\ x_{k+1} & = \operatorname{prox}_{\gamma_k G}\left(y_k^a - \gamma_k \nabla F(y_k^b)\right) \end{cases}$$

- PPA ($F=0$) : [Alvarez and Attouch 2001].
- $a_k = b_k = 0$: FB [Lions and Mercier 1979].
- $b_k = 0$: [Moudafi and Oliny 2003] (heavy ball method if $G=0$ [Attouch et al. 2000]).
- $a_k = b_k$:
  - [Lorenz and Pock 2014].
  - FISTA (Beck-Teboulle, Chambolle-Dossal).
  - $\gamma_k \in]0,\beta]$ and $a_k \to 1$: FISTA-like [Beck and Teboulle 2009, Chambolle and Dossal 2015].



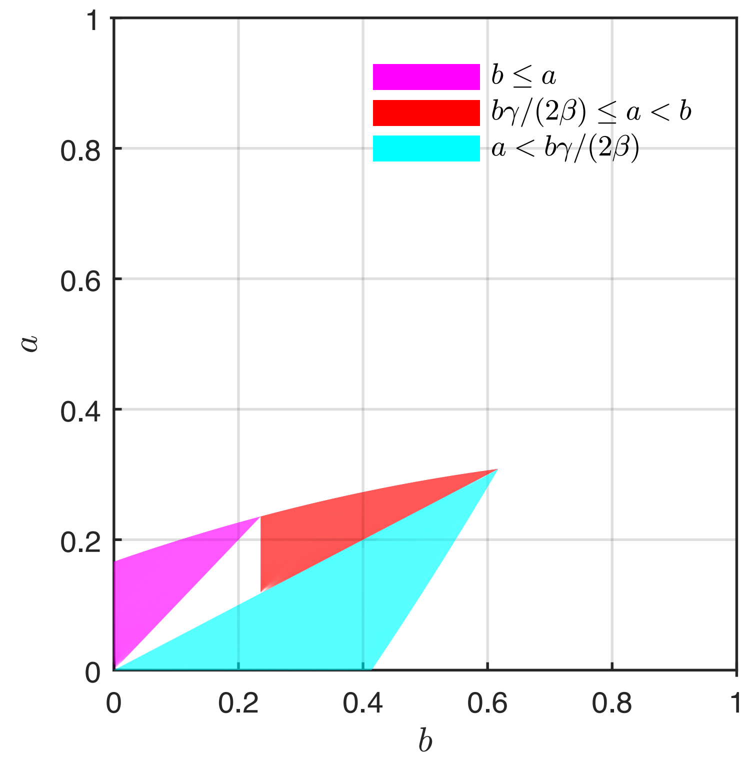
Global convergence

Theorem *Let* $\epsilon \in]0, \beta[$ *and* $\bar{a} < 1$ $\bar{b} < 1$. *Suppose that* $\gamma_k \in [\epsilon, 2\beta - \epsilon]$,$a_k \in]0, \bar{a}]$, $b_k \in]0, \bar{b}]$, $\tau > 0$ *is such that either of the following holds :*

(i) $(1 + a_k) - \frac{\gamma_k}{2\beta}(1 + b_k)^2 > \tau$ *: for* $a_k < \frac{\gamma_k}{2\beta} b_k$ *; or*

(ii) $(1 - 3a_k) - \frac{\gamma_k}{2\beta}(1 - b_k)^2 > \tau$ *: for* $b_k \leq a_k$ *or* $\frac{\gamma_k}{2\beta} b_k \leq a_k < b_k$.

Then $(x_k)_{k\in\mathbb{N}}$ *is asymptotically regular and converges to* $x^\star \in \operatorname{Argmin}(F + G)$.



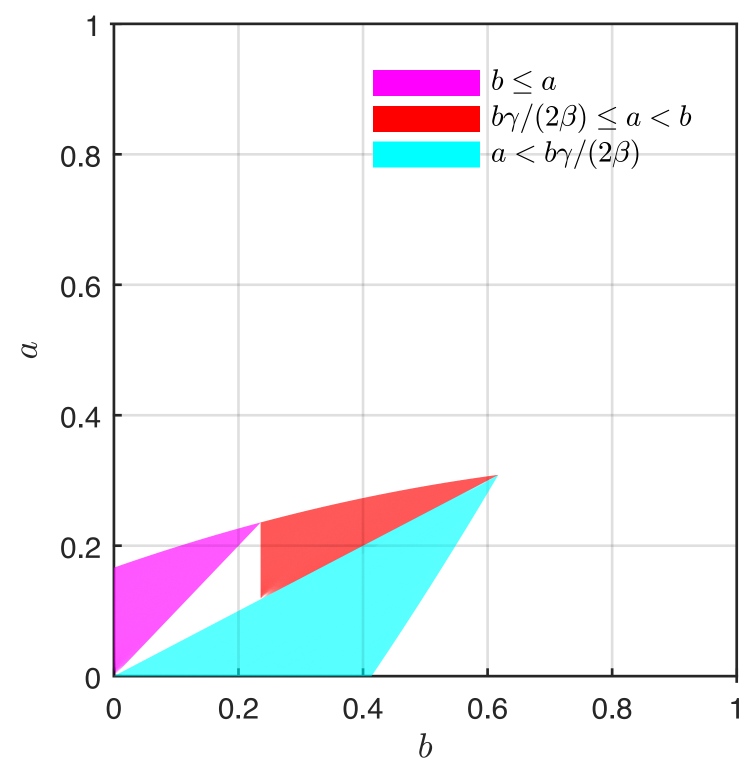
Global convergence

Theorem *Let* $\epsilon \in]0, \beta[$ *and* $\bar{a} < 1$ $\bar{b} < 1$. *Suppose that* $\gamma_k \in [\epsilon, 2\beta - \epsilon]$,$a_k \in]0, \bar{a}]$, $b_k \in]0, \bar{b}]$, $\tau > 0$ *is such that either of the following holds :*

