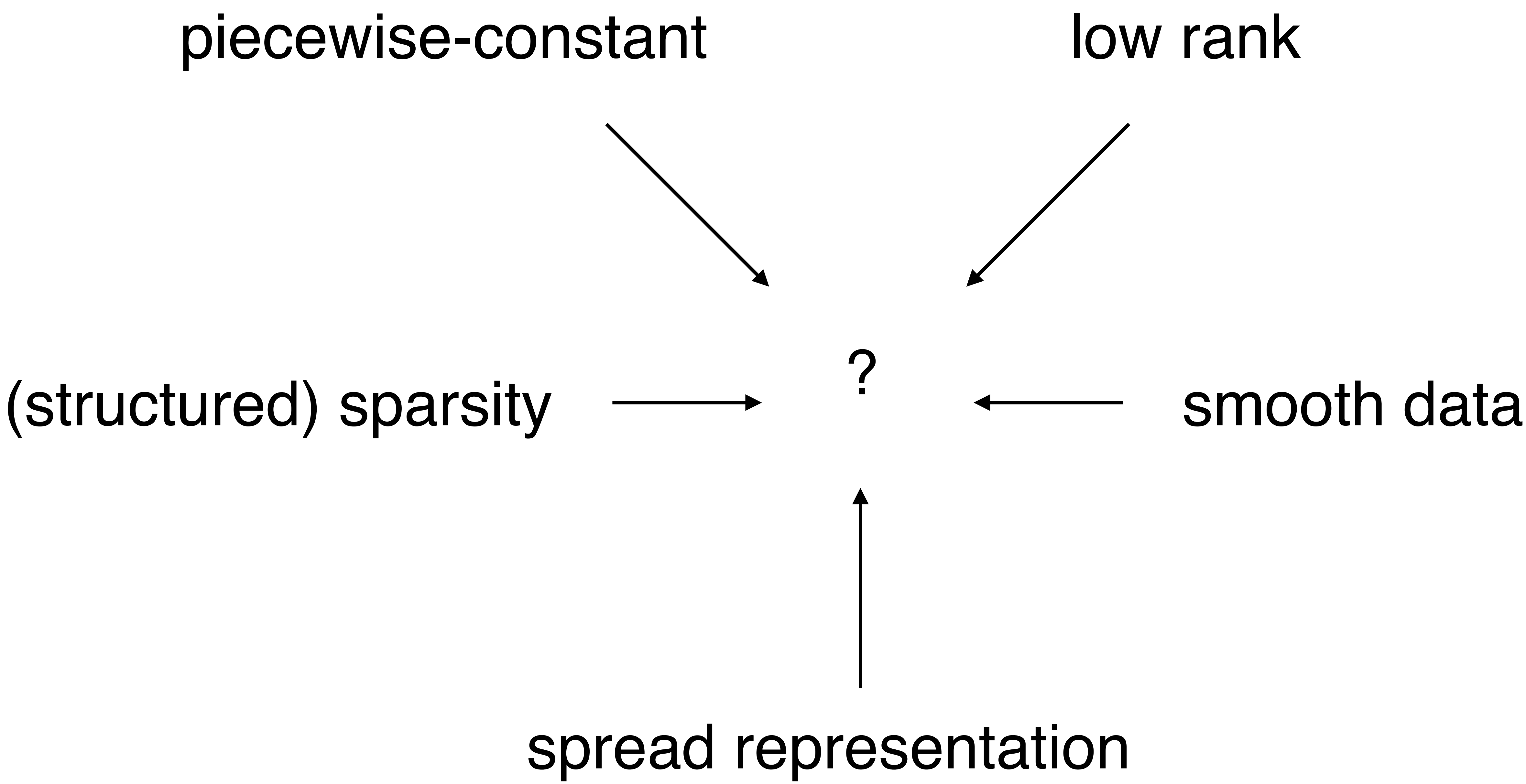
# Model Selection for Low Complexity Priors

**Samuel Vaiter**[1], Jalal Fadili and Gabriel Peyré

[1]CNRS & IMB, Dijon, (Burgundy) France

June 7, 2018
@SIAM IS

# Context




[Negahban et al. ’12], [Candes-Recht ’13], [Chandrasekaran et al. ’10], …

# Inverse Problem and Variational Methods

*Inverse problem / regression setting*

$$y = \mathbf{\Phi} x_0 + w$$

observations $\mathbb{R}^n$

degradation operator $\mathbb{R}^p \to \mathbb{R}^n$

ground truth $\mathbb{R}^p$

noise

*Variational methods*

compromise

$$\hat{x}_\lambda(y) \in \underset{x \in \mathbb{R}^p}{\text{Argmin}}\ F(\mathbf{\Phi} x, y) + \lambda J(x)$$

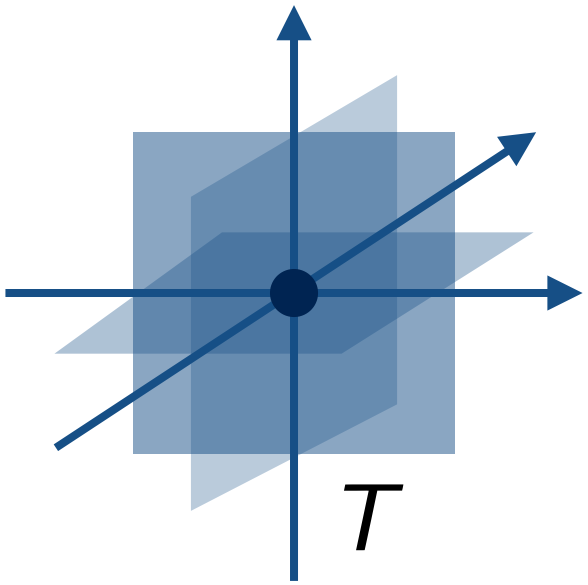
data fidelity

convex "regularization"

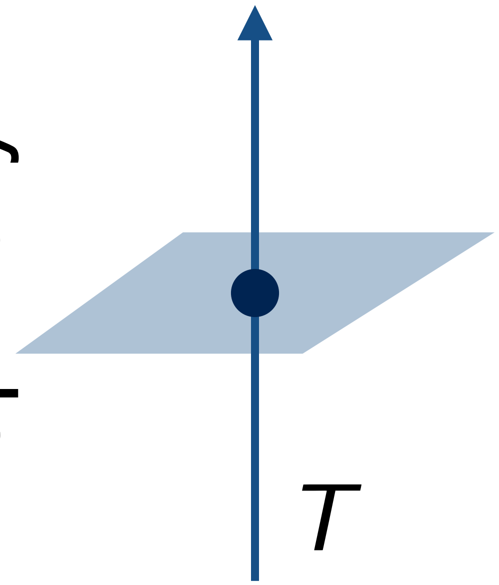
BPDN / LASSO
Total Variation
Nuclear
...

