

# Geometry of First-order Proximal Splitting Methods and Beyond

Activity identification, local linear convergence, adaptive acceleration

---

**Jingwei Liang**

Department of Applied Mathematics and Theoretical Physics

Joint work with: Clarice Poon (Bath)

Carola Schönlieb (Cambridge)

Jalal Fadili (CNRS, ENSICAEN)

Gabriel Peyré (CNRS, ENS-Paris)

Russell Luke (Göttingen)

Adrian Lewis (Cornell)

## Table of contents

- 1 First-order Proximal Splitting Methods
- 2 Geometry of Non-smooth Regularisation
- 3 Local Analysis of FoM
- 4 Numerical Experiments
- 5 Conclusions

## Sparse logistic regression [Friedman et al, 2001]

$(z_i, y_i) \in \mathbb{R}^n \times \{\pm 1\}, i = 1, \dots, m,$

$$\min_{(b,x) \in \mathbb{R} \times \mathbb{R}^n} \mu \|x\|_1 + \frac{1}{m} \sum_{i=1}^m f(\langle x, z_i \rangle + b, y_i),$$

where  $f(u_i, y_i) = \log(1 + e^{-u_i y_i})$ .

$$\|x\|_1 = \sum_{j=1}^n |x_j|.$$

### TV based Image deblur [Rudin et al, 1992]

$$w = Hx_{\text{ob}} + \omega,$$

where  $H \in \mathbb{R}^{m \times n}$  is blur kernel,  $\omega \in \mathbb{R}^m$  is additive noise.

$$\text{TV}(x) = \|\nabla x\|_1$$

$x_{\text{ob}}$

$w$

recovered  $x$

### Principal component pursuit [Candès et al, 2011]

$$w = x_{ob,I} + x_{ob,S} + \omega,$$

$x_{ob,I} \in \mathbb{R}^{m \times n}$  is low-rank,  $x_{ob,S} \in \mathbb{R}^{m \times n}$  is sparse and  $\omega \in \mathbb{R}^{m \times n}$  is noise.

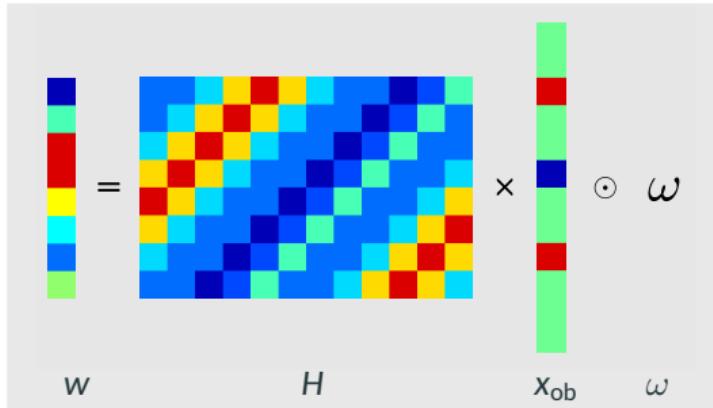
$$\|x\|_* = \sum_{j=1}^{\text{rank}(x)} \sigma_j(x).$$

$w$

$x_I$

$x_S$

## Example: inverse problems



**Forward model:**

$$w = (Hx_{\text{ob}}) \odot \omega.$$

**Goal:** recover  $x_{\text{ob}}$

**Challenge:** ill-posed

**Hope:** prior knowledge of  $x_{\text{ob}}$

- Regularisation: promoting low-complexity structure to the solution...
- Examples:

**Sparsity**  $\ell_1$ -norm,  $\ell_{1,2}$ -norm,  $\ell_p$ -norm,  $\ell_0$  pseudo-norm

**Analysis sparsity** total variation, wavelet, dictionary...

**Low-rank** nuclear norm, rank function

**Constraints** simplex, non-negativity...

**Nerual networks** CNN...

## Non-smooth optimisation problem

$$\min_{x \in \mathbb{R}^n} \left\{ \Phi(x) \stackrel{\text{def}}{=} F(x) + \sum_{i=1}^r R_i(x) \right\}.$$

$F$ : data fidelity term...

$R_i$ : non-smooth regularisation terms...

### Image deblur

$$\min_{x \in \mathbb{R}^n} \mu \|\nabla x\|_1 + \frac{1}{2} \|Hx - w\|^2.$$

### Sparse logistic regression

$$\min_{x \in \mathbb{R}^n} \mu \|x\|_1 + \frac{1}{m} \sum_{i=1}^m f(\langle x, z_i \rangle + b, y_i).$$

### Principal component pursuit

$$\min_{x_l, x_s \in \mathbb{R}^{m \times n}} \mu_1 \|x_s\|_1 + \mu_2 \|x_l\|_* + \frac{1}{2} \|w - x_l - x_s\|^2.$$

**Non-smooth, (non-convex), composite, high dimension**

### Gradient descent

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

where  $F$  is convex smooth differentiable with  $\nabla F$  being  $L$ -Lipschitz

$$x_{k+1} = x_k - \gamma_k \nabla F(x_k), \quad \gamma_k \in ]0, 2/L[.$$

### Proximal point algorithm [Rockafellar, 1976]

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

with  $R$  being only convex. Define “proximity operator” by

$$\text{prox}_{\gamma R}(v) \stackrel{\text{def}}{=} \operatorname{argmin}_{x \in \mathbb{R}^n} \gamma R(x) + \frac{1}{2} \|x - v\|^2.$$

Then

$$x_{k+1} = \text{prox}_{\gamma_k R}(x_k), \quad \gamma_k > 0.$$

FoM [Bauschke and Combettes, 2011, Raguet et al, 2011]...

$F + R$  Forward–Backward splitting (FB), inertial FB, Nesterov, FISTA

$F = \frac{1}{m} \sum_i f_i$ : stochastic gradient methods

$R_1 + R_2$  Douglas–Rachford splitting

$F + R(\mathcal{W}\cdot)$  Class of Primal–Dual splitting

Alternating Direction Method of Multipliers (ADMM)

$F + \sum_{i=1}^r R_i$  Three-operator splitting ( $r = 2$ )

Forward–Douglas–Rachford ( $r = 2, R_2 = \iota_{\mathcal{V}}(\cdot)$ )

Generalized Forward–Backward splitting ( $r \geq 2$ )

- ...

Origins from numerical PDE dates back to 1950s, now becomes ubiquitous in signal/image processing, inverse problems, data science, statistics and machine learning, game theory...