(i) $(1 + a_k) - \frac{\gamma_k}{2\beta}(1 + b_k)^2 > \tau$ *: for* $a_k < \frac{\gamma_k}{2\beta} b_k$ *; or*

(ii) $(1 - 3a_k) - \frac{\gamma_k}{2\beta}(1 - b_k)^2 > \tau$ *: for* $b_k \le a_k$ *or* $\frac{\gamma_k}{2\beta} b_k \le a_k < b_k$.

Then $(x_k)_{k\in\mathbb{N}}$ *is asymptotically regular and converges to* $x^\star \in \mathrm{Argmin}(F + G)$.



Theorem (**[Chambolle and Dossal 2015, Attouch's talk]**) *Suppose that* $\gamma_k \in]0, \beta]$, *and take* $a_k = b_k = \frac{k-1}{k+p}$, $\forall p > 2$. *Then* $(x_k)_{k\in\mathbb{N}}$ *is asymptotically regular and converges to* $x^\star \in \mathrm{Argmin}(F + G)$.

Identification and local linear convergence

Theorem *Let the iFB be used to create a sequence x_k which converges to $x^\star \in \operatorname{Argmin}(F+G)$, such that $R \in \mathrm{PS}_{x^\star}(\mathcal{M}_{x^\star})$, F is C^2 near $x^\star$ and*

$$-\nabla F(x^\star) \in \operatorname{ri}\left(\partial G(x^\star)\right).$$

Then the following holds,

(1) The iFB has the finite identification property, i.e. $x_k \in \mathcal{M}_{x^\star}$ for k large enough. If $\mathcal{M}_{x^\star}$ is affine (or linear), then also y_k^a and $y_k^b \in \mathcal{M}_{x^\star}$ for large k.

(2) Suppose moreover there exists $\alpha \geq 0$ such that

$$\mathrm{P}_T \, \nabla^2 F(x^\star) \, \mathrm{P}_T \succ \alpha \mathrm{Id}, \qquad T := T_{x^\star} \,.$$

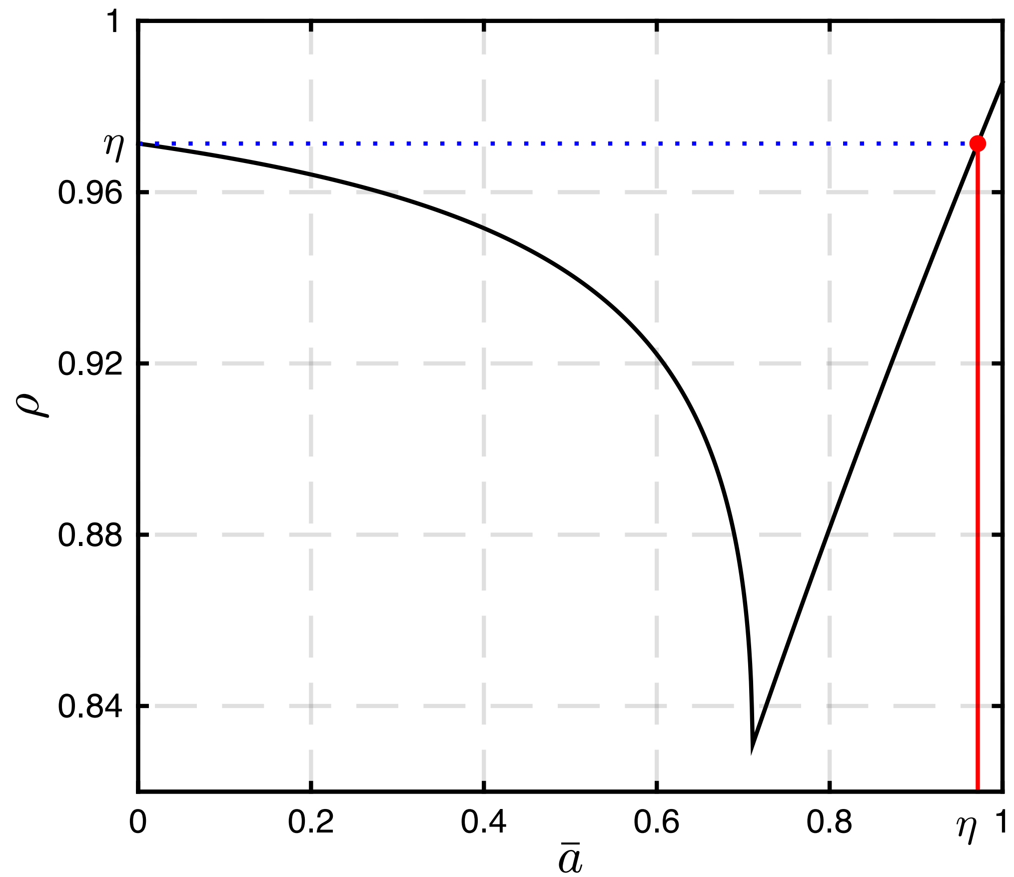
Then $\forall k$ large enough, the following holds.

