Low Complexity Regularization of Inverse Problems

Cours #2
Recovery Guarantees

Gabriel Peyré

www.numerical-tours.com

Overview of the Course

- Course #1: Inverse Problems

- **Course #2: Recovery Guarantees**

- Course #3: Proximal Splitting Methods

Overview

- Low-complexity Regularization with Gauges

- Performance Guarantees

- Grid-free Regularization

Inverse Problem Regularization

$Observations:\ y = \Phi x_0 + w \in \mathbb{R}^P.$

$Estimator:\ x(y)$ depends only on → observations y ; → parameter λ

Inverse Problem Regularization

Observations: $y = \Phi x_0 + w \in \mathbb{R}^P$.

Estimator: $x(y)$ depends only on
- observations y
- parameter λ

Example: variational methods

$$x(y) \in \underset{x \in \mathbb{R}^N}{\operatorname{argmin}} \frac{1}{2}\|y - \Phi x\|^2 + \lambda J(x)$$

Data fidelity — Regularity

Inverse Problem Regularization

Observations: $y = \Phi x_0 + w \in \mathbb{R}^P$.

Estimator: $x(y)$ depends only on
- observations y
- parameter λ

Example: variational methods

$$x(y) \in \underset{x \in \mathbb{R}^N}{\text{argmin}} \; \frac{1}{2}\|y - \Phi x\|^2 + \lambda \, J(x)$$

Data fidelity — Regularity

Choice of λ: tradeoff

Noise level $\|w\|$	⟷	Regularity of x_0 $J(x_0)$

Inverse Problem Regularization

Observations: $y = \Phi x_0 + w \in \mathbb{R}^P$.

Estimator: $x(y)$ depends only on
- observations y
- parameter λ

Example: variational methods

$$x(y) \in \underset{x \in \mathbb{R}^N}{\text{argmin}} \frac{1}{2}\|y - \Phi x\|^2 + \lambda J(x)$$

Data fidelity — Regularity

Choice of λ: tradeoff — Noise level $\|w\|$ ↔ Regularity of x_0 $J(x_0)$

No noise: $\lambda \to 0^+$, minimize $x^\star \in \underset{x \in \mathbb{R}^Q, \mathcal{K}x = y}{\text{argmin}} J(x)$

Inverse Problem Regularization

Observations: $y = \Phi x_0 + w \in \mathbb{R}^P$.

Estimator: $x(y)$ depends only on
- observations y
- parameter λ

Example: variational methods

$$x(y) \in \underset{x \in \mathbb{R}^N}{\text{argmin}} \; \frac{1}{2}\|y - \Phi x\|^2 + \lambda \, J(x)$$

Data fidelity — Regularity

Choice of λ: tradeoff

Noise level $\|w\|$ ⟷ Regularity of x_0 $J(x_0)$

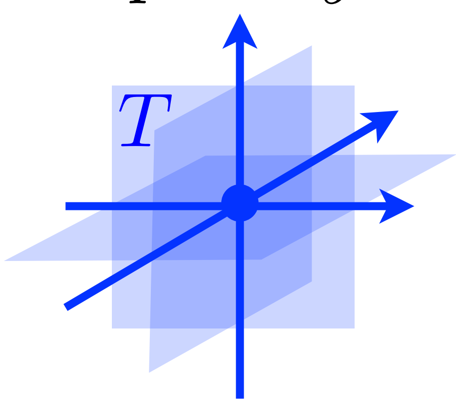
No noise: $\lambda \to 0^+$, minimize $x^\star \in \underset{x \in \mathbb{R}^Q, \mathcal{K}x = y}{\text{argmin}} \; J(x)$

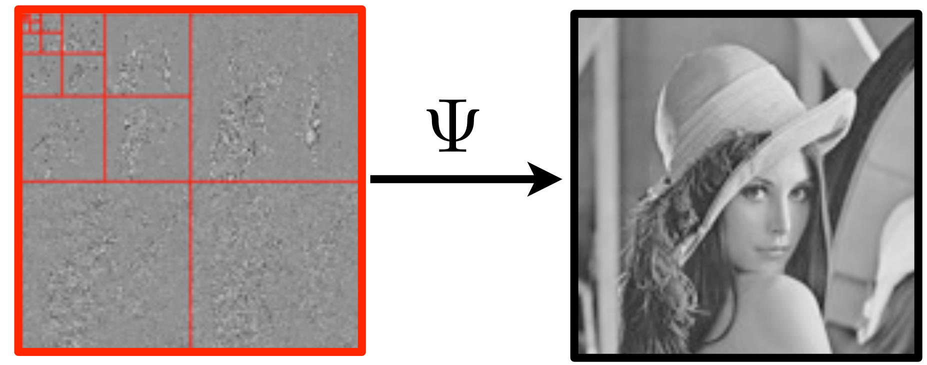
This course:
- **Performance analysis.**
- **Fast computational scheme.**

Union of Linear Models for Data Processing

Union of models: $T \in \mathcal{T}$ linear spaces.