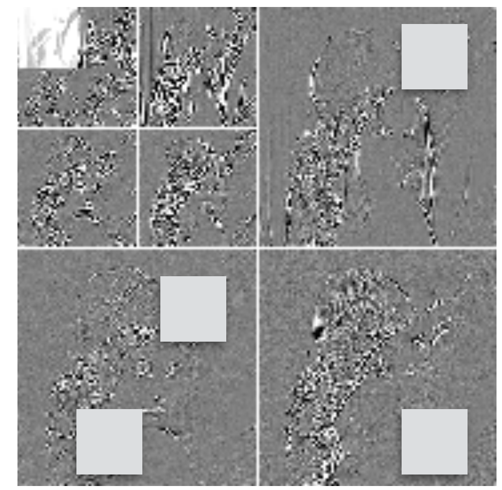
# Union of Models

Sparsity




Group Sparsity



$x$

$\Psi x$

⟶ combinatorial candidate $\|x\|_0 = |\operatorname{supp}(x)|$

⟶ convex candidate $\|x\|_1 = \sum |x_i|$

How to relate the model $T$ and the functional $J$ ?

# What kind of results ?

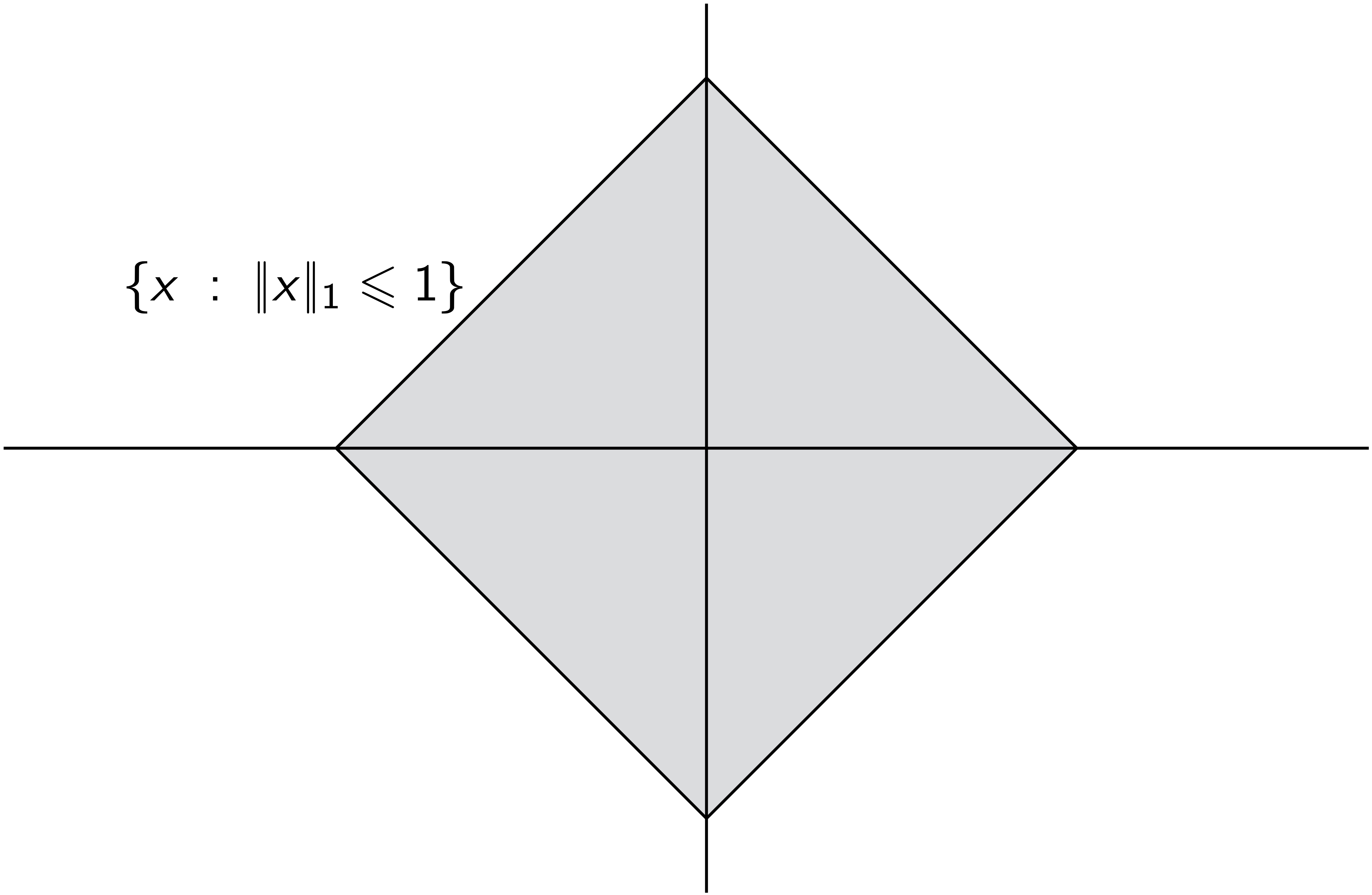
- (deterministic and non-uniform) $\ell^2$ stability