(i) Q-linear convergence *: if $0 < \underline{\gamma} \leq \gamma_k \leq \bar{\gamma} < \min(2\alpha\beta^2, 2\beta)$, then $\|x_{k+1} - x^\star\|$ converges Q-linearly.*

(ii) R-linear convergence *: if $\mathcal{M}_{x^\star}$ is affine (or linear), then $\|x_{k+1} - x^\star\|$ converges R-linearly.*

Identification and local linear convergence

- The rates are expressed analytically (see [Liang, Fadili and Peyré 2015]).
- $\mathcal{M}_{x^\star}$ affine/linear : the rate estimate is tight.
- G locally polyhedral at $x^\star$:
  - the rate estimate is optimal.
  - the restricted injectivity assumption can be removed (less sharp rate).
  - with $F = \frac{1}{2}\|y - \mathrm{A}\cdot\|_2^2$: explicit equation/finite termination.
- Though iFB can be globally faster than FB, the situation changes locally : for $\gamma_k \in]0, \beta]$, $\rho_k \in]\eta_k, \sqrt{\eta_k}]$ for $a_k > \eta_k$.



Random convex programs

$$\min_{x \in \mathbb{R}^n} \frac{1}{2} \|y - \mathrm{A}x\|_2^2 + \lambda R(x) \qquad (\mathsf{P}_\lambda)$$

$$\min_{x \in \mathbb{R}^n} R(x) \quad \text{s.t.} \quad y = \mathrm{A}x \qquad (\mathrm{BP}_\mathrm{R})$$

Theorem *Let $x^\star$ be a feasible point of (BP_R) such that $R \in \mathrm{PS}_{x^\star}(\mathcal{M}_{x^\star})$ and that*

$$\ker(\mathrm{A}) \cap T_{x^\star} = \{0\}, \quad \text{and} \quad (\mathrm{A}_T^+ \mathrm{A})^* e_{x^\star} \in \mathrm{ri}(\partial R(x^\star)). \tag{1}$$

Then, for λ sufficiently small, (P_λ) has a unique minimizer, and the iFB applied to solve it identifies $\mathcal{M}_{x^\star}$ in finite time, and then converges locally linearly.

Random convex programs

$$\min_{x\in\mathbb{R}^n} \frac{1}{2}\|y-\mathrm{A}x\|_2^2+\lambda R(x) \qquad (\mathsf{P}_\lambda)$$

$$\min_{x\in\mathbb{R}^n} R(x) \quad \text{s.t.} \quad y=\mathrm{A}x \qquad (\mathrm{BP}_\mathrm{R})$$

Theorem *Let $x^\star$ be a feasible point of (BP_R) such that $R \in \mathrm{PS}_{x^\star}(\mathcal{M}_{x^\star})$ and that*

$$\ker(\mathrm{A}) \cap T_{x^\star} = \{0\}, \quad \text{and} \quad (\mathrm{A}_T^+\mathrm{A})^* e_{x^\star} \in \mathrm{ri}(\partial R(x^\star)). \tag{1}$$

Then, for λ sufficiently small, (P_λ) has a unique minimizer, and the iFB applied to solve it identifies $\mathcal{M}_{x^\star}$ in finite time, and then converges locally linearly.

Proposition (Gaussian measurements) *Choose A from the standard Gaussian ensemble (iid $\sim \mathcal{N}(0,1)$ entries).*

(i) $R=\|\cdot\|_1$: let $s=\|x^\star\|_0$. If $m > 2\beta s\log(n)+s$ for some $\beta>1$, then (1) is in force w.o.p..

(ii) $R=\|\cdot\|_*$: let $r=\mathrm{rank}(x^\star)$, $x^\star\in\mathbb{R}^{n\times n}$. If $m \geq \beta r(6n-5r)$ for some $\beta>1$, then (1) is in force w.o.p..

Random convex programs

$$\min_{x\in\mathbb{R}^n} \frac{1}{2}\|y - \mathrm{A}x\|_2^2 + \lambda R(x) \qquad (\mathsf{P}_\lambda)$$

$$\min_{x\in\mathbb{R}^n} R(x) \quad \text{s.t.} \quad y = \mathrm{A}x \qquad (\mathrm{BP_R})$$

Theorem *Let $x^\star$ be a feasible point of $(\mathrm{BP_R})$ such that $R \in \mathrm{PS}_{x^\star}(\mathcal{M}_{x^\star})$ and that*

$$\ker(\mathrm{A}) \cap T_{x^\star} = \{0\}, \quad \text{and} \quad (\mathrm{A}_T^+\mathrm{A})^* e_{x^\star} \in \mathrm{ri}(\partial R(x^\star)). \tag{1}$$

Then, for λ sufficiently small, (P_λ) has a unique minimizer, and the iFB applied to solve it identifies $\mathcal{M}_{x^\star}$ in finite time, and then converges locally linearly.