# Outline

1 First-order Proximal Splitting Methods

2 Geometry of Non-smooth Regularisation

3 Local Analysis of FoM

4 Numerical Experiments

5 Conclusions

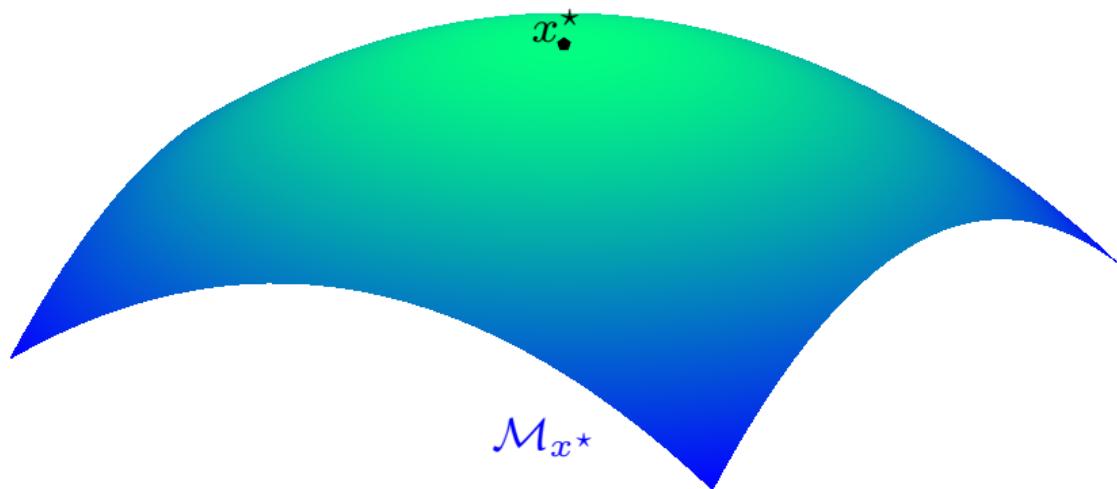
When regularises

**Goal:** find  $x^*$  which has low-complexity, e.g.  $x^* \in \mathcal{M}_{x^*}$ .

# What happens when regularises...

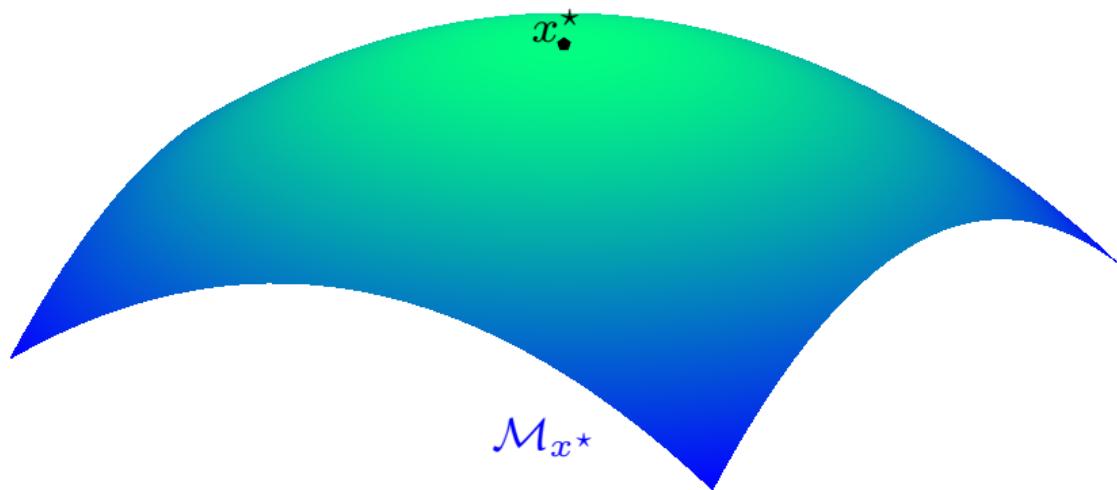
When regularises

**Goal:** find  $x^*$  which has low-complexity, e.g.  $x^* \in \mathcal{M}_{x^*}$ .



When regularises

**Goal:** find  $x^*$  which has low-complexity, e.g.  $x^* \in \mathcal{M}_{x^*}$ .



**FoM:** generates  $x_k$  that converges to  $x^*$ .

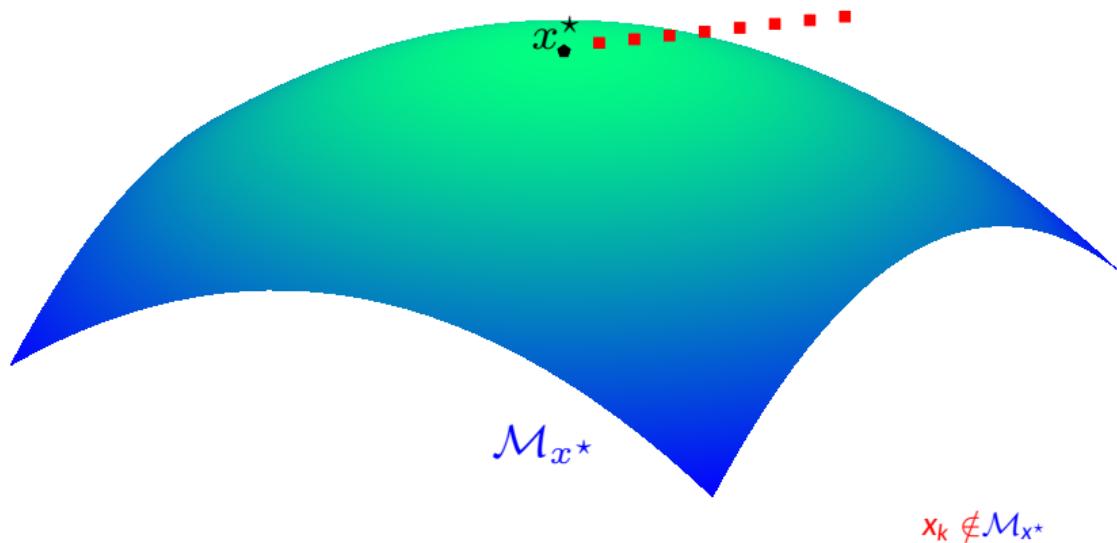
How about

$$x_k ? \mathcal{M}_{x^*}$$

# What happens when regularises...

When regularises

**Goal:** find  $x^*$  which has low-complexity, e.g.  $x^* \in \mathcal{M}_{x^*}$ .



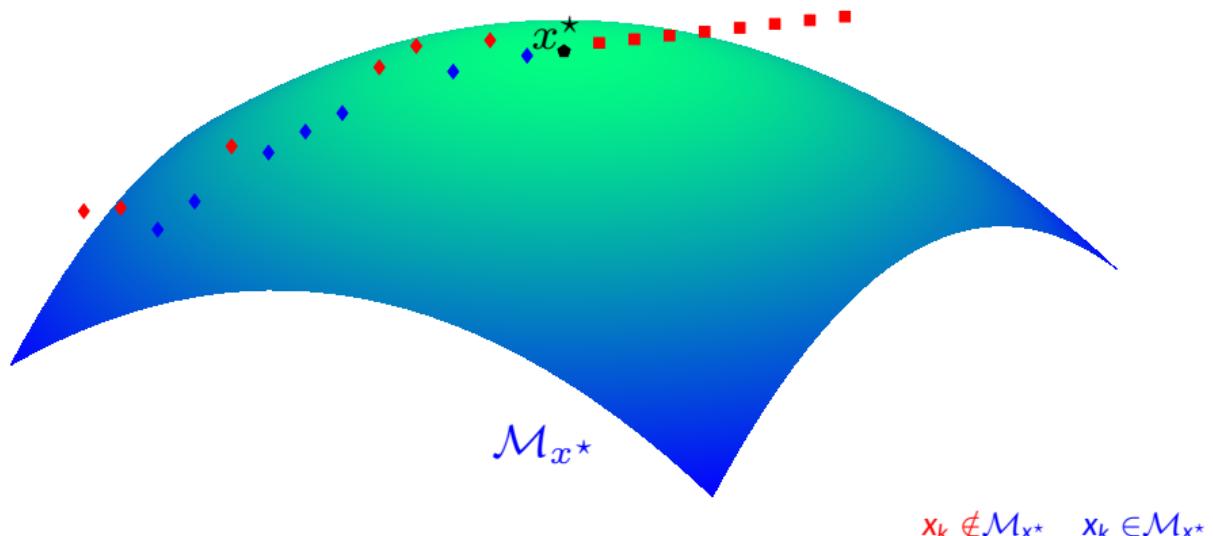
**FoM:** generates  $x_k$  that converges to  $x^*$ .