- (random/deterministic and non-uniform) model stability

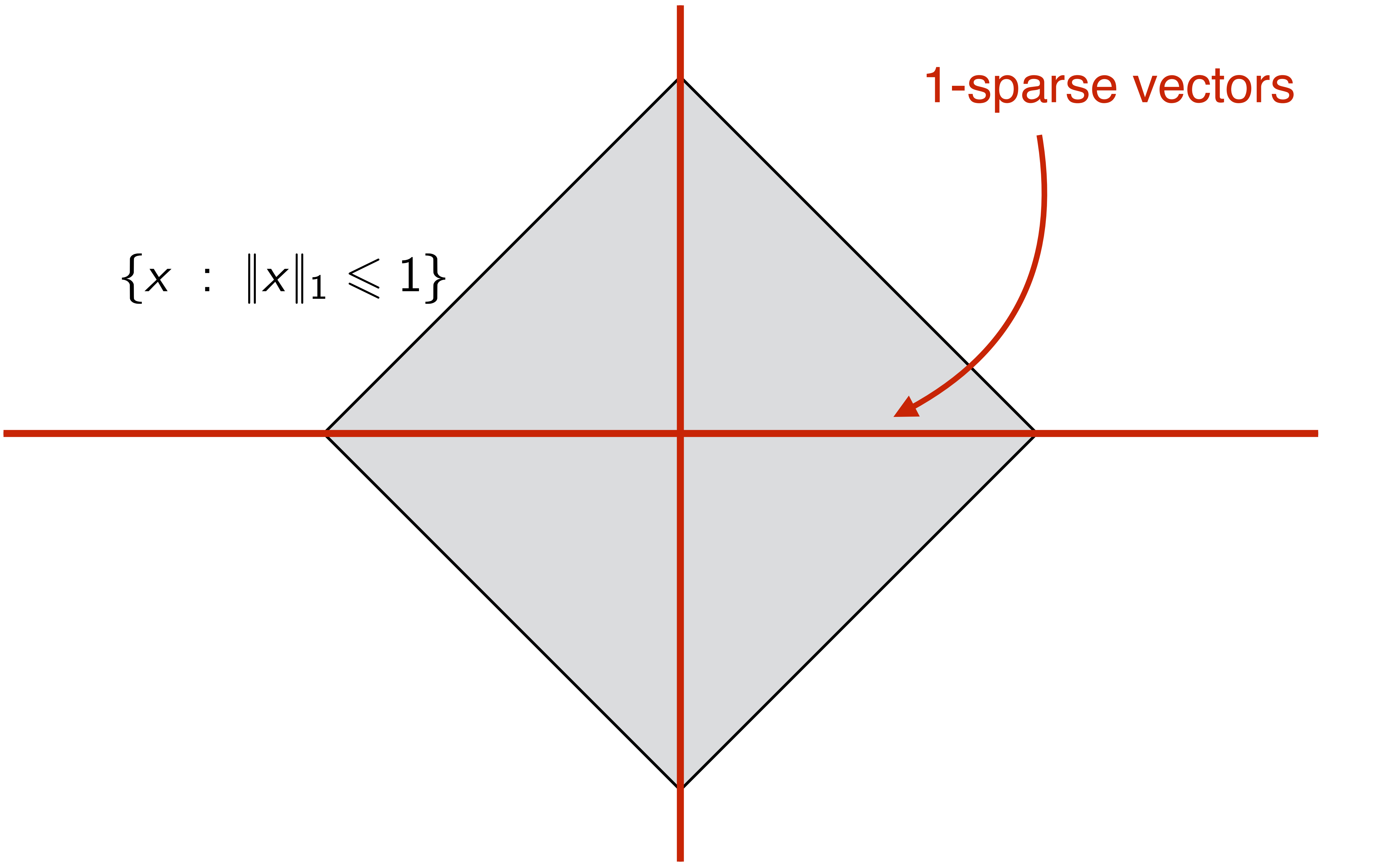
- degrees of freedom (DoF) computation

- efficient risk estimation

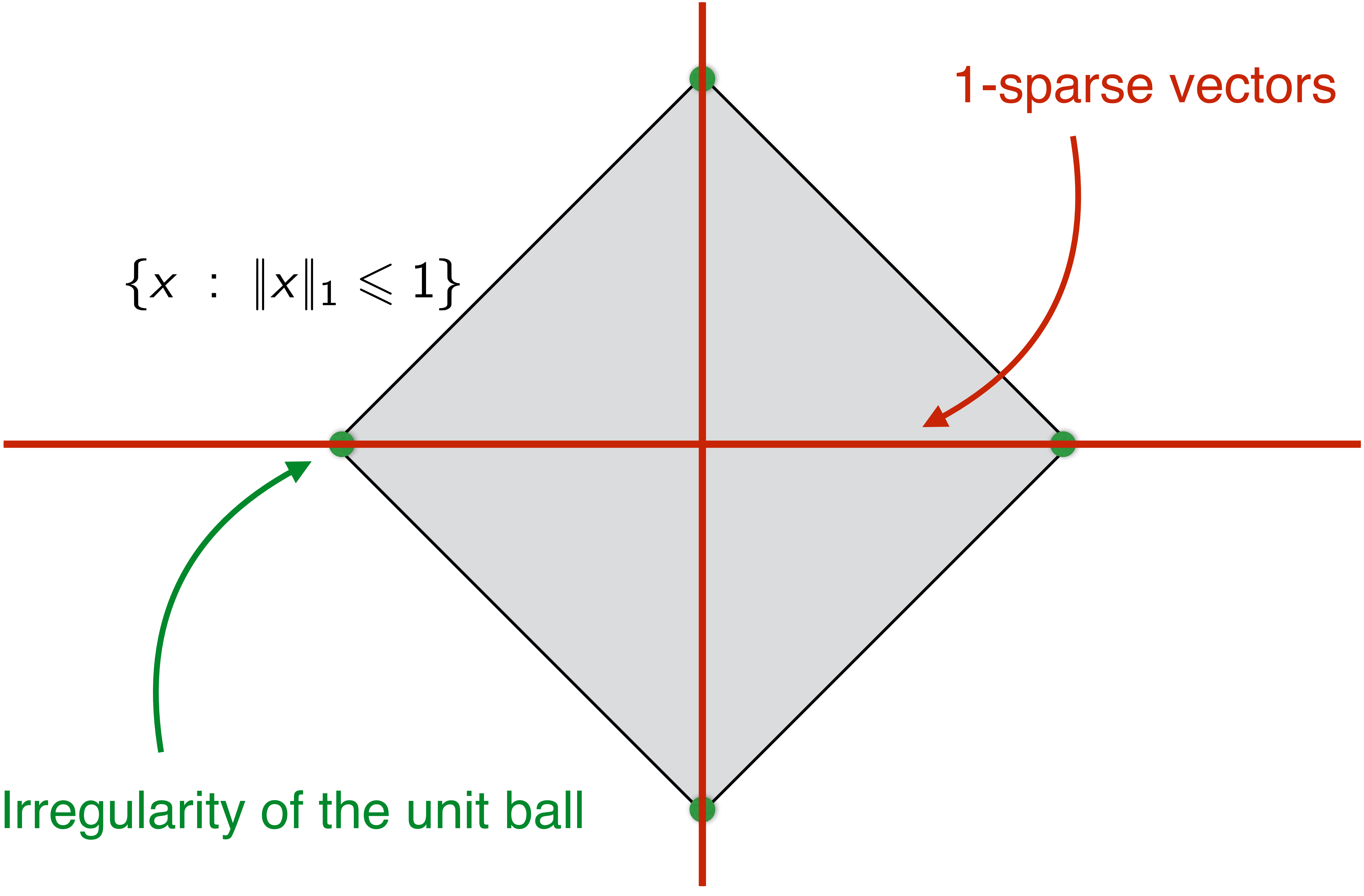
# Combinatorial and Convex Objects



# Combinatorial and Convex Objects



# Combinatorial and Convex Objects



# A Decomposability Point of View

Proposition

$$\partial\|\cdot\|_1(x) = \{\eta \in \mathbb{R}^n \;:\; \eta_I = \mathrm{sign}(x)_I \quad \text{and} \quad \|\eta_J\|_\infty \leqslant 1\}$$

# A Decomposability Point of View

Proposition

$$\partial\|\cdot\|_1(x) = \{\eta \in \mathbb{R}^n \;:\; \eta_I = \text{sign}(x)_I \quad \text{and} \quad \|\eta_J\|_\infty \leqslant 1\}$$