Proposition (Gaussian measurements) *Choose A from the standard Gaussian ensemble (iid $\sim \mathcal{N}(0,1)$ entries).*

(i) $R = \|\cdot\|_1$ *: let $s = \|x^\star\|_0$. If $m > 2\beta s \log(n) + s$ for some $\beta > 1$, then (1) is in force w.o.p..*

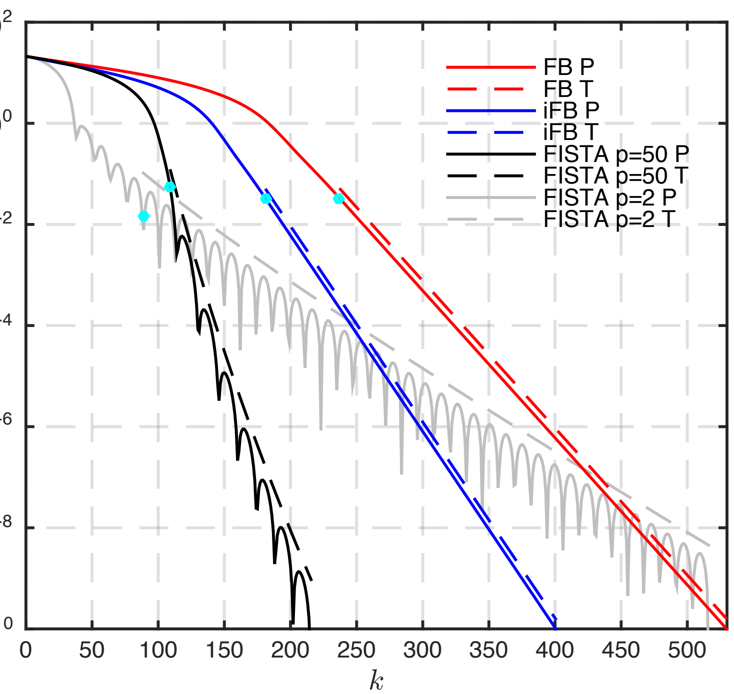
(ii) $R = \|\cdot\|_*$ *: let $r = \mathrm{rank}(x^\star)$, $x^\star \in \mathbb{R}^{n\times n}$. If $m \geq \beta r(6n - 5r)$ for some $\beta > 1$, then (1) is in force w.o.p..*