Synthesis sparsity:



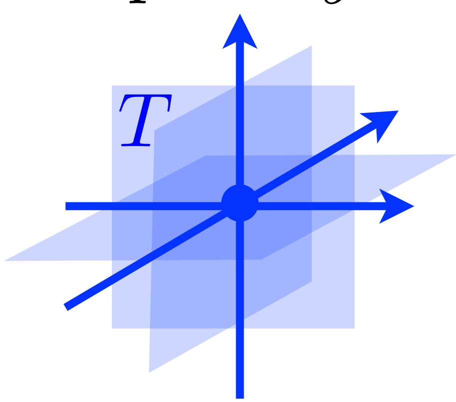


Coefficients x Image Ψx

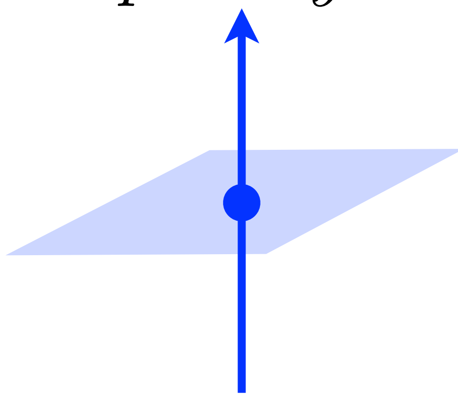
Union of Linear Models for Data Processing

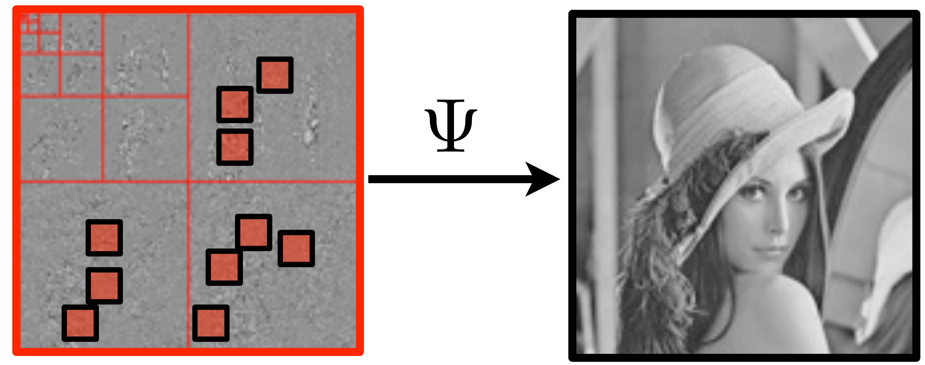
Union of models: $T \in \mathcal{T}$ linear spaces.

Synthesis sparsity:



Structured sparsity:





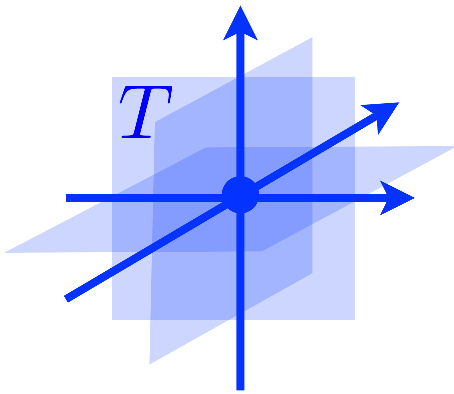
Coefficients x

Image Ψx

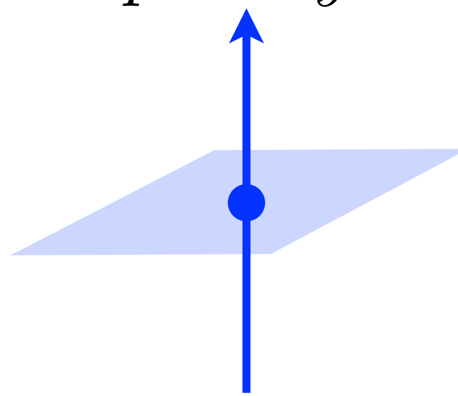
Union of Linear Models for Data Processing