Definition (*Decomposable norm* [Candes-Recht '12])

$$\partial\|\cdot\|(x) = \left\{\eta \in \mathbb{R}^n \;:\; \eta_{T_x} = e_x \quad \text{and} \quad \|\eta_{T_x^\perp}\|_* \leqslant 1\right\}$$

# A Decomposability Point of View

**Proposition**

$$\partial \| \cdot \|_1(x) = \{\eta \in \mathbb{R}^n \;:\; \eta_I = \text{sign}(x)_I \quad \text{and} \quad \|\eta_J\|_\infty \leqslant 1\}$$

**Definition** (*Decomposable norm* [Candes-Recht '12])

$$\partial \| \cdot \|(x) = \left\{\eta \in \mathbb{R}^n \;:\; \eta_{T_x} = e_x \quad \text{and} \quad \|\eta_{T_x^\perp}\|_* \leqslant 1\right\}$$

**Proposition** [V. et al. '15]

$$\partial J(x) = \left\{\eta \in \mathbb{R}^n \;:\; \eta_{T_x} = e_x \quad \text{and} \quad f_x(\eta_{T_x^\perp}) \leqslant 1\right\}$$

# A Decomposability Point of View

**Proposition**

$$\partial\|\cdot\|_1(x) = \{\eta \in \mathbb{R}^n \;:\; \eta_I = \mathrm{sign}(x)_I \quad \text{and} \quad \|\eta_J\|_\infty \leqslant 1\}$$

**Definition** (*Decomposable norm* [Candes-Recht '12])

$$\partial\|\cdot\|(x) = \left\{\eta \in \mathbb{R}^n \;:\; \eta_{T_x} = e_x \quad \text{and} \quad \|\eta_{T_x^\perp}\|_* \leqslant 1\right\}$$

**Proposition** [V. et al. '15]

$$\partial J(x) = \left\{\eta \in \mathbb{R}^n \;:\; \eta_{T_x} = e_x \quad \text{and} \quad f_x(\eta_{T_x^\perp}) \leqslant 1\right\}$$

**Definition** (*Partly smooth wrt linear model* [V. et al. '15])

$T_x$, $e_x$, $f_x$ "Lipschitz"-continuous

# A Decomposability Point of View

**Proposition**

$$\partial \| \cdot \|_1(x) = \{\eta \in \mathbb{R}^n \ : \ \eta_I = \text{sign}(x)_I \quad \text{and} \quad \|\eta_J\|_\infty \leqslant 1\}$$

**Definition** (*Decomposable norm* [Candes-Recht '12])

$$\partial \| \cdot \|(x) = \left\{\eta \in \mathbb{R}^n \ : \ \eta_{T_x} = e_x \quad \text{and} \quad \|\eta_{T_x^\perp}\|_* \leqslant 1\right\}$$

**Proposition** [V. et al. '15]

$$\partial J(x) = \left\{\eta \in \mathbb{R}^n \ : \ \eta_{T_x} = e_x \quad \text{and} \quad f_x(\eta_{T_x^\perp}) \leqslant 1\right\}$$

**Definition** (*Partly smooth wrt linear model* [V. et al. '15])

$T_x$, $e_x$, $f_x$ "Lipschitz"-continuous

**Proposition** [V. et al. '15]

Almost all regularizers are PSL

# A Decomposability Point of View

**Proposition**

$$\partial\|\cdot\|_1(x) = \{\eta \in \mathbb{R}^n \;:\; \eta_I = \mathrm{sign}(x)_I \quad \text{and} \quad \|\eta_J\|_\infty \leqslant 1\}$$

**Definition** (*Decomposable norm* [Candes-Recht '12])

$$\partial\|\cdot\|(x) = \left\{\eta \in \mathbb{R}^n \;:\; \eta_{T_x} = e_x \quad \text{and} \quad \|\eta_{T_x^\perp}\|_* \leqslant 1\right\}$$

**Proposition** [V. et al. '15]

$$\partial J(x) = \left\{\eta \in \mathbb{R}^n \;:\; \eta_{T_x} = e_x \quad \text{and} \quad f_x(\eta_{T_x^\perp}) \leqslant 1\right\}$$

**Definition** (*Partly smooth wrt linear model* [V. et al. '15])

$T_x$, $e_x$, $f_x$ "Lipschitz"-continuous

**Proposition** [V. et al. '15]

Almost all regularizers are PSL … except the nuclear norm :(

# Combinatorial and Convex Objects

$\mathrm{Sym}_2(\mathbb{R})$ $\qquad\qquad$ $\mathbb{R}^3$

$$\begin{pmatrix} a & b \\ b & c \end{pmatrix} \longrightarrow \begin{pmatrix} a \\ b \\ c \end{pmatrix}$$

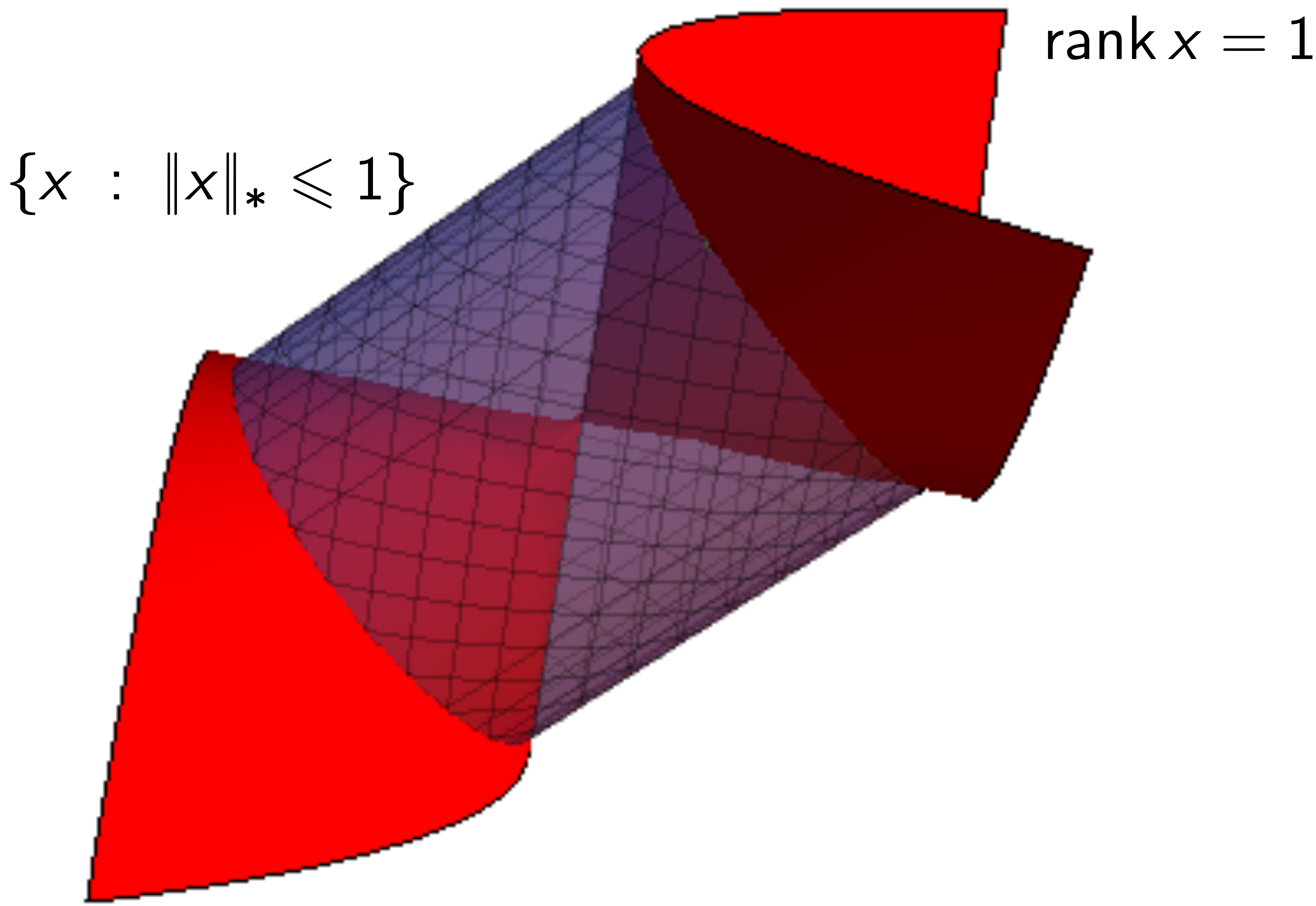
Matrix of rank 1 (+ zero vector)