$$m \geq C \dim(T_{x^\star})\mathrm{polylog}(n)$$

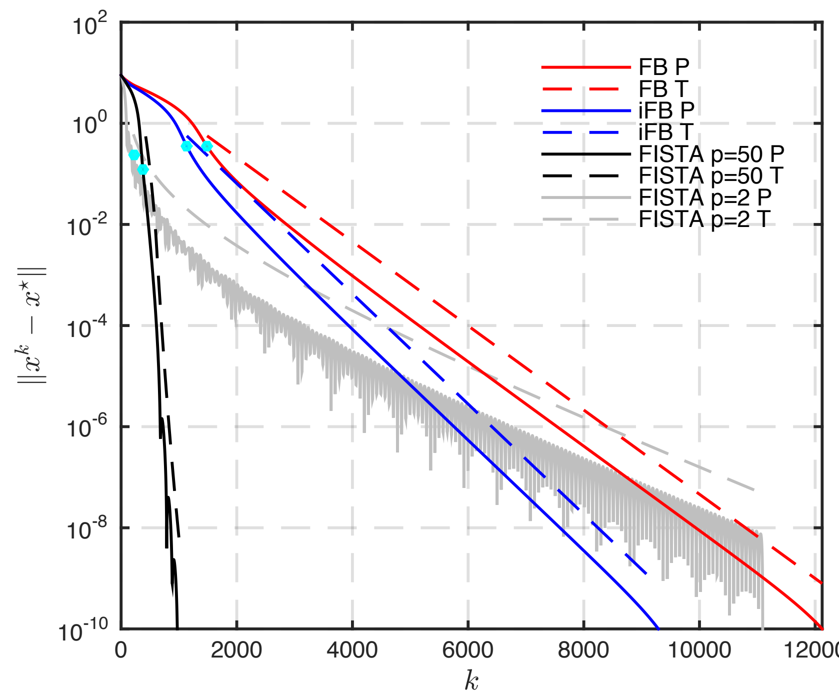
Stylized applications

$$\min_{x\in\mathbb{R}^n} \frac{1}{2}\|y - \mathrm{A}x\|_2^2 + \lambda R(x)$$

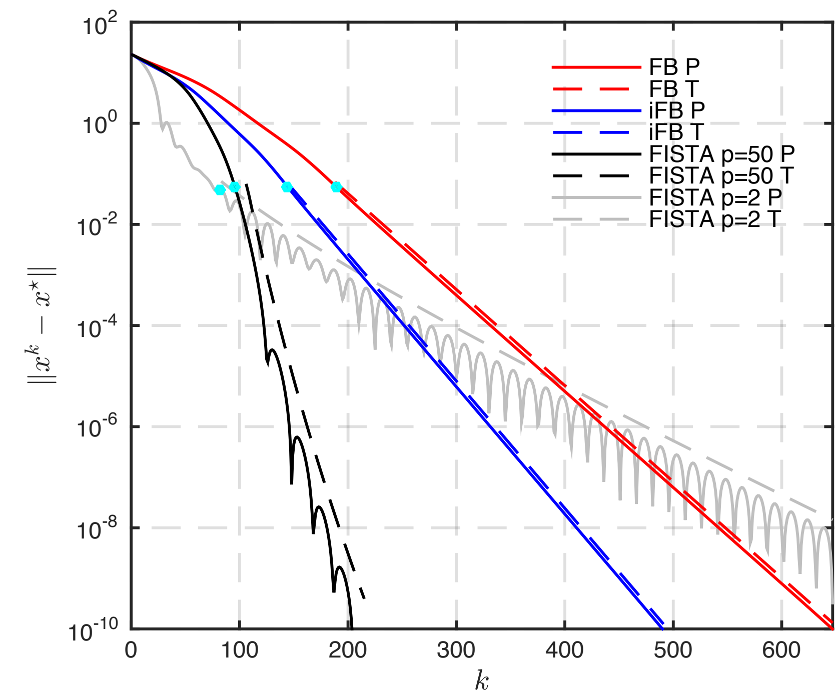
ℓ_1



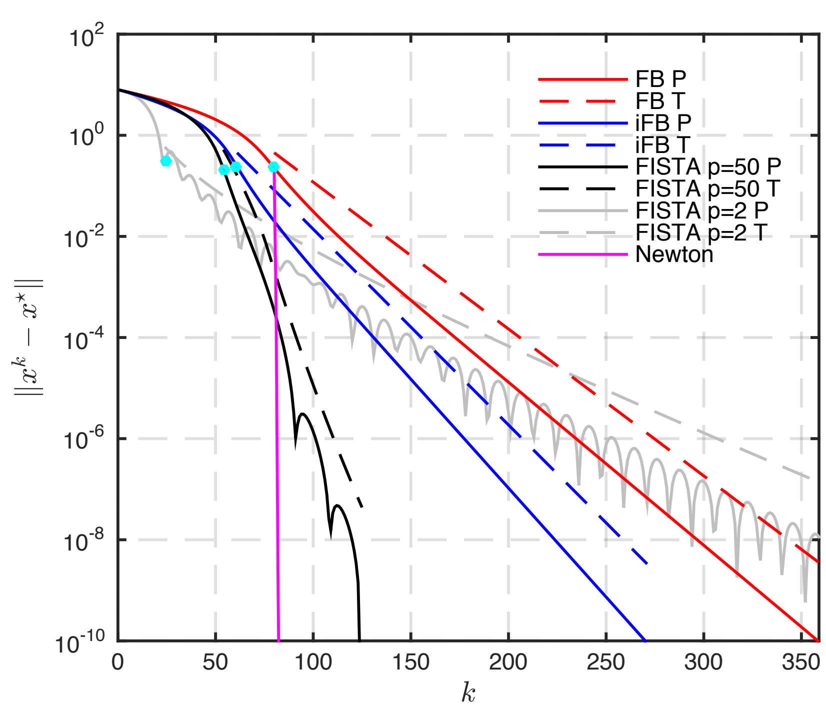
ℓ_∞



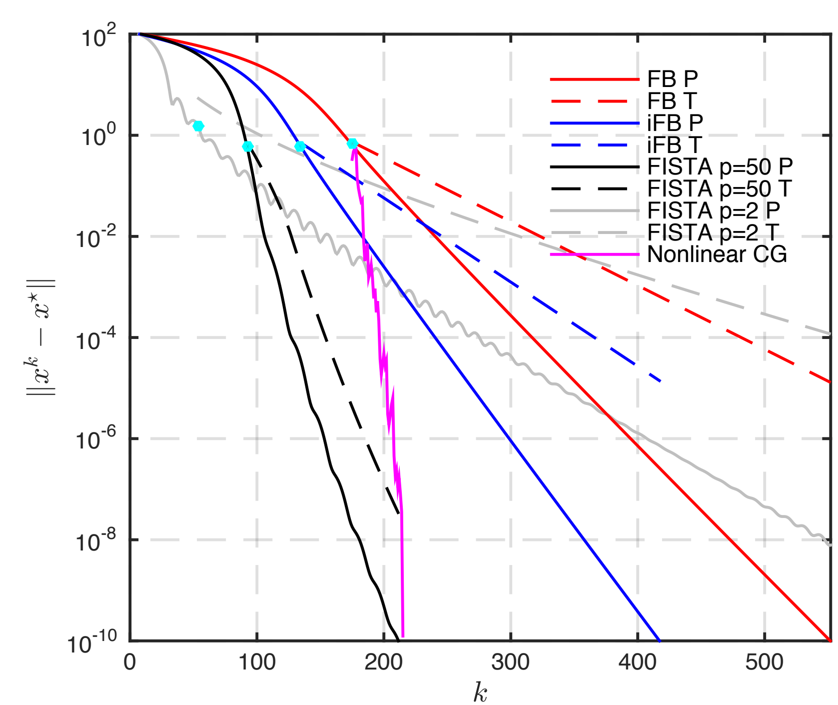
TV



$\ell_1 - \ell_2$



Nuclear norm



Outline

- Partial smoothness.
- Inertial Forward-Backward.
- Douglas-Rachford.
- Conclusion and future work.

Douglas-Rachford

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

(**A**.1) F and $G \in \Gamma_0(\mathbb{R}^n)$, $\mathrm{ri}\,(\mathrm{dom}(F)) \cap \mathrm{ri}\,(\mathrm{dom}(G)) \neq \emptyset$.