How about

$$x_k ? \mathcal{M}_{x^*}$$

When regularises

**Goal:** find  $x^*$  which has low-complexity, e.g.  $x^* \in \mathcal{M}_{x^*}$ .

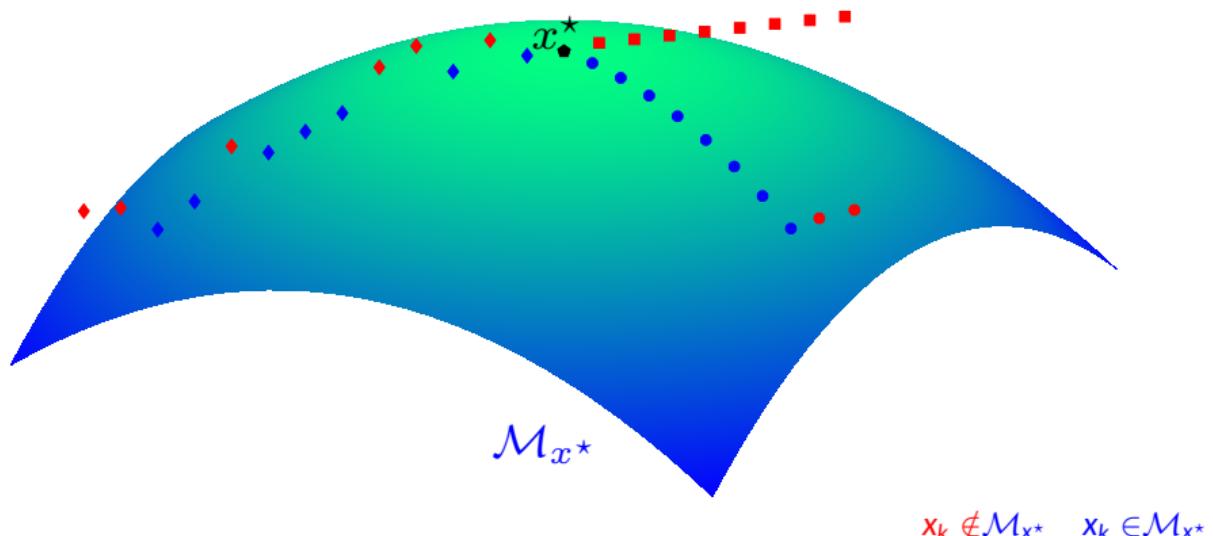


**FoM:** generates  $x_k$  that converges to  $x^*$ .  
How about

$x_k ? \mathcal{M}_{x^*}$

When regularises

**Goal:** find  $x^*$  which has low-complexity, e.g.  $x^* \in \mathcal{M}_{x^*}$ .



**FoM:** generates  $x_k$  that converges to  $x^*$ .

How about

$x_k ? \mathcal{M}_{x^*}$

Consider low-rank recovery

$$\min_{x \in \mathbb{R}^{n \times n}} \mu \|x\|_* + \frac{1}{2} \|Hx - w\|^2.$$

“Activity” of  $x_k$ : let  $q = \text{rank}(x^*)$ ,

- $\text{rank}(x) \in ]q, n[ : k \leq K$
- $\text{rank}(x) = q : k \geq K$

Rate of convergence:

- Sub-linear:  $k \leq K$
- Linear:  $k \geq K$

Phase transition of convergence rate coincides with that of “activity”.

- What are the possible mechanisms underlying the identification of “activity”?
- How fast is the global sub-linear convergence rate?
- How to explain the local linear convergence?
- What is the relation between local linear convergence and identification of “activity”?
- Can we accelerate
  - the local convergence rate?
  - higher-order methods?

Specific problems (e.g.  $\ell_1$ -norm)

Specific algorithms (e.g. FB)

Cannot explain “phase transition”



A unified framework is missing!

- Global  $o(1/\sqrt{k})$  sub-linear convergence rate for  $\|x_k - x_{k-1}\|$ .
- A unified framework for:
  - Finite time activity identification
  - Local linear convergence
  - Relation between the two...
- Covers both **deterministic** and **stochastic** setting.
- Geometry based acceleration which bridges 1st-order and 2nd-order.

# Outline

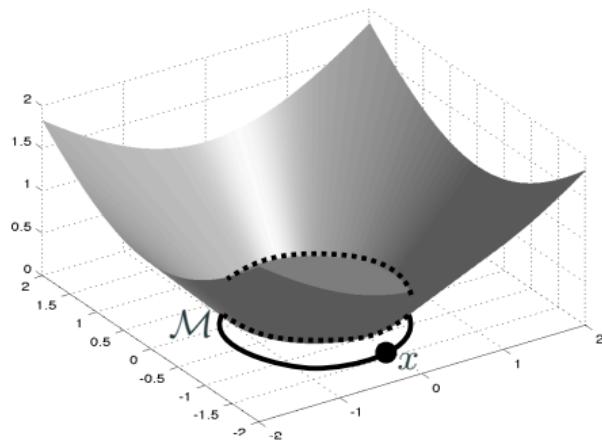
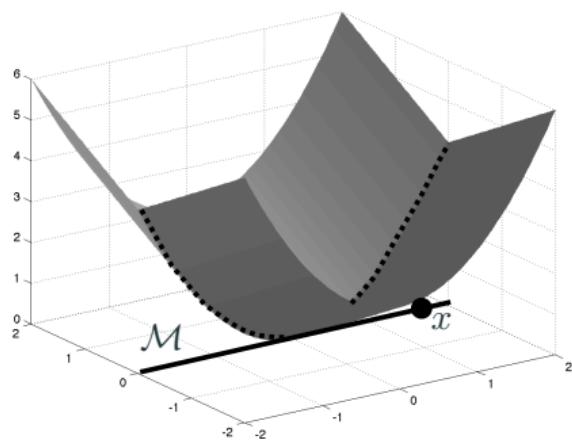
1 First-order Proximal Splitting Methods

2 Geometry of Non-smooth Regularisation

3 Local Analysis of FoM

4 Numerical Experiments

5 Conclusions



- A partly smooth function behaves smoothly along a manifold  $\mathcal{M}$ , and sharply normal to it.
- The behaviours of the function and its minimizers depend essentially on their restrictions to the manifold.
- It offers a powerful framework for algorithmic and sensitivity analysis.

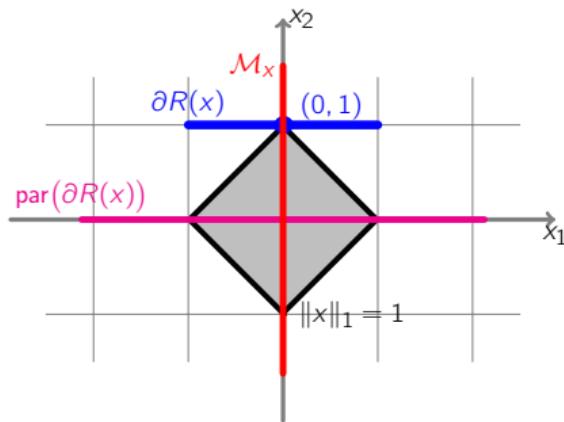
## Partly smooth function [Lewis, 2003]

Let  $R \in \Gamma_0 \mathbb{R}^n$ ,  $R$  is *partly smooth at  $x$  relative to a set  $\mathcal{M}_x$*  containing  $x$  if  $\partial R(x) \neq \emptyset$  and

**Smoothness:**  $\mathcal{M}_x$  is a  $C^2$ -manifold,  $R|_{\mathcal{M}_x}$  is  $C^2$  near  $x$ .

**Sharpness:** Tangent space  $T_{\mathcal{M}_x}(x)$  is  $T_x \stackrel{\text{def}}{=} \text{par}(\partial R(x))^\perp$ .

**Continuity:**  $\partial R : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$  is continuous along  $\mathcal{M}_x$  near  $x$ .



$\text{par}(C)$ : sub-space parallel to  $C$ , where  $C \subset \mathbb{R}^n$  is a non-empty convex set.

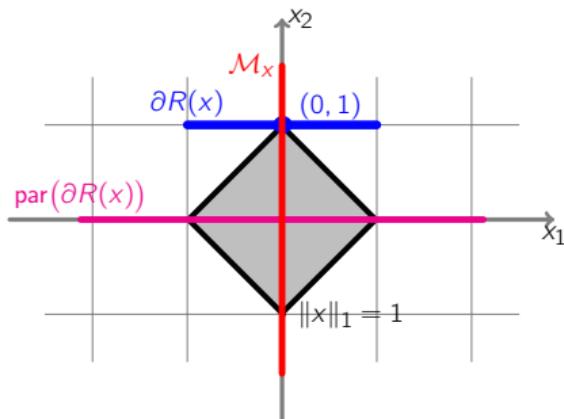
## Partly smooth function [Lewis, 2003]

Let  $R \in \Gamma_0 \mathbb{R}^n$ ,  $R$  is *partly smooth at  $x$  relative to a set  $\mathcal{M}_x$*  containing  $x$  if  $\partial R(x) \neq \emptyset$  and

**Smoothness:**  $\mathcal{M}_x$  is a  $C^2$ -manifold,  $R|_{\mathcal{M}_x}$  is  $C^2$  near  $x$ .

**Sharpness:** Tangent space  $T_{\mathcal{M}_x}(x)$  is  $T_x \stackrel{\text{def}}{=} \text{par}(\partial R(x))^\perp$ .

**Continuity:**  $\partial R : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$  is continuous along  $\mathcal{M}_x$  near  $x$ .



### Calculus rules:

- Sum and composition
- Smooth perturbation
- Spectral lifting

$\text{par}(C)$ : sub-space parallel to  $C$ , where  $C \subset \mathbb{R}^n$  is a non-empty convex set.

## Partly smooth function [Lewis, 2003]

Let  $R \in \Gamma_0 \mathbb{R}^n$ ,  $R$  is *partly smooth at  $x$  relative to a set  $\mathcal{M}_x$*  containing  $x$  if  $\partial R(x) \neq \emptyset$  and

**Smoothness:**  $\mathcal{M}_x$  is a  $C^2$ -manifold,  $R|_{\mathcal{M}_x}$  is  $C^2$  near  $x$ .

**Sharpness:** Tangent space  $T_{\mathcal{M}_x}(x)$  is  $T_x \stackrel{\text{def}}{=} \text{par}(\partial R(x))^\perp$ .