$$ac - b^2 = 0 \rightarrow \text{curve of degree 2}$$

Matrix of unit nuclear norm

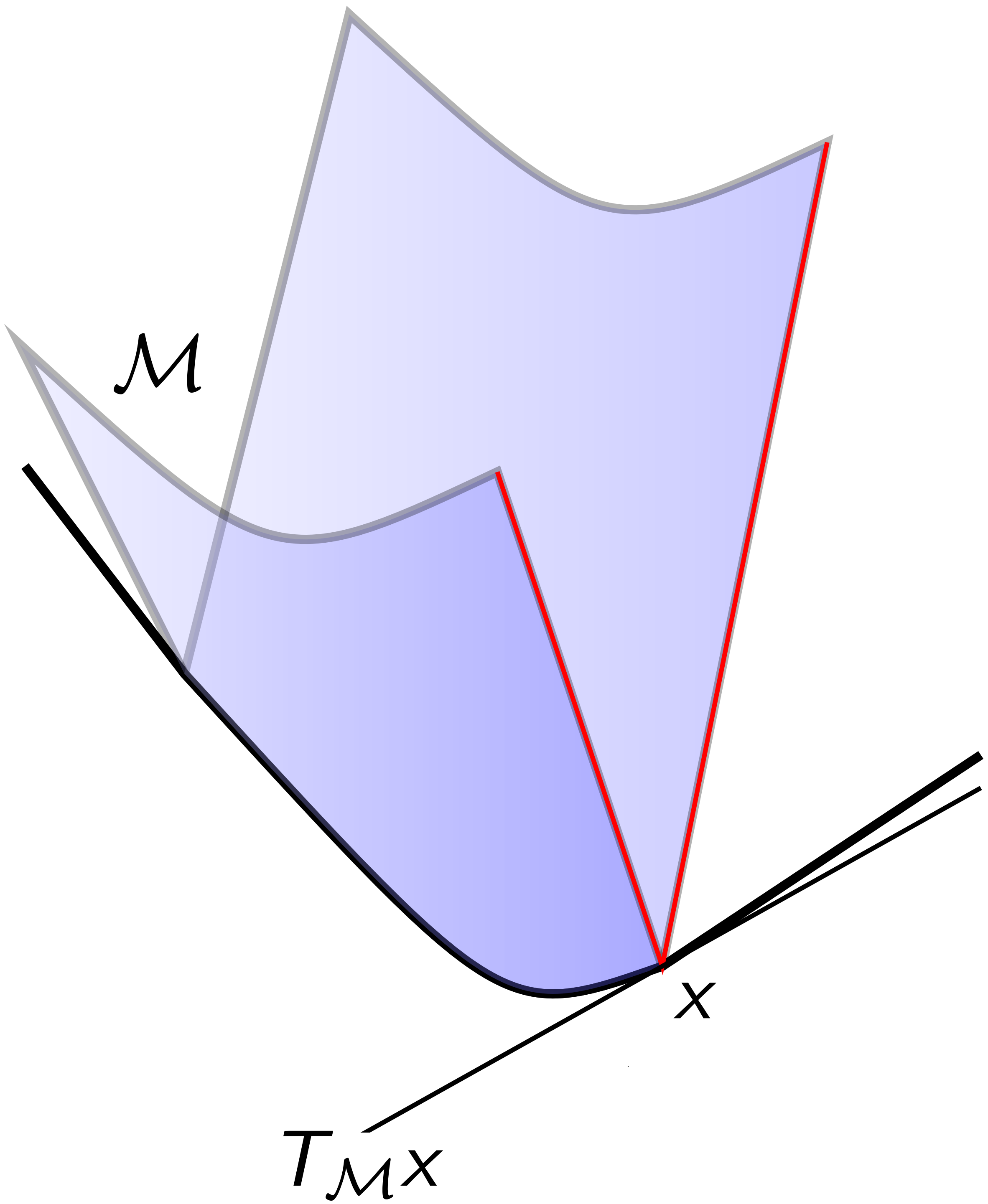
finite cylinder

# Combinatorial and Convex Objects



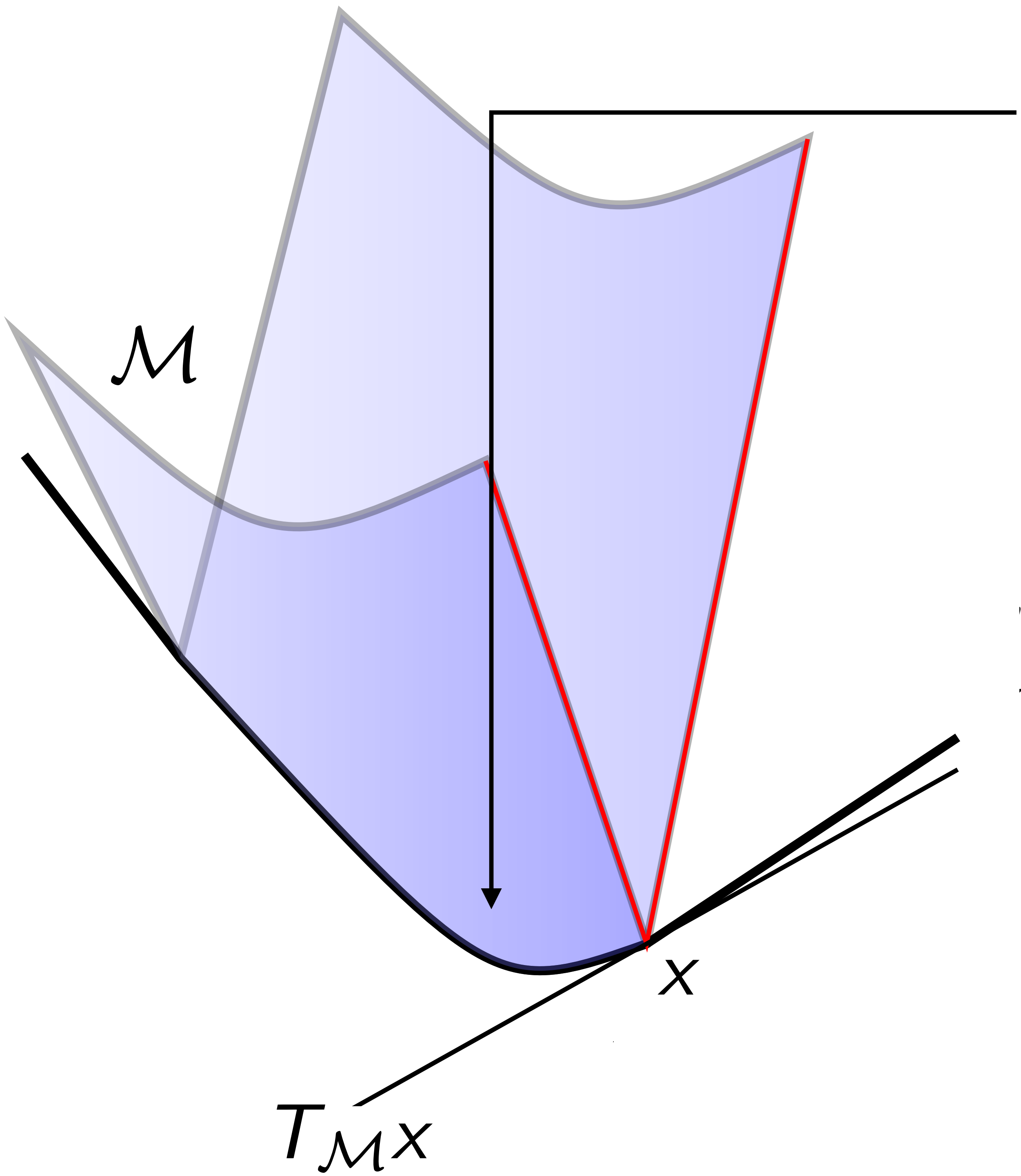