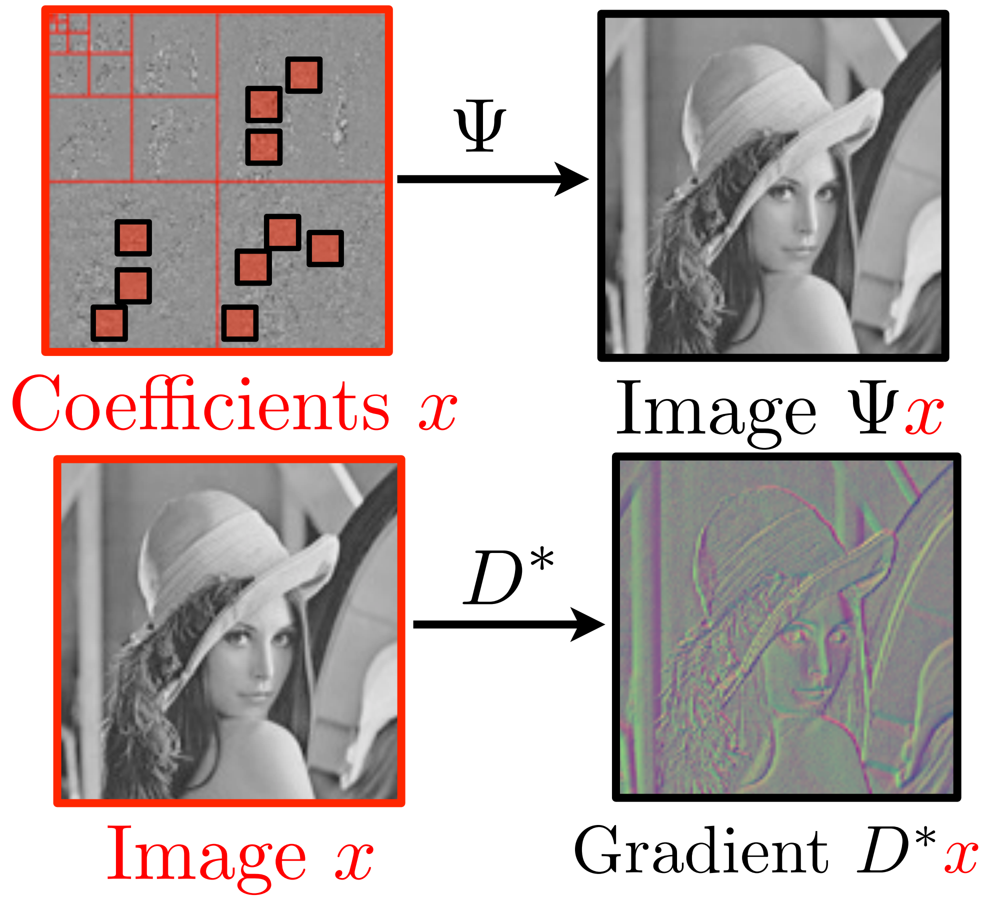
Union of models: $T \in \mathcal{T}$ linear spaces.

Synthesis sparsity:

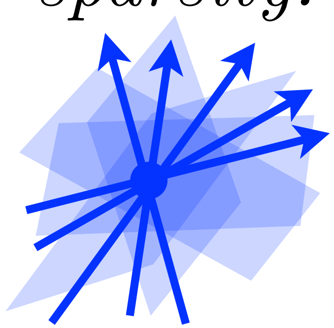


Structured sparsity:





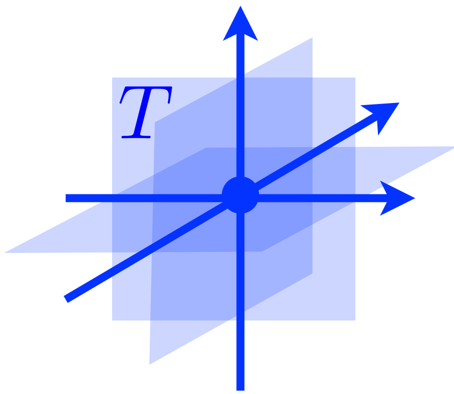
Analysis sparsity:



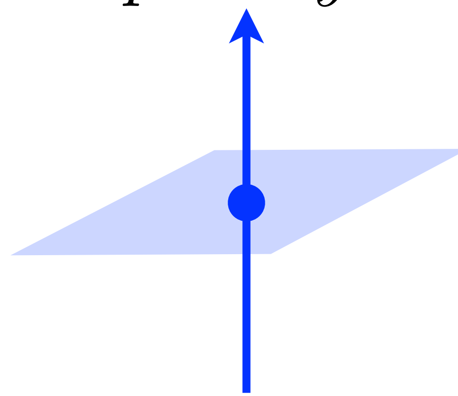
Union of Linear Models for Data Processing

Union of models: $T \in \mathcal{T}$ linear spaces.

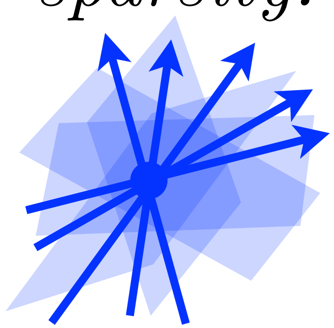
Synthesis sparsity:



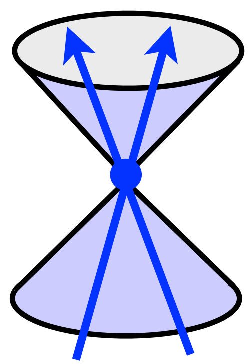
Structured sparsity:

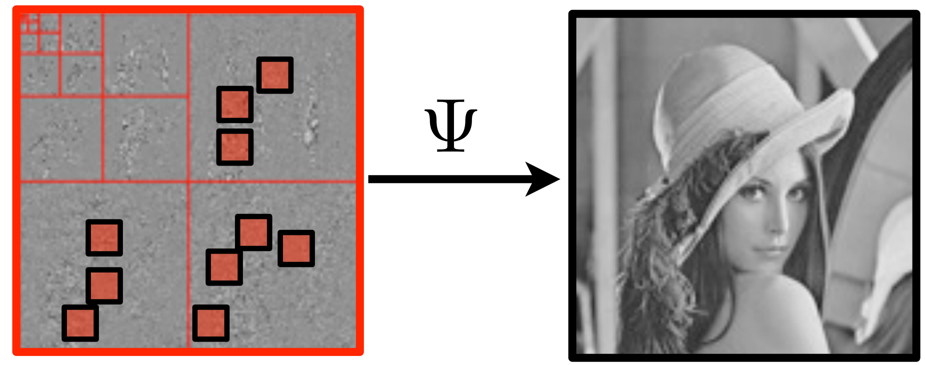


Analysis sparsity:



Low-rank:





Coefficients x Image Ψx

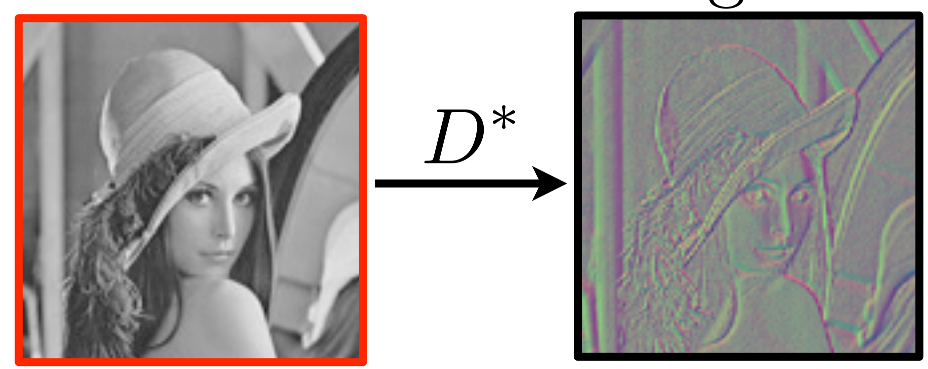
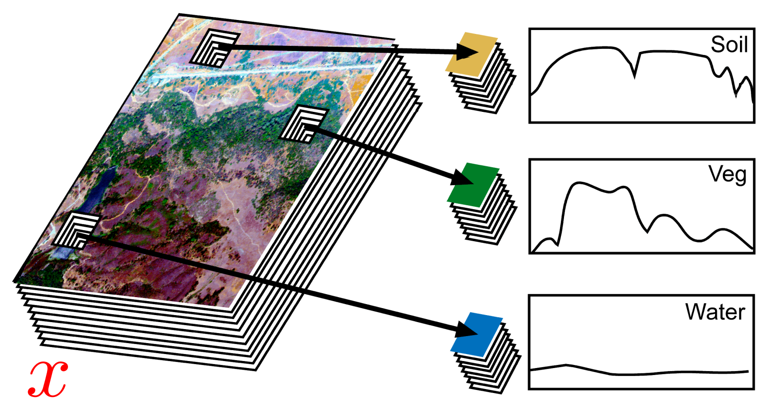


Image x Gradient $D^* x$

Multi-spectral imaging:

$$x_{i,\cdot} = \sum_{j=1}^{r} A_{i,j} S_{j,\cdot}$$



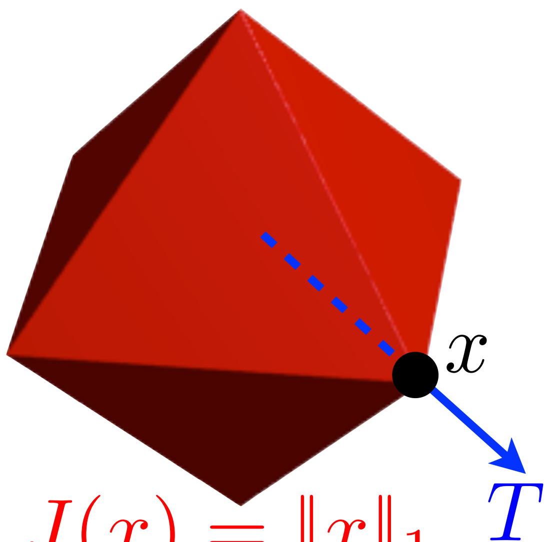
Gauges for Union of Linear Models

$$\textit{Gauge:} \quad J : \mathbb{R}^N \to \mathbb{R}^+ \quad \left| \begin{array}{l} \text{Convex} \\ \forall\, \alpha \in \mathbb{R}^+,\ J(\alpha x) = \alpha J(x) \end{array} \right.$$

Gauges for Union of Linear Models

Gauge: $J : \mathbb{R}^N \to \mathbb{R}^+$ | Convex
$\forall \alpha \in \mathbb{R}^+,\ J(\alpha x) = \alpha J(x)$

Piecewise regular ball $\Leftrightarrow$ Union of linear models $(T)_{T \in \mathcal{T}}$