**Continuity:**  $\partial R : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$  is continuous along  $\mathcal{M}_x$  near  $x$ .

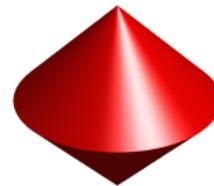
$\text{PSF}_x(\mathcal{M}_x)$ : partly smooth function at  $x$  relative to  $\mathcal{M}_x$



$\ell_1$ -norm



$\ell_\infty$ -norm



$\ell_{1,2}$ -norm



Nuclear norm

**Lemma ([L, Jalal & Peyré, 14])**

Let  $\mathcal{M}_x$  be a  $C^2$ -smooth manifold around  $x$ . Then for any  $x' \in \mathcal{M}_x$  near  $x$ , the projection  $\mathcal{P}_{\mathcal{M}_x}(x')$  is uniquely valued and  $C^1$  around  $x$ , and thus

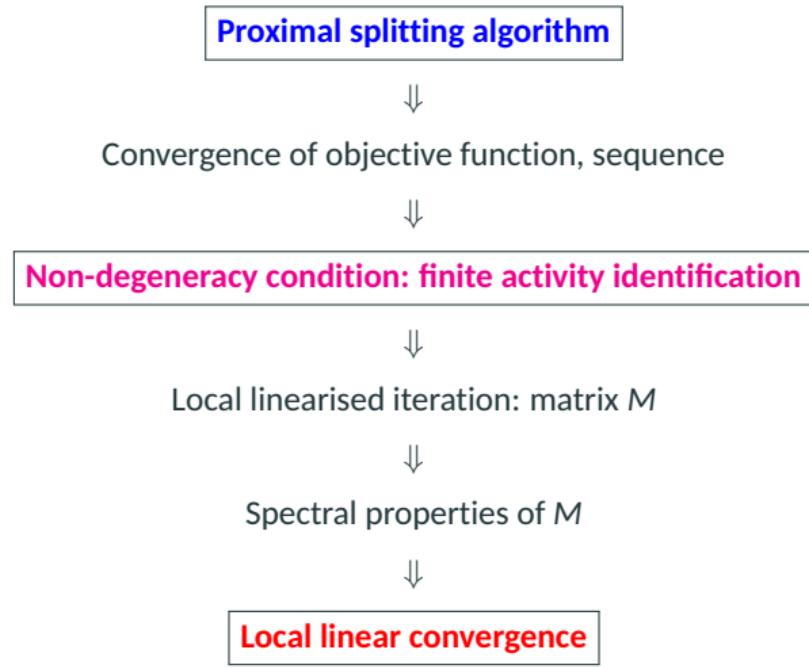
$$x' - x = \mathcal{P}_{T_{\mathcal{M}_x}(x)}(x' - x) + o(\|x' - x\|).$$

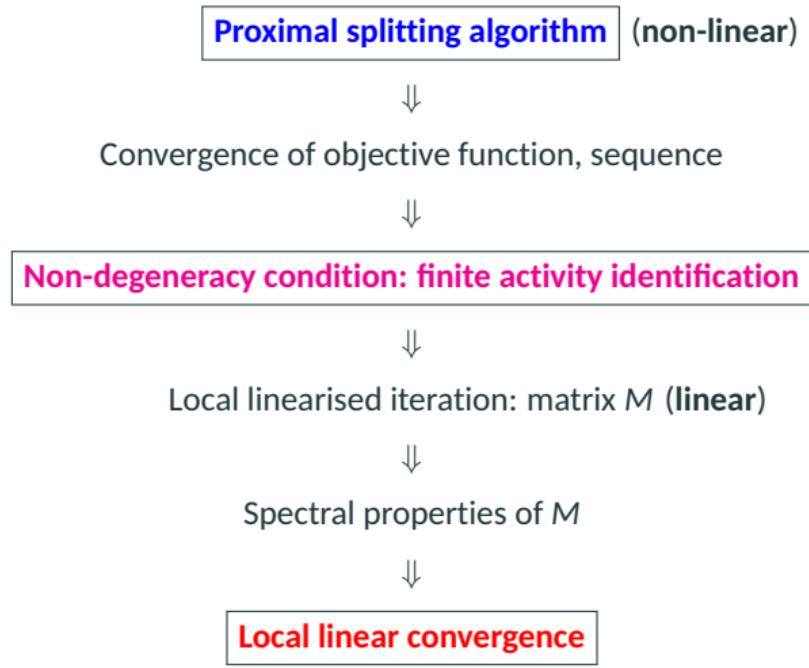
**Lemma ([Poon, L, 19])**

Let  $R \in \text{PSF}_{\bar{x}}(\mathcal{M}_{\bar{x}})$  and assume  $\bar{u} \in \text{ri}(\partial R(\bar{x}))$ . Suppose  $\exists \bar{z} \in \mathbb{R}^n$  such that  $\bar{x} = \text{prox}_R(\bar{z})$  and  $\bar{u} = \bar{z} - \bar{x}$ . Let  $\{z_k\}_{k \in \mathbb{N}}$  be a sequence such that  $z_k \rightarrow \bar{z}$  and  $x_k = \text{prox}_R(z_k) \rightarrow \bar{x}$ , then for all  $k$  large enough,

$$x_k - x_{k-1} = M_R(z_k - z_{k-1}) + o(\|z_k - z_{k-1}\|),$$

where  $M_R = \mathcal{P}_{T_{\bar{x}}}(\text{Id} + H_R)^{-1}\mathcal{P}_{T_{\bar{x}}}$  with  $H_R \stackrel{\text{def}}{=} \mathcal{P}_{T_{\bar{x}}}\nabla_{\mathcal{M}_{\bar{x}}}^2 R(\bar{x})\mathcal{P}_{T_{\bar{x}}}$ .





- Forward–Backward-type: FB, inertial FB, FISTA [[L, Jalal & Peyré, 14, 17, Poon, L & Schönlieb, 18](#)]
- Douglas–Rachford splitting, ADMM [[L, Jalal, Peyré & Luke, 15, L, Jalal & Peyré, 16](#)]
- Class of Primal–Dual splitting methods [[L, Jalal & Peyré, 18](#)]
- Forward–Douglas–Rachford splitting, Generalized Forward–Backward splitting, Three-operator splitting [[Molinari, L & Jalal, 18](#)]

- Forward–Backward-type: **FB**, inertial FB, FISTA [[L, Jalal & Peyré, 14, 17](#), [Poon, L & Schönlieb, 18](#)]
- Douglas–Rachford splitting, ADMM [[L, Jalal, Peyré & Luke, 15](#), [L, Jalal & Peyré, 16](#)]
- Class of Primal–Dual splitting methods [[L, Jalal & Peyré, 18](#)]
- Forward–Douglas–Rachford splitting, Generalized Forward–Backward splitting, Three-operator splitting [[Molinari, L & Jalal, 18](#)]

Recall the optimisation problem

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

## Inexact Forward–Backward (iFB)

Let  $\gamma_k \in ]0, 2/L[$  and  $\epsilon_k \in \mathbb{R}^n$ :

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

**Remark** For  $F = \frac{1}{m} \sum_i f_i$ :  $i_k \in \{1, \dots, m\}$

- Stochastic gradient descent (SGD):

$$\epsilon_k = \nabla f_{i_k}(x_k) - \nabla F(x_k).$$

- SAGA [Defazio et al, 14]:

$$\epsilon_k = \nabla f_{i_k}(x_k) - \nabla f_{i_k}(x_{k-1}) + \frac{1}{m} \sum_{j=1}^m \nabla f_{i_{k-j}}(x_{k-j}) - \nabla F(x_k).$$

- Prox-SVRG [Xiao & Zhang, 14]

$$\epsilon_k = \nabla f_{i_k}(x_k) - \nabla f_{i_k}(\tilde{x}_\ell) + \nabla F(\tilde{x}_\ell) - \nabla F(x_k).$$

Let  $x^* \in \text{Argmin}(\Phi)$  be a global minimiser.