# Partial Smoothness

*Partly smooth function* [Lewis '02]



# Partial Smoothness

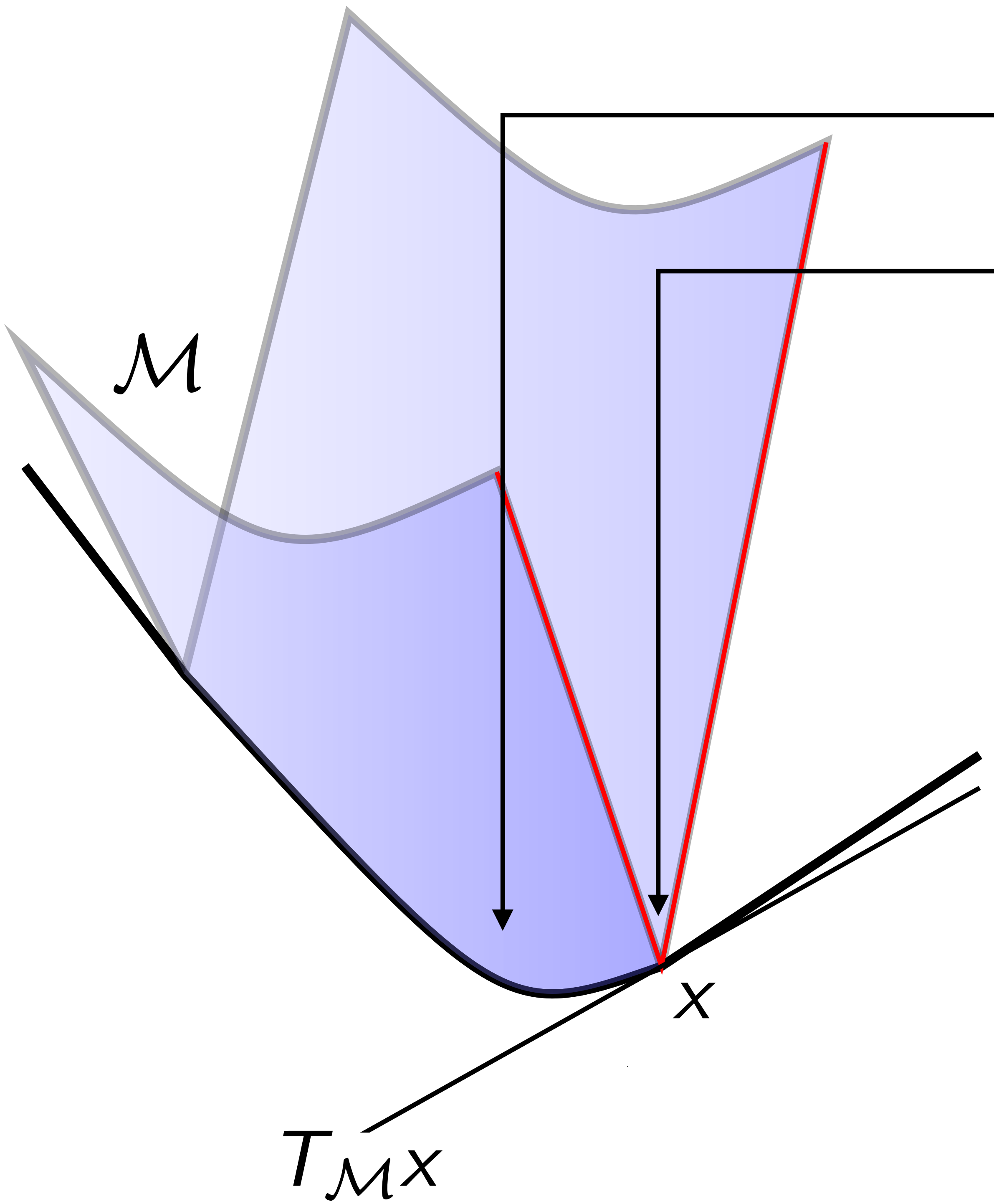
*Partly smooth function* [Lewis '02]