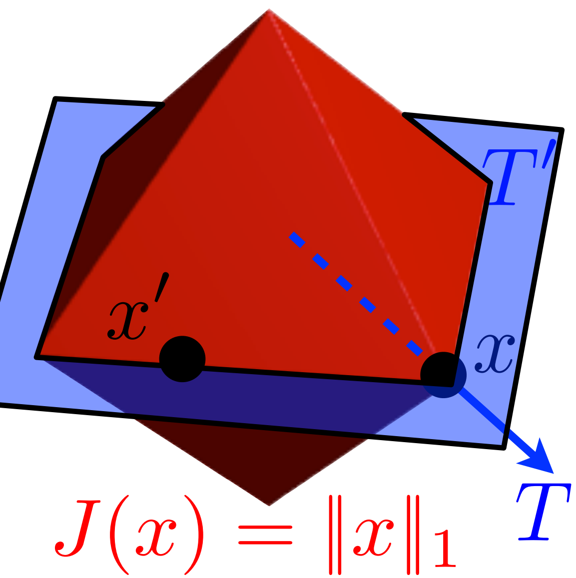


$\mathcal{T}$ = sparse vectors

Gauges for Union of Linear Models

Gauge: $J : \mathbb{R}^N \to \mathbb{R}^+$ | Convex; $\forall\, \alpha \in \mathbb{R}^+,\ J(\alpha x) = \alpha J(x)$

Piecewise regular ball $\Leftrightarrow$ Union of linear models $(T)_{T \in \mathcal{T}}$

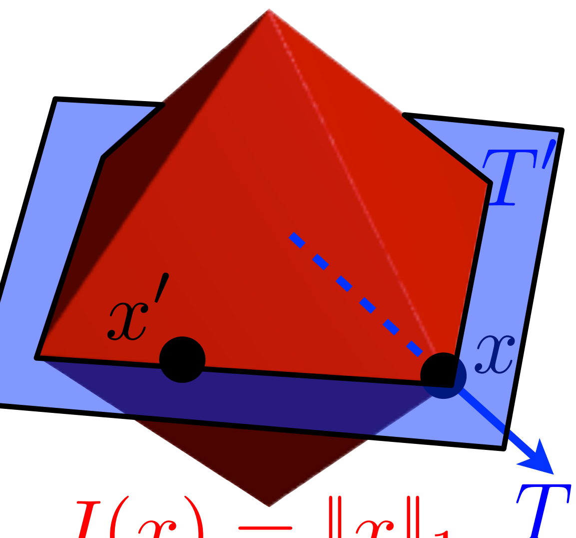


$\mathcal{T}$ = sparse vectors

Gauges for Union of Linear Models

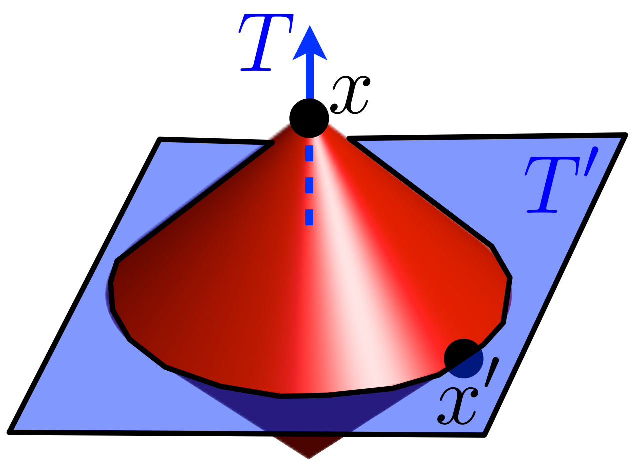
Gauge: $J : \mathbb{R}^N \to \mathbb{R}^+$ | Convex; $\forall\, \alpha \in \mathbb{R}^+,\ J(\alpha x) = \alpha J(x)$

Piecewise regular ball $\Leftrightarrow$ Union of linear models $(T)_{T \in \mathcal{T}}$