### Convergence of iFB [L, Jalal & Peyré, 16, Poon, L & Schönlieb, 18]

Deterministic:  $x_k \rightarrow x^*$  if

$$\gamma_k \in ]0, 2/L[ \quad \text{and} \quad \sum_k \|\epsilon_k\| < +\infty.$$

Stochastic:  $x_k \rightarrow x^*$  almost surely if

$$\gamma_k \equiv \gamma \in ]0, 1/L[ \quad \text{and} \quad \sum_k \mathbb{E}[\|\epsilon_k\|^2] < +\infty.$$

### Remark

- The convergence of  $\Phi(x_k)$  follows that of  $x_k$
- SGD:

$$\mathbb{E}[\|\epsilon_k\|] \in ]0, +\infty[.$$

- SAGA/Prox-SVRG:

$$\mathbb{E}[\|\epsilon_k\|] \rightarrow 0.$$

Let  $x^* \in \text{Argmin}(\Phi)$ , then

$$0 \in \nabla F(x^*) + \partial R(x^*).$$

### Finite identification [L, Jalal & Peyré, 17, Poon, L & Schönlieb, 18]

Let the convergence of iFB hold. Suppose that  $R \in \text{PSF}_{x^*}(\mathcal{M}_{x^*})$ , and the non-degeneracy condition

$$0 \in \text{ri}(\nabla F(x^*) + \partial R(x^*)), \quad (\text{ND})$$

holds. Then, there exists a  $K \geq 0$  such that for all  $k \geq K$ :

$$x_k \in \mathcal{M}_{x^*}.$$

#### Remark

- A bound on  $K$  can be provided.
- Stochastic proximal gradient does **NOT** have finite identification.
- The identification of SAGA/Prox-SVRG is **almost surely**.

### Local linearisation [L, Jalal & Peyré, 17]

For the iFB iteration, suppose the Identification theorem holds. If  $F$  is locally  $C^2$  around  $x^*$ ,

$$\gamma_k \rightarrow \gamma \in ]0, 2/L[,$$

then for all  $k$  large enough, there exist a matrix  $M$  such that

$$x_{k+1} - x^* = M(x_k - x^*) + o(\|x_k - x^*\|) + \epsilon_k.$$

#### Remark

- $o(\|x_k - x^*\|)$  vanishes if  $R$  is locally polyhedral around  $x^*$ , and

$$\gamma_k \equiv \gamma.$$

- $M$  is similar to a symmetric positive semidefinite matrix.
- For stochastic algorithms, no linearisation is needed.

Restricted injectivity:  $\exists \alpha > 0$  such that  $\forall h \in T_{x^*}$ ,

$$\langle h, \nabla^2 F(x^*)h \rangle \geq \alpha \|h\|^2. \quad (\text{RI})$$

### Spectral radius of $M$ [L, Jalal & Peyré, 17]

For the matrix  $M$ , suppose that (RI) holds, then  $\rho(M) < 1$  as long as

$$\gamma \in ]0, 2/L[,$$

and  $\rho(M)$  can be given explicitly.

### Local linear convergence [L, Jalal & Peyré, 17]

Suppose iFB creates a sequence  $x_k \rightarrow x^* \in \text{Argmin}(\Phi)$  such that the Identification, Linearisation and Spectral radius theorems hold. Then given any  $\rho \in [\rho(M), 1[$ , there is  $K$  large enough such that for all  $k \geq K$ ,

$$\|x_k - x^*\| = O(\rho^k).$$

### Quadratic growth [L, Jalal & Peyré, 17]

Let  $x^* \in \text{Argmin}(\Phi)$  be a minimiser such that conditions (ND) and (RI) are fulfilled and  $R \in \text{PSF}_{x^*}(\mathcal{M}_{x^*})$ , then  $x^*$  is the unique minimiser of  $\Phi$  and there exist  $\alpha > 0$  and  $r > 0$  such that

$$\Phi(x) - \Phi(x^*) \geq \alpha \|x - x^*\|^2 : \forall x \text{ s.t. } \|x - x^*\| \leq r.$$

### Local linear convergence [Poon, L & Schönlieb, 18]

Suppose iFB creates a sequence  $x_k \rightarrow x^* \in \text{Argmin}(\Phi)$  such that the Identification, Linearisation and Spectral radius theorems hold. Then there exists  $\rho < 1$  such that for all  $k$  large enough,

$$\mathbb{E}[\|x_k - x^*\|] = O(\rho^k).$$

**Remark** The theoretical rate estimation of in general is not as tight as their deterministic counterparts.



**Local condition:** better Lipschitz constant along  $\mathcal{M}_{x^*}$

**Locally polyhedral:** finite termination if  $F$  is quadratic

**General manifold:** Newton-like, Conjugate gradient, Manifold based optimisation methods

# Outline

1 First-order Proximal Splitting Methods

2 Geometry of Non-smooth Regularisation

3 Local Analysis of FoM

4 Numerical Experiments

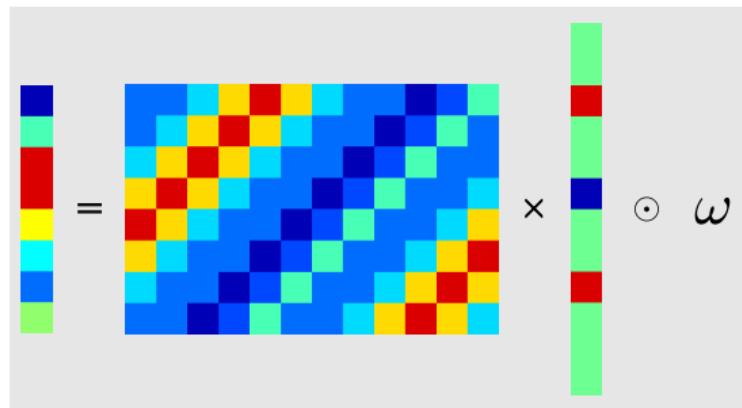
5 Conclusions

## Examples

**Sparse LR**  $(z_i, y_i) \in \mathbb{R}^n \times \{\pm 1\}$ ,  $m = 64$ ,  $n = 96$

$$\min_{(b,x) \in \mathbb{R} \times \mathbb{R}^n} \frac{1}{m} \sum_{i=1}^m f(\langle x, z_i \rangle + b, y_i) + \mu \|x\|_1,$$

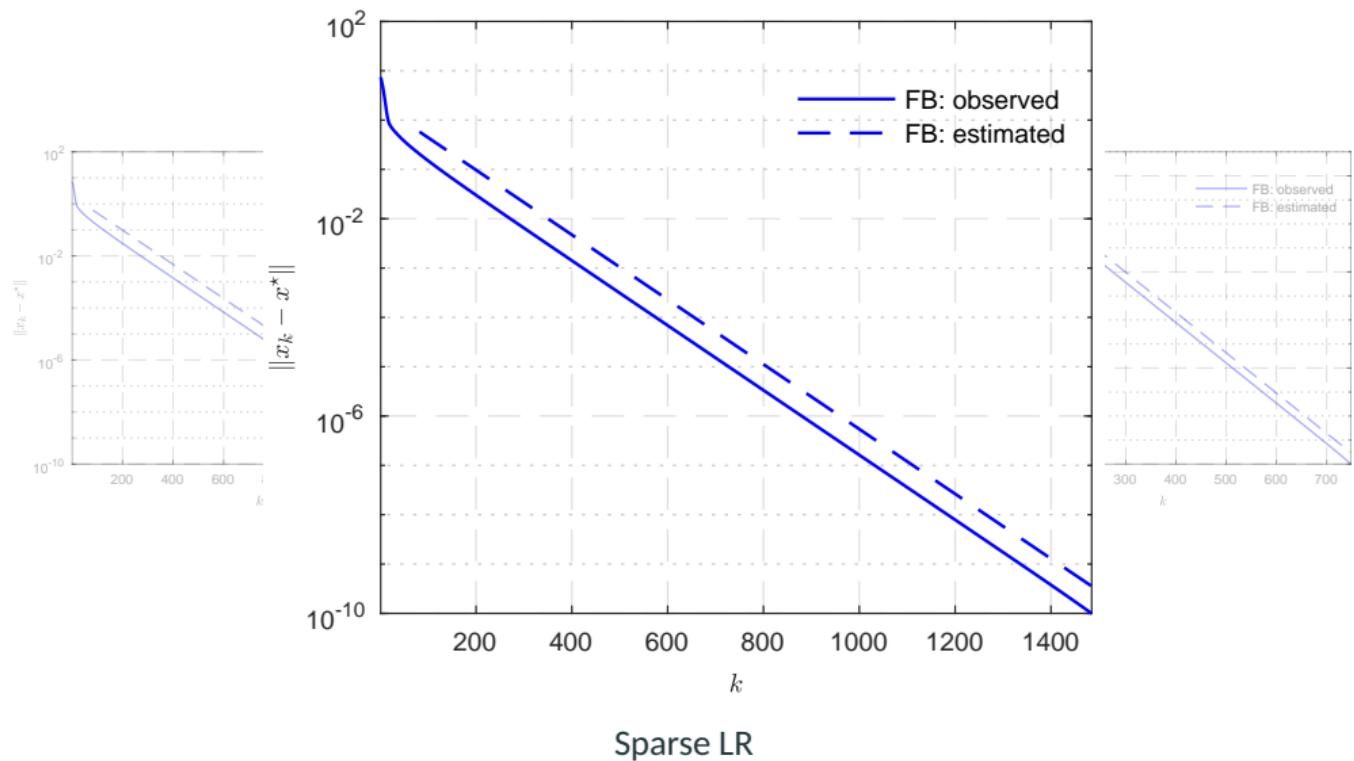
where  $f(w_i, y_i) = \log(1 + e^{-w_i y_i})$ .