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

Primal form
$$\begin{cases} v_{k+1} = \mathrm{prox}_{\gamma G}\left(2x_k - z_k\right), \\ z_{k+1} = (1-\lambda_k)z_k + \lambda_k\left(z_k + v_{k+1} - x_k\right), \\ x_{k+1} = \mathrm{prox}_{\gamma F}\, z_{k+1}, \end{cases}$$

$\gamma > 0$

$\lambda_k \in]0,2], \sum_{k\in\mathbb{N}} \lambda_k(2-\lambda_k) = +\infty$

A primal-dual form
$$\begin{cases} u_{k+1} = \mathrm{prox}_{G^*/\gamma}\left(\left(2x_k - z_k\right)/\gamma\right), \\ z_{k+1} = (1-\lambda_k)z_k + \lambda_k\left(x_k - \gamma u_{k+1}\right), \\ x_{k+1} = \mathrm{prox}_{\gamma F}\left(z_{k+1}\right). \end{cases}$$

Identification and local linear convergence

Theorem *Suppose that DR is used to create a sequence (x_k, u_k) which converges to a primal-dual Kuhn-Tucker pair $(x^\star, u^\star)$ such that $F \in \mathrm{PS}_{x^\star}(\mathcal{M}^F_{x^\star})$ and $G^* \in \mathrm{PS}_{u^\star}(\mathcal{M}^{G^*}_{u^\star})$, and*

$$-u^\star \in \mathrm{ri}\,(\partial F(x^\star)) \quad \text{and} \quad x^\star \in \mathrm{ri}\,(\partial G^*(u^\star))\,.$$

(1) The DR scheme has the finite activity identification property, i.e. $\forall k$ large enough, $(x_k, u_k) \in \mathcal{M}^F_{x^\star} \times \mathcal{M}^{G^*}_{u^\star}$.

(2) Suppose furthermore that $\lambda_k \equiv 1$ and F is locally polyhedral around $x^\star$. Denote $d_k = \begin{pmatrix} \gamma\,(u_k - u^\star) \\ x_{k-1} - x^\star \end{pmatrix}$. Then,

(i) Q-linear convergence *: given any ρ such that $1 > \rho > \sin\theta_F(T^F_{x^\star}, T^{G^*}_{u^\star})$, we have*

$$\|d_{k+1}\| \le \rho\,\|d_k\| \quad \text{and} \quad \|z_k - z^\star\| = O(\rho^k)\,.$$

(ii) R-linear convergence *: if $\mathcal{M}_{x^\star}$ is affine/linear, then*

$$\|d_{k+1}\| \le \sin\theta_F(T^F_{x^\star}, T^{G^*}_{u^\star})\,\|d_{k+1}\| \quad \text{and} \quad \|z_k - z^\star\| = O\left(\sin^k\theta_F(T^F_{x^\star}, T^{G^*}_{u^\star})\right)\,.$$

Identification and local linear convergence

Theorem *Suppose that DR is used to create a sequence (x_k, u_k) which converges to a primal-dual Kuhn-Tucker pair $(x^\star, u^\star)$ such that $F \in \mathrm{PS}_{x^\star}(\mathcal{M}^F_{x^\star})$ and $G^* \in \mathrm{PS}_{u^\star}(\mathcal{M}^{G^*}_{u^\star})$, and*

$$-u^\star \in \mathrm{ri}\,(\partial F(x^\star)) \quad \text{and} \quad x^\star \in \mathrm{ri}\,(\partial G^*(u^\star))\,.$$

(1) The DR scheme has the finite activity identification property, i.e. $\forall k$ large enough, $(x_k, u_k) \in \mathcal{M}^F_{x^\star} \times \mathcal{M}^{G^*}_{u^\star}$.

(2) Suppose furthermore that $\lambda_k \equiv 1$ and F is locally polyhedral around $x^\star$. Denote $d_k = \begin{pmatrix} \gamma\,(u_k - u^\star) \\ x_{k-1} - x^\star \end{pmatrix}$. Then,

(i) Q-linear convergence *: given any ρ such that $1 > \rho > \sin\theta_F(T^F_{x^\star}, T^{G^*}_{u^\star})$, we have*

$$\|d_{k+1}\| \le \rho\,\|d_k\| \quad \text{and} \quad \|z_k - z^\star\| = O(\rho^k)\,.$$

(ii) R-linear convergence *: if $\mathcal{M}_{x^\star}$ is affine/linear, then*

$$\|d_{k+1}\| \le \sin\theta_F(T^F_{x^\star}, T^{G^*}_{u^\star})\,\|d_{k+1}\| \quad \text{and} \quad \|z_k - z^\star\| = O\left(\sin^k\theta_F(T^F_{x^\star}, T^{G^*}_{u^\star})\right)\,.$$

The Friedrichs angle $\theta_F(U, V)$ is the $\dim(U \cap V) + 1$ principal angle [Bauschke et al. 2014].
and $\theta_F(U, V) > 0$

Identification and local linear convergence

- If G^* is also locally polyhedral at $x^\star$: the rate estimate is optimal.
- Encompasses some previous results [Demanet and Zhang 2013], [Bauschke et al. 2013, 2014] (in finite dimension), [Boley et al. 2014].
- Extends readily to the case of more than two functions with the product space trick [Liang, Fadili, Peyré and Luke 2015].
- Extends easily to ADMM (DR on the dual).