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

$\mathcal{T}$ = sparse vectors



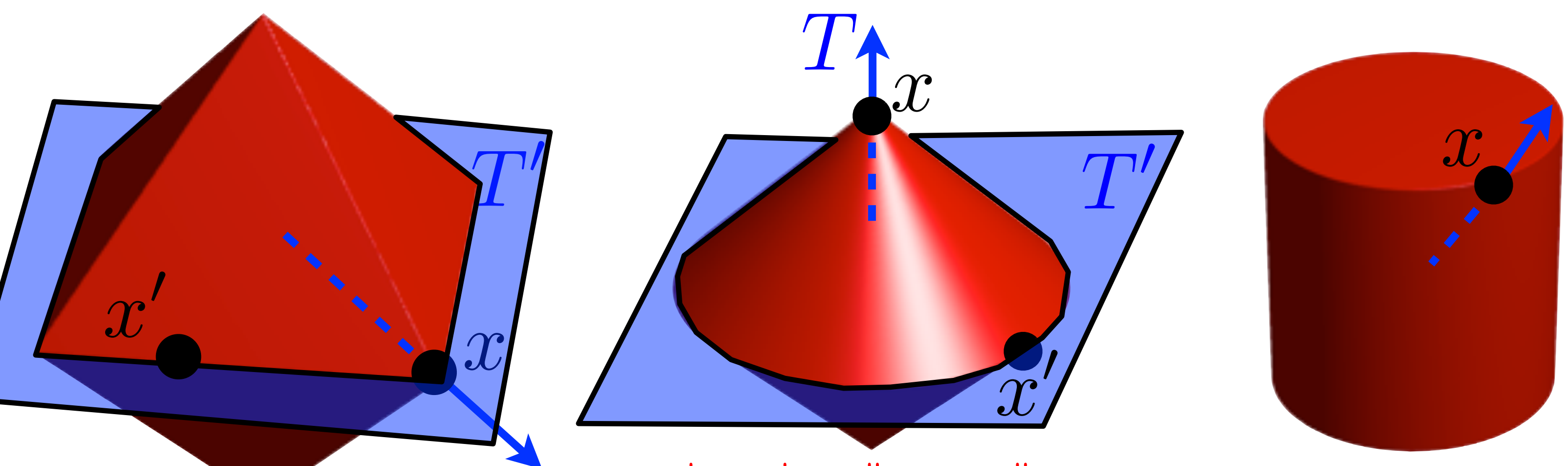
$|x_1| + \|x_{2,3}\|$

$\mathcal{T}$ = block sparse vectors

Gauges for Union of Linear Models

Gauge: $J : \mathbb{R}^N \to \mathbb{R}^+$ | Convex; $\forall\, \alpha \in \mathbb{R}^+,\; J(\alpha x) = \alpha J(x)$

Piecewise regular ball $\Leftrightarrow$ Union of linear models $(T)_{T \in \mathcal{T}}$



$J(x) = \|x\|_1$ — $\mathcal{T}$ = sparse vectors

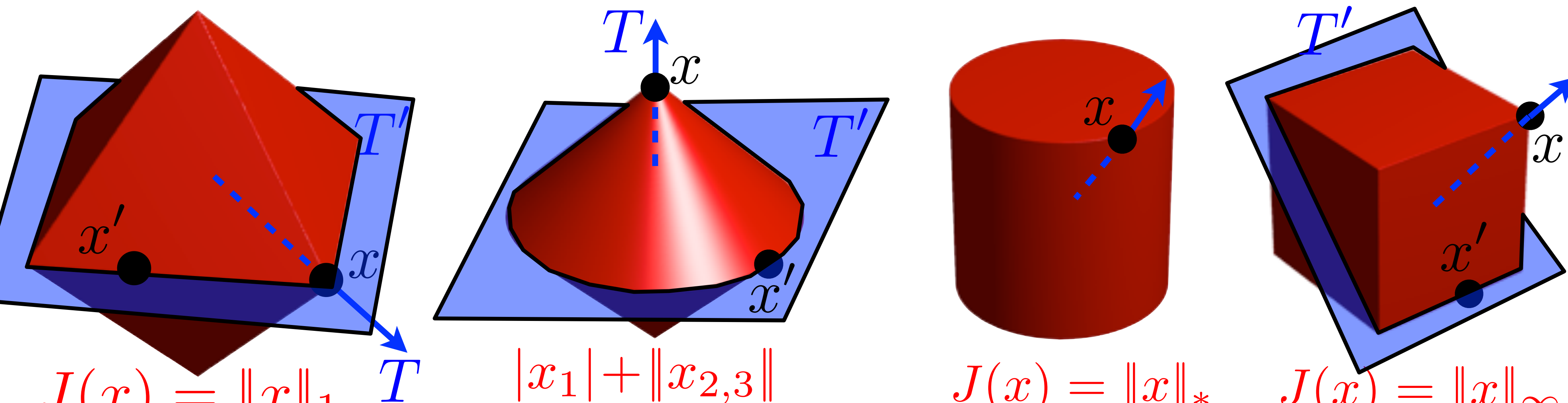
$|x_1| + \|x_{2,3}\|$ — $\mathcal{T}$ = block sparse vectors

$J(x) = \|x\|_*$ — $\mathcal{T}$ = low-rank matrices

Gauges for Union of Linear Models

Gauge: $J : \mathbb{R}^N \to \mathbb{R}^+$ | Convex; $\forall\, \alpha \in \mathbb{R}^+,\ J(\alpha x) = \alpha J(x)$

Piecewise regular ball $\Leftrightarrow$ Union of linear models $(T)_{T \in \mathcal{T}}$



$J(x) = \|x\|_1$ — $\mathcal{T}$ = sparse vectors

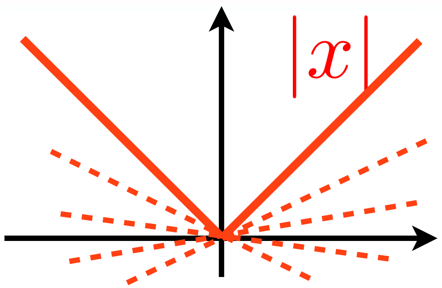
$|x_1| + \|x_{2,3}\|$ — $\mathcal{T}$ = block sparse vectors

$J(x) = \|x\|_*$ — $\mathcal{T}$ = low-rank matrices

$J(x) = \|x\|_\infty$ — $\mathcal{T}$ = anti-sparse vectors

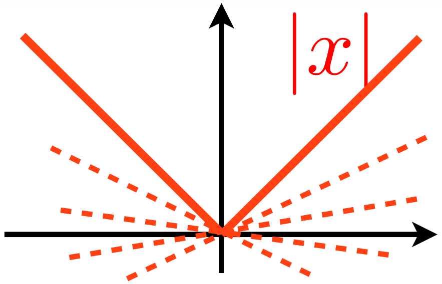
Subdifferentials and Models

$$\partial J(x) = \{\eta \setminus \forall\, y, J(y) \geqslant J(x) + \langle \eta,\, y - x\rangle\}$$