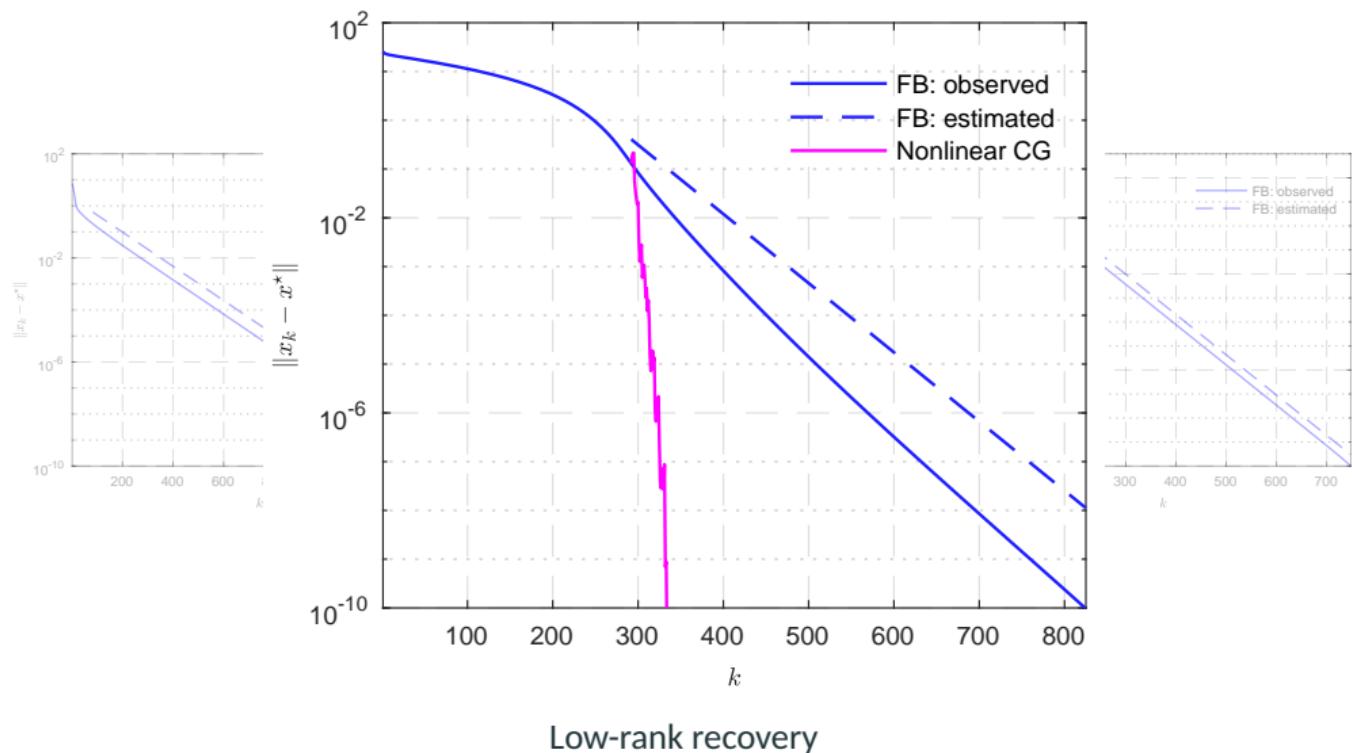


$$\min_x \mu R(x) + \frac{1}{2} \|Hx - w\|^2$$

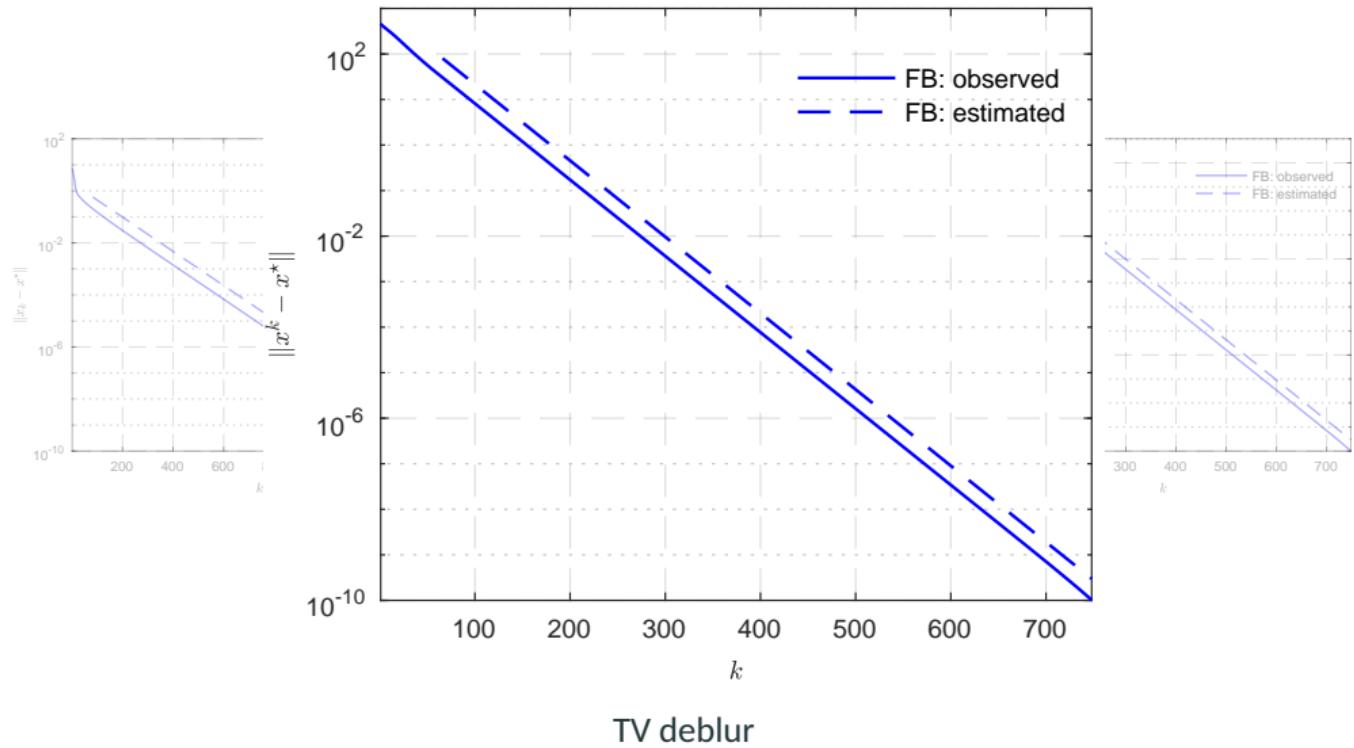


$$\min_x \mu \|\nabla x\|_1 + \frac{1}{2} \|Hx - w\|^2$$

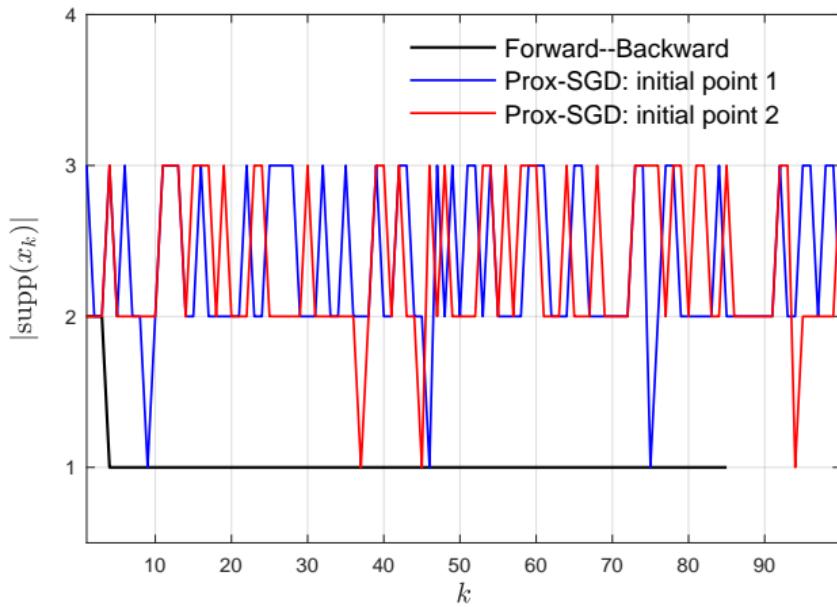


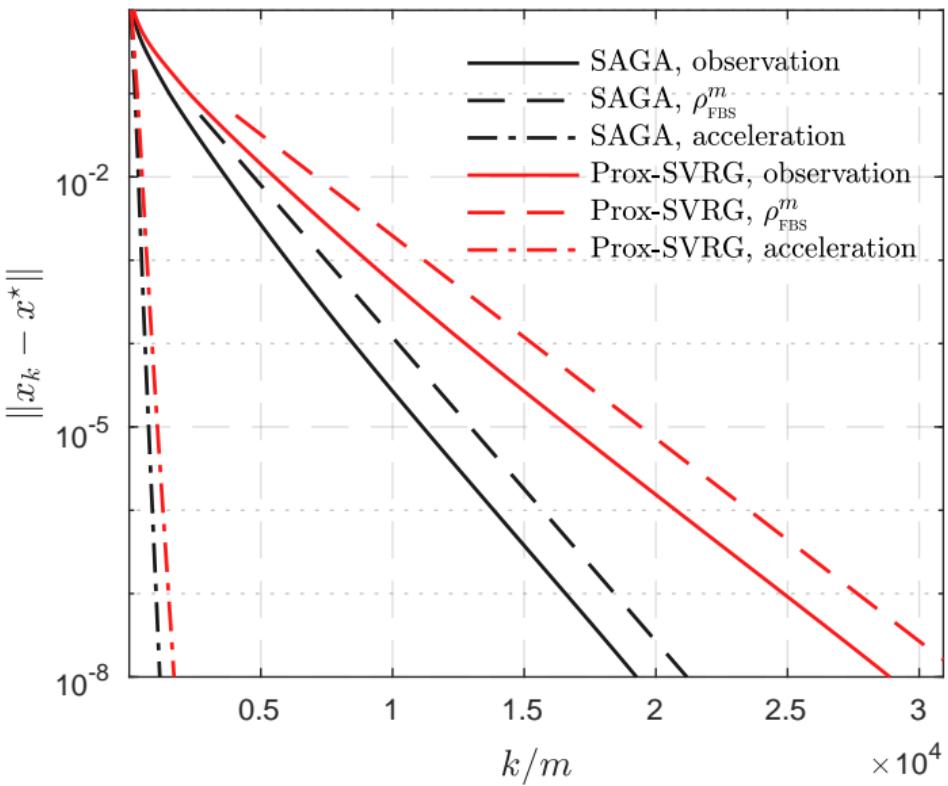


# Numerical result: deterministic



$$\min_{x \in \mathbb{R}^3} \frac{1}{3} \|x\|_1 + \frac{1}{3} \sum_{i=1}^3 \frac{1}{2} \|H_i x - b_i\|^2, \quad H = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \sqrt{2} & 0 \\ 0 & 0 & \sqrt{3} \end{bmatrix} \text{ and } b = \begin{pmatrix} 2 \\ \sqrt{2}/3 \\ \sqrt{3}/4 \end{pmatrix}.$$