Affine constrained problems

$$\min_{x \in \mathbb{R}^n} G(x) \quad \text{s.t.} \quad y = \mathrm{A}x \qquad (\mathrm{BP}_{\mathrm{R}})$$

Theorem *Suppose that DR is used to create a sequence* (x_k, u_k) *that converges to* $(x^\star, u^\star)$ *such that* $G^* \in \mathrm{PS}_{u^\star}(\mathcal{M}_{u^\star}^{G^*})$ *and*

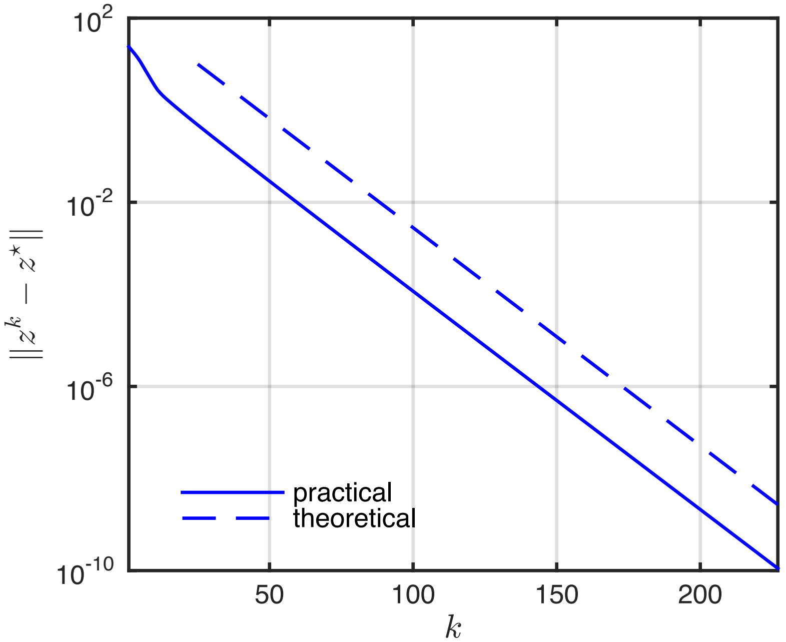
$$x^\star \in \left(\mathrm{A}^+ y + \ker(\mathrm{A})\right) \cap \mathrm{ri}\,(\partial G^*(u^\star)) \ . \qquad (1)$$

Then, $u_k \in \mathcal{M}_{u^\star}^{G^*}$ *for* k *large enough, and DR converges locally linearly with rate given by* $\cos\theta_F(\ker(A), S_{u^\star}^{G^*})$, $S_{u^\star}^{G^*} = {T_{u^\star}^{G^*}}^\perp$.

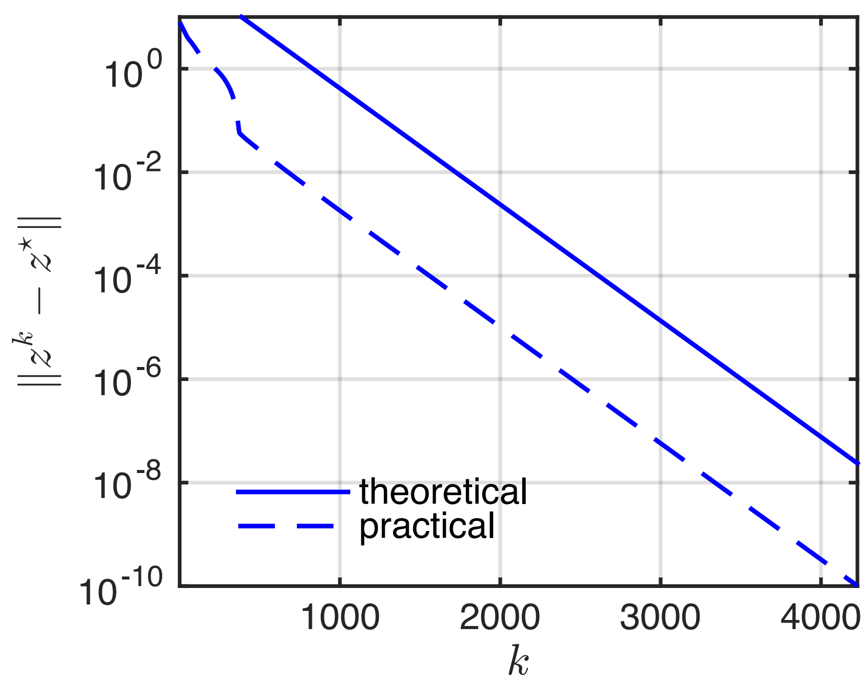
Stylized applications

$$\min_{x \in \mathbb{R}^n} G(x) \quad \text{s.t.} \quad y = Ax \qquad (\mathrm{BP_R})$$