Subdifferentials and Models

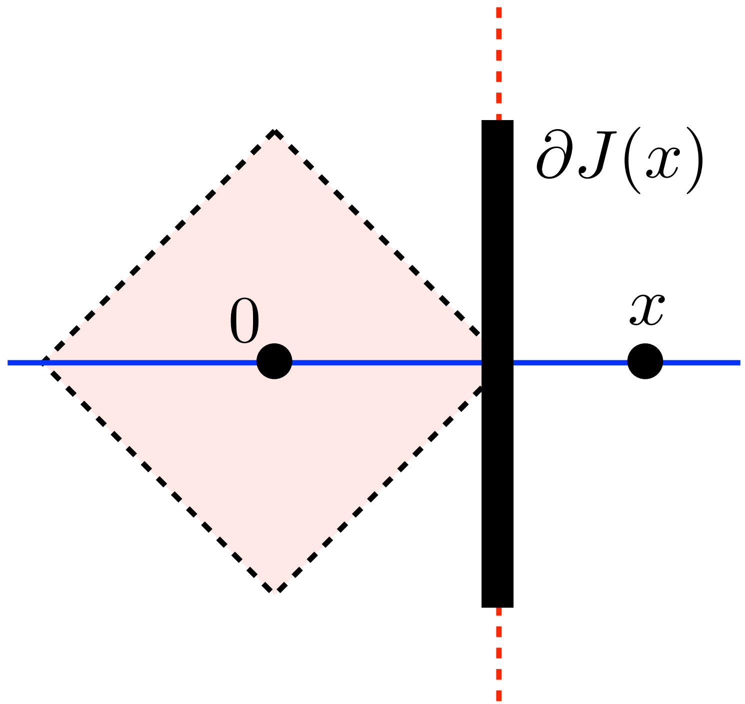
$$\partial J(x) = \{\eta \setminus \forall\, y, J(y) \geqslant J(x) + \langle \eta,\, y - x\rangle\}$$



Example: $J(x) = \|x\|_1$

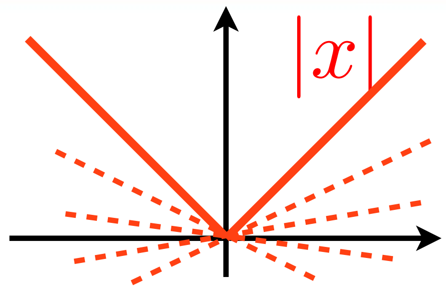
$$\partial \|x\|_1 = \left\{ \eta \setminus \begin{array}{l} \operatorname{supp}(\eta) = I, \\ \forall\, j \notin I, |\eta_j| \leqslant 1 \end{array} \right\}$$

$$I = \operatorname{supp}(x) = \{i \setminus x_i \neq 0\}$$



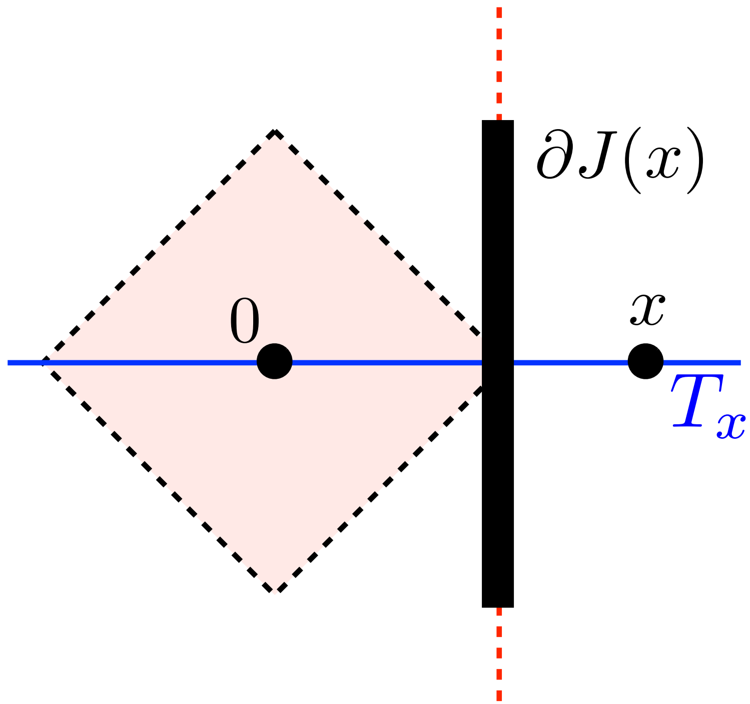
Subdifferentials and Models

$$\partial J(x) = \{\eta \setminus \forall\, y, J(y) \geqslant J(x) + \langle \eta,\, y - x \rangle\}$$



Example: $J(x) = \|x\|_1$

$$\partial \|x\|_1 = \left\{ \eta \setminus \begin{array}{l} \operatorname{supp}(\eta) = I, \\ \forall\, j \notin I, |\eta_j| \leqslant 1 \end{array} \right\}$$