$J$ restricted to $\mathcal{M}$ is $C^2$

# Partial Smoothness

*Partly smooth function* [Lewis '02]

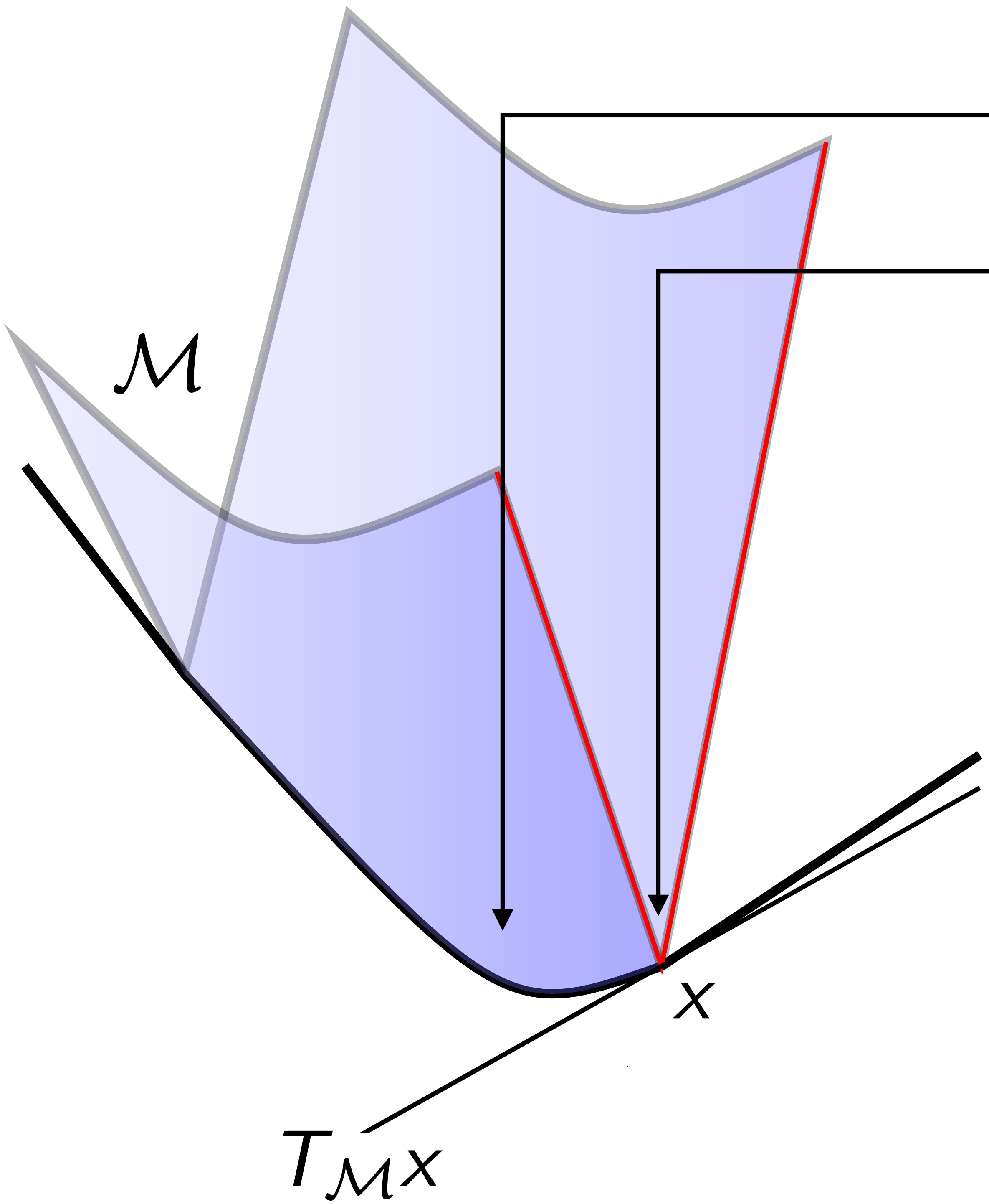



$J$ restricted to $\mathcal{M}$ is $C^2$

$\forall h \in (T_{\mathcal{M}}x)^{\perp}, t \mapsto J(x + th)$
not smooth at 0

# Partial Smoothness

*Partly smooth function* [Lewis '02]




$J$ restricted to $\mathcal{M}$ is $C^2$

$\forall h \in (T_{\mathcal{M}}x)^{\perp}, t \mapsto J(x+th)$
not smooth at 0

*Examples*

$\|\cdot\|_1, \|\nabla\cdot\|_1, \|\cdot\|_{1,2}, \|\cdot\|_*, \|\cdot\|_\infty, \cdots$

*Algebraic stability*

Sum, precomposition by a linear operator and spectral lift are stable within the class of PS.

# Linearized Precertificate

Certificates are dual solutions of the noiseless problem, i.e.,

$$p \text{ s.t. } \Phi^* p \in \partial J(x_0)$$

# Linearized Precertificate

Certificates are dual solutions of the noiseless problem, i.e.,

$$p \text{ s.t. } \Phi^* p \in \partial J(x_0)$$

Minimal norm certificate

$$p_0 = \underset{\Phi^* p \in \partial J(x_0)}{\operatorname{argmin}} \|p\|$$

Linearized precertificate

$$p_F = \underset{\Phi^* p \in \operatorname{Aff} \partial J(x_0)}{\operatorname{argmin}} \|p\|$$

# Linearized Precertificate

Certificates are dual solutions of the noiseless problem, i.e.,

$$p \text{ s.t. } \Phi^* p \in \partial J(x_0)$$

**Minimal norm certificate**

$$p_0 = \underset{\Phi^* p \in \partial J(x_0)}{\text{argmin}} \|p\|$$

**Linearized precertificate**

$$p_F = \underset{\Phi^* p \in \text{Aff}\,\partial J(x_0)}{\text{argmin}} \|p\|$$

**Proposition**

$$\Phi^* p_0 \in \text{ri}\, \partial J(x_0)$$

$$\Updownarrow$$

$$\Phi^* p_F \in \text{ri}\, \partial J(x_0)$$

# Linearized Precertificate

Certificates are dual solutions of the noiseless problem, i.e.,

$$p \text{ s.t. } \Phi^* p \in \partial J(x_0)$$

**Minimal norm certificate**

$$p_0 = \underset{\Phi^* p \in \partial J(x_0)}{\operatorname{argmin}} \|p\|$$

**Linearized precertificate**

$$p_F = \underset{\Phi^* p \in \operatorname{Aff} \partial J(x_0)}{\operatorname{argmin}} \|p\|$$