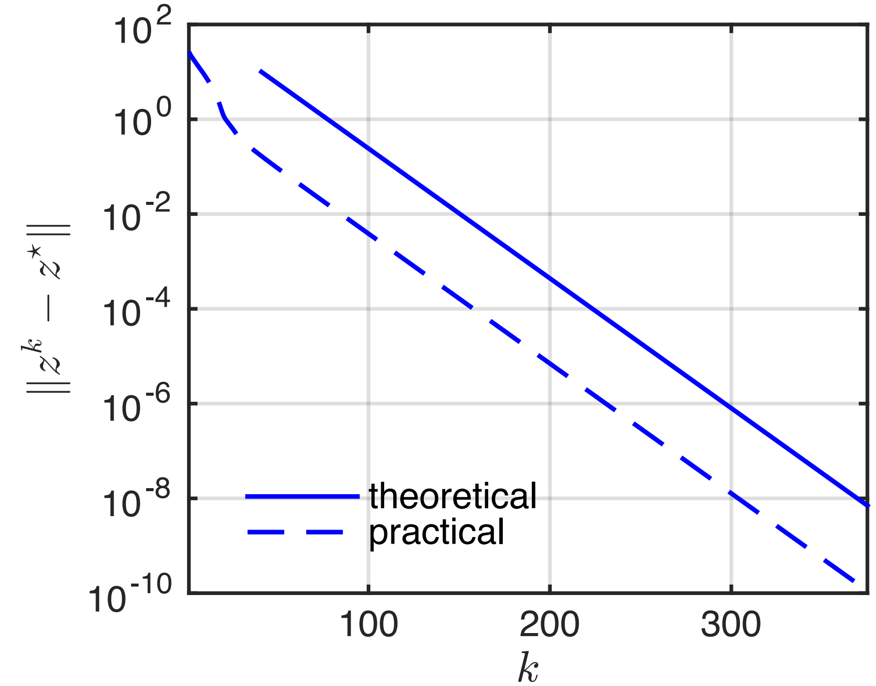
ℓ_1



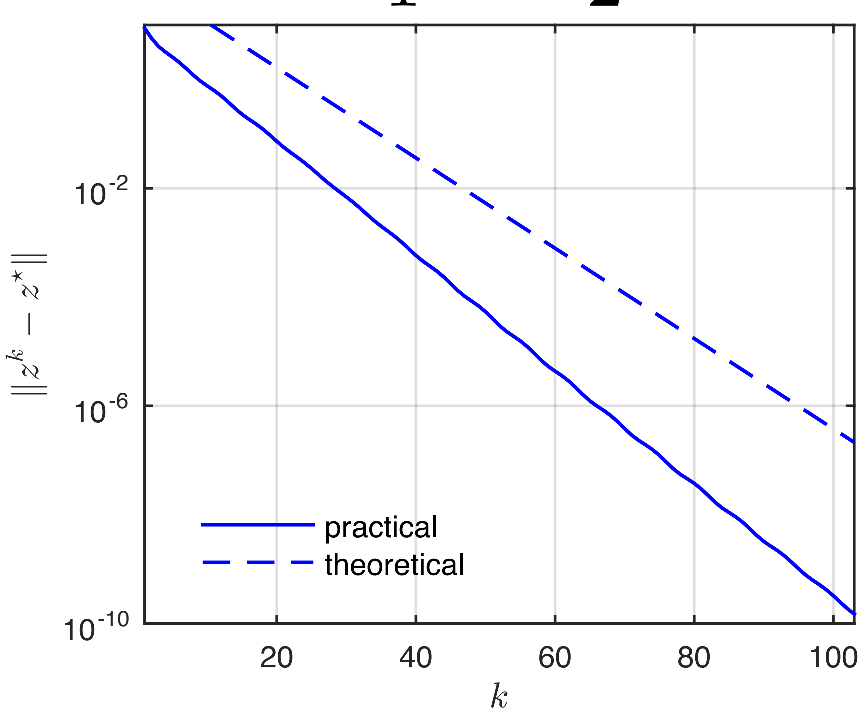
ℓ_∞



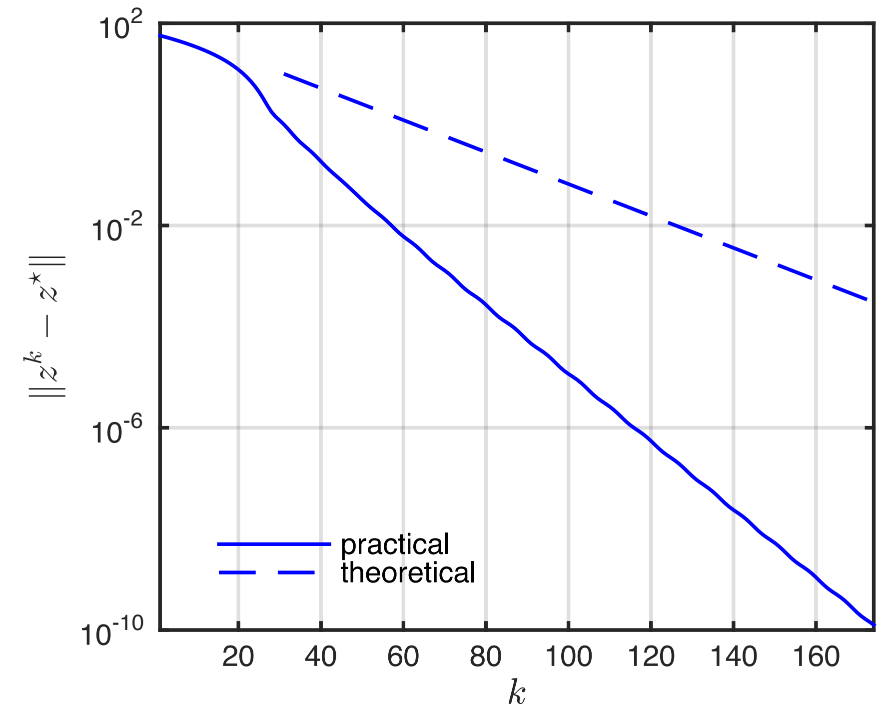
TV



$\ell_1 - \ell_2$



Nuclear norm



Outline

- Partial smoothness.
- Inertial Forward-Backward.
- Douglas-Rachford.
- Conclusion and future work.

Take away messages

- Finite activity identification and local linear convergence of proximal splitting algorithms.
- Explains the behaviour typically observed in many applications.
- The key: partial smoothness (a powerful framework for local convergence analysis).
- Many other splitting algorithms.

- Beyond convexity.
- Beyond non-degeneracy (polyhedral case and stratification).
- Infinite dimensional case.

Preprints on arxiv and papers on

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

Thanks
Any questions ?