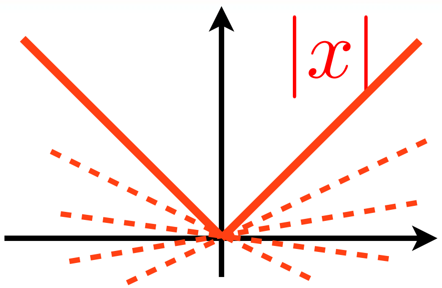
$$I = \operatorname{supp}(x) = \{i \setminus x_i \neq 0\}$$

$$T_x = \{\eta \setminus \operatorname{supp}(\eta) = I\}$$

Definition: $T_x = \operatorname{VectHull}(\partial J(x))^\perp$

Subdifferentials and Models

$$\partial J(x) = \{\eta \setminus \forall\, y,\, J(y) \geqslant J(x) + \langle \eta,\, y - x \rangle\}$$



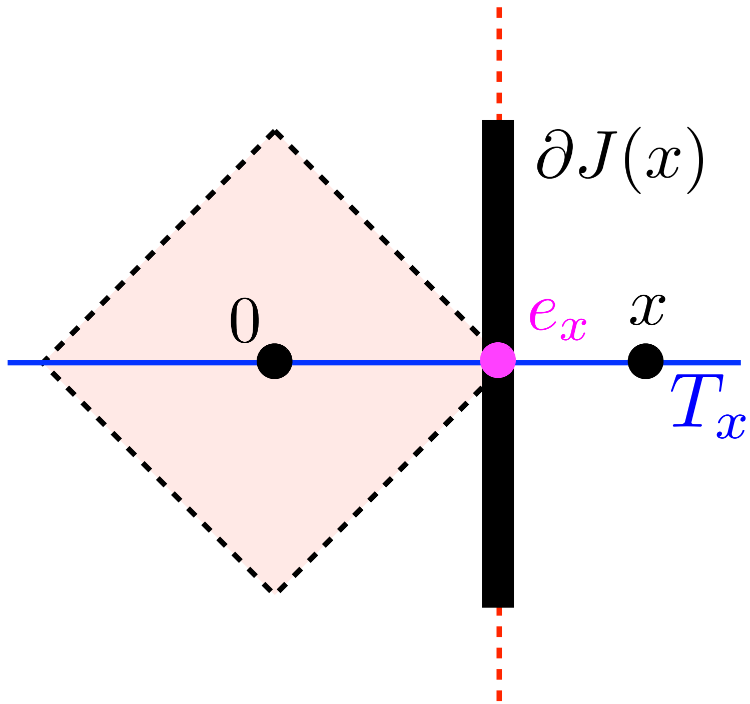
Example: $J(x) = \|x\|_1$

$$\partial \|x\|_1 = \left\{ \eta \setminus \begin{array}{l} \operatorname{supp}(\eta) = I, \\ \forall\, j \notin I,\, |\eta_j| \leqslant 1 \end{array} \right\}$$

$$I = \operatorname{supp}(x) = \{i \setminus x_i \neq 0\}$$

$$T_x = \{\eta \setminus \operatorname{supp}(\eta) = I\}$$

$$e_x = \operatorname{sign}(x)$$



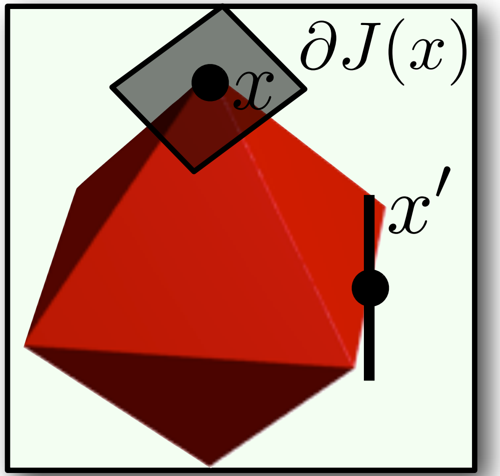
Definition: $T_x = \operatorname{VectHull}(\partial J(x))^\perp$

$$\eta \in \partial J(x) \implies \operatorname{Proj}_{T_x}(\eta) = e_x$$

Examples

ℓ^1 *sparsity:* $J(x) = \|x\|_1$

$$e_x = \text{sign}(x) \qquad T_x = \{z \setminus \text{supp}(z) \subset \text{supp}(x)\}$$



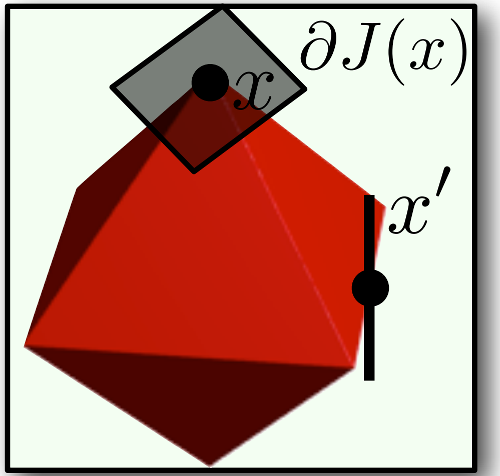
Examples