# Outline

1 First-order Proximal Splitting Methods

2 Geometry of Non-smooth Regularisation

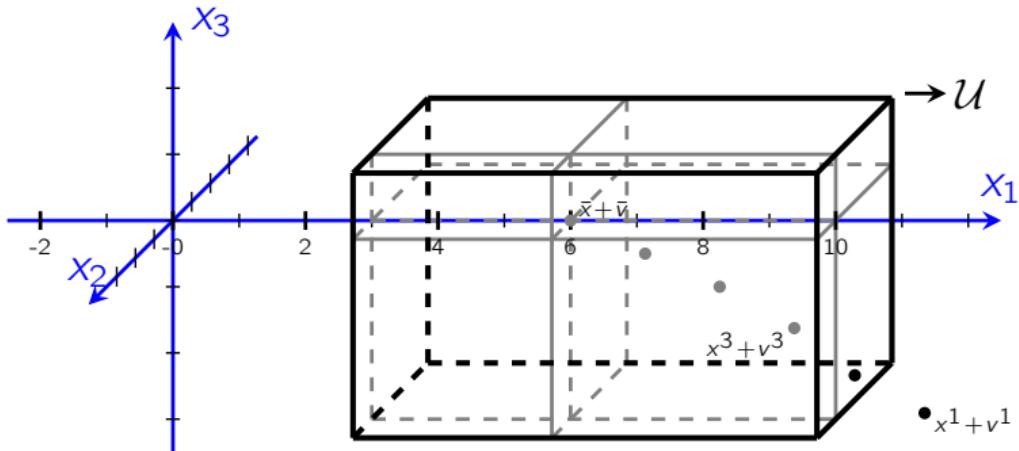
3 Local Analysis of FoM

4 Numerical Experiments

5 Conclusions

## Extensions

- Beyond non-degeneracy: enlarged manifold (with J. Fadili)
- Beyond optimization: set-valued operators (with A. Lewis)

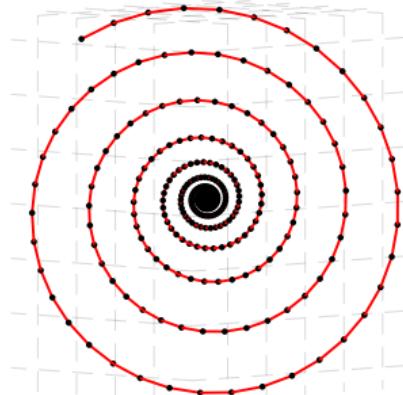
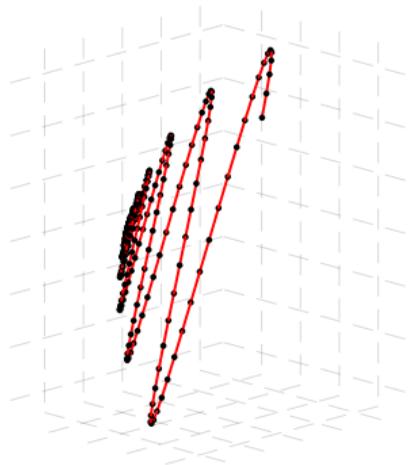


Let  $\bar{x} = (5; 0; 0)$ ,  $\mathcal{M}_x = [\mathbb{R}; 0; 0]$ ,  $A = \partial \|\cdot\|_1$  and  $\bar{v} \in \text{ri}(A(\bar{x}))$ :