**Proposition**

$$\Phi^* p_0 \in \operatorname{ri} \partial J(x_0)$$

$$\Updownarrow$$

$$\Phi^* p_F \in \operatorname{ri} \partial J(x_0)$$

*Example: Sparse Regularization (BPDN, Lasso)*

$$\Phi^* p_F \in \operatorname{ri} \partial \|\cdot\|_1(x_0) \Leftrightarrow \max_{j \in I^c} |\langle \Phi_j, \Phi_I (\Phi_I^* \Phi_I)^{-1} \operatorname{sign}(x_0)_I \rangle| < 1$$

[Fuchs '04], [Tropp '05], …

# Model Stability

**Theorem**

Assume $J$ is PS at $x_0$ relative to $\mathcal{M}$. Suppose

$$\Phi^* p_F \in \operatorname{ri} \partial J(x_0) \quad \text{and} \quad \operatorname{Ker} \Phi \cap T_{x_0} = \{0\}$$

There exists $C > 0$ such that if $\max(\lambda, \|w\|/\lambda) \leqslant C$, the unique solution of $x^\star$ of $(\mathcal{P}_{y,\lambda})$ satisfies

| | |
|---|---|
| $x^\star \in \mathcal{M}_{x_0}$ | model stability |
| $\|x^\star - x_0\| = O(\|w\|)$ | $\ell^2$-stability |

# Model Stability

**Theorem**

Assume $J$ is PS at $x_0$ relative to $\mathcal{M}$. Suppose

$$\Phi^* p_F \in \text{ri}\, \partial J(x_0) \quad \text{and} \quad \text{Ker}\, \Phi \cap T_{x_0} = \{0\}$$

There exists $C > 0$ such that if $\max(\lambda, \|w\|/\lambda) \leqslant C$, the unique solution of $x^\star$ of $(\mathcal{P}_{y,\lambda})$ satisfies

$x^\star \in \mathcal{M}_{x_0}$ model stability

$\|x^\star - x_0\| = O(\|w\|)$ $\ell^2$-stability

Tightness

If $\Phi^* p_F \notin \partial J(x_0) \Rightarrow x^\star \notin \mathcal{M}_{x_0}$ no model stability

If $\Phi^* p_F \in \text{bd}\, \partial J(x_0) \Rightarrow$ case-by-case

# Identification in Finite Time

Forward—Backward scheme

$$x^{k+1} = \mathrm{Prox}_{\rho J}(x^k - \rho \nabla F(x^k))$$

# Identification in Finite Time

Forward—Backward scheme

$$x^{k+1} = \text{Prox}_{\rho J}(x^k - \rho \nabla F(x^k))$$

**Theorem**

Assume $J$ is PS at $x_0$ relative to $\mathcal{M}$. Suppose

$$\Phi^* p_F \in \text{ri}\, \partial J(x_0) \quad \text{and} \quad \text{Ker}\, \Phi \cap T_{x_0} = \{0\}$$

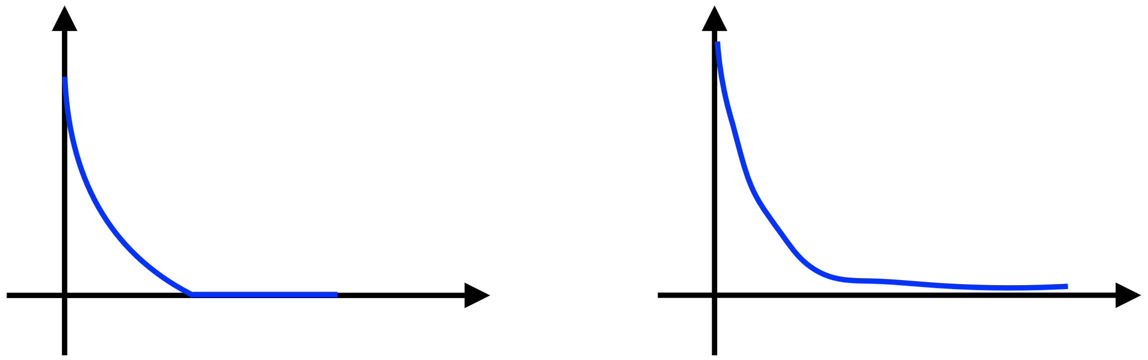
There exists $C > 0$ such that if $\max(\lambda, \|w\|/\lambda) \leqslant C$, and $\rho < 2/\|\Phi\|$, one has

$$x^k \to x_0 \qquad \text{convergence}$$

$$\exists k_0, \forall k \geqslant k_0, x^k \in \mathcal{M}_{x_0} \qquad \text{finite identification}$$

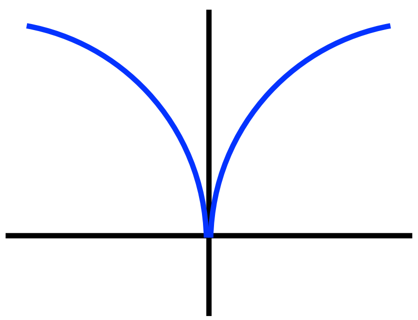
# Perspectives

*approximate model*



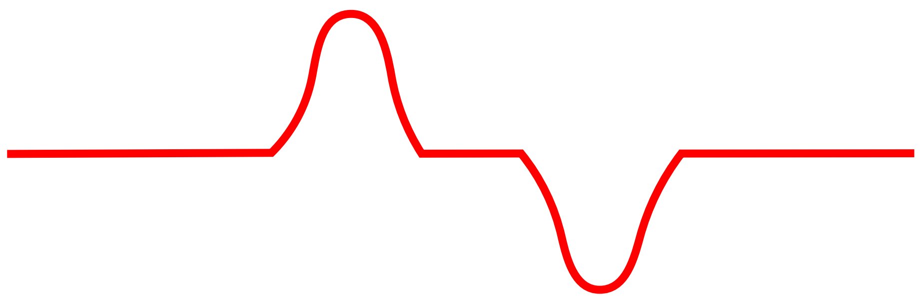
e.g. weak $\ell^p$ space

*non-convex regularization*



e.g. $\|\cdot\|_p$ with $0 < p < 1$

*infinite dimension*



e.g. C-BP, C-ROF

*algorithmic implication*

other algorithms,
accelerations

see Liang et al.

# Thanks for your attention

Want more ?

Review book chapter:

*V., G. Peyré, J. Fadili, Low Complexity Regularizations, LNCS, 2014*

Model selection/consistency:

*V., G. Peyré, J. Fadili, Manifold Consistency with Partly Smooth Regularizers, TIT 2017*

Special case for analysis sparsity:

*V., C. Dossal, G. Peyré, J. Fadili, Robust Sparse Analysis Regularization, TIT, 2013*

# *Example:* Nuclear Norm Regularization