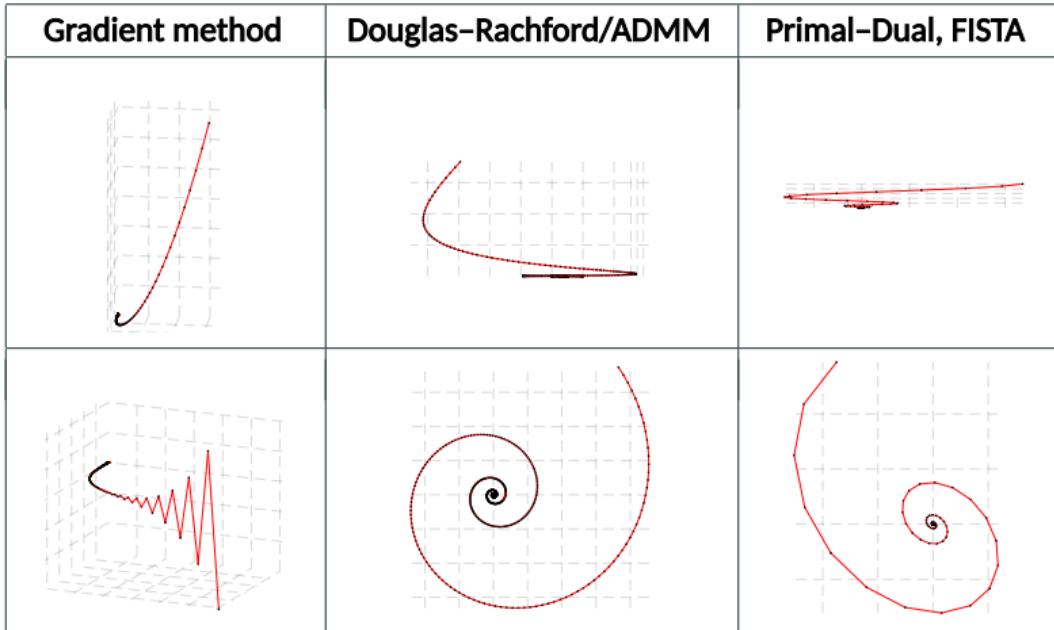
$$\mathcal{U} = \cup_{x \in \mathcal{M}_x} (x + \partial \|\cdot\|_1(x)) \rightarrow \bar{x} + \bar{v} \in \text{int}(\mathcal{U}).$$

Non-smooth opt.  $\xrightarrow{\text{structure}}$  FoM  $\xrightarrow{\text{geometry?}}$  Tra. of Seq.  $\implies$  Acceleration?

Power iteration in 3D: 2nd biggest eigenvalue is complex; the trajectory of eigenvector of the biggest eigenvalue.

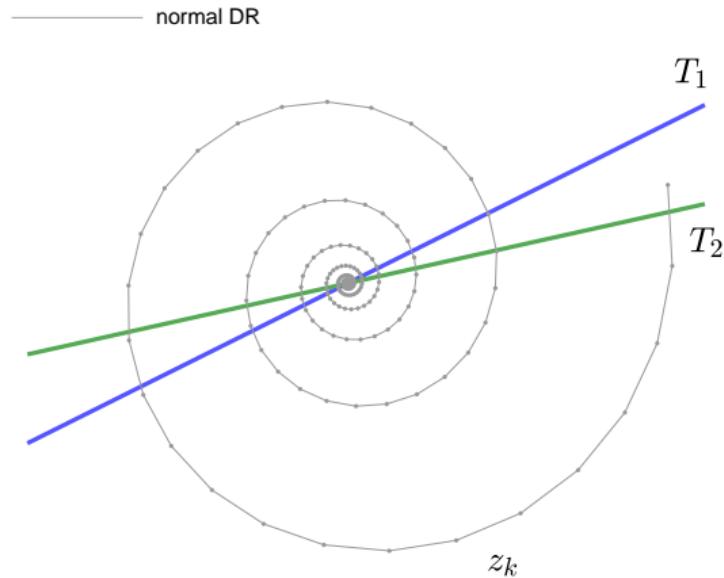


Non-smooth opt.  $\xrightarrow{\text{structure}}$  FoM  $\xrightarrow{\text{geometry?}}$  Tra. of Seq.  $\implies$  Acceleration?



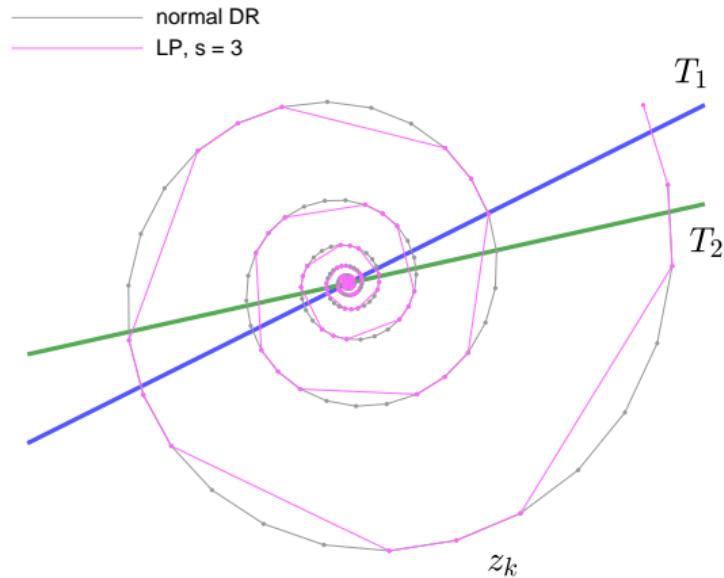
Non-smooth opt.  $\xrightarrow{\text{structure}}$  FoM  $\xrightarrow{\text{geometry?}}$  Tra. of Seq.  $\implies$  Acceleration?

Ada-acceleration (with C. Poon): adaptive acceleration based on trajectory.



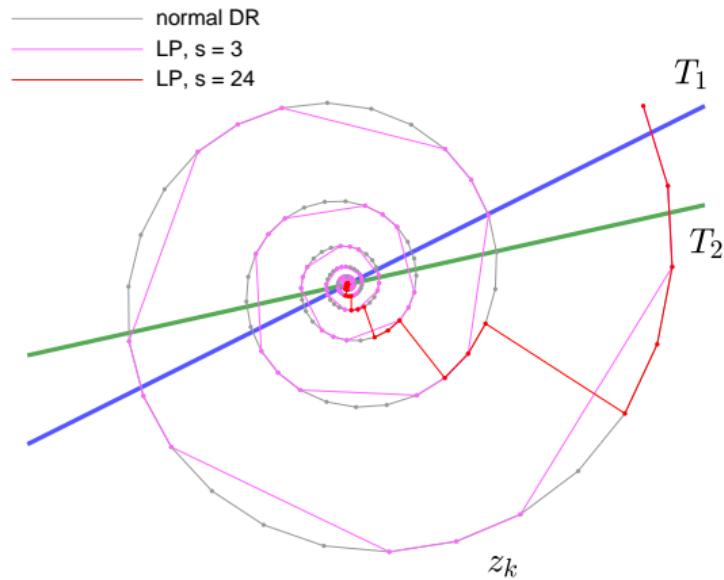
Non-smooth opt.  $\xrightarrow{\text{structure}}$  FoM  $\xrightarrow{\text{geometry?}}$  Tra. of Seq.  $\implies$  Acceleration?

Ada-acceleration (with C. Poon): adaptive acceleration based on trajectory.



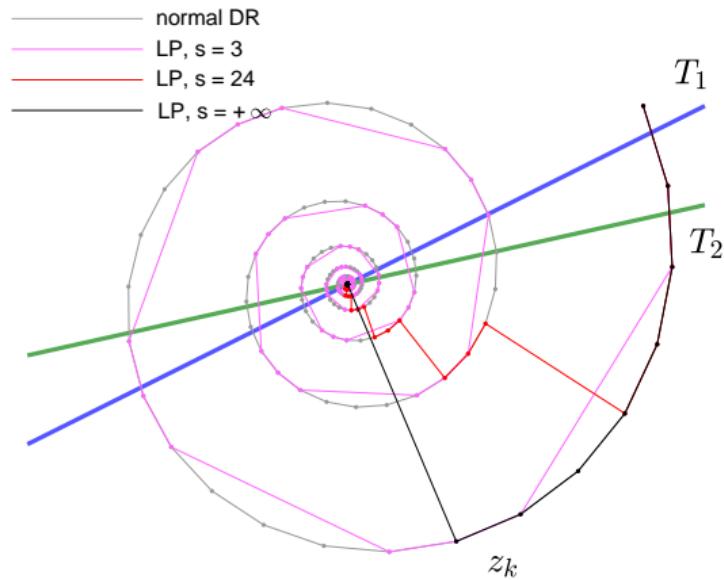
Non-smooth opt.  $\xrightarrow{\text{structure}}$  FoM  $\xrightarrow{\text{geometry?}}$  Tra. of Seq.  $\implies$  Acceleration?

Ada-acceleration (with C. Poon): adaptive acceleration based on trajectory.



Non-smooth opt.  $\xrightarrow{\text{structure}}$  FoM  $\xrightarrow{\text{geometry?}}$  Tra. of Seq.  $\implies$  Acceleration?

Ada-acceleration (with C. Poon): adaptive acceleration based on trajectory.



Partial smoothness builds an elegant connection between functions and the underlying Riemannian geometry

A unified framework for local analysis

Higher-order acceleration

Better understanding of existing algorithms

Steer new direction for designing accelerated schemes

**Thank you very much!**

<https://jliang993.github.io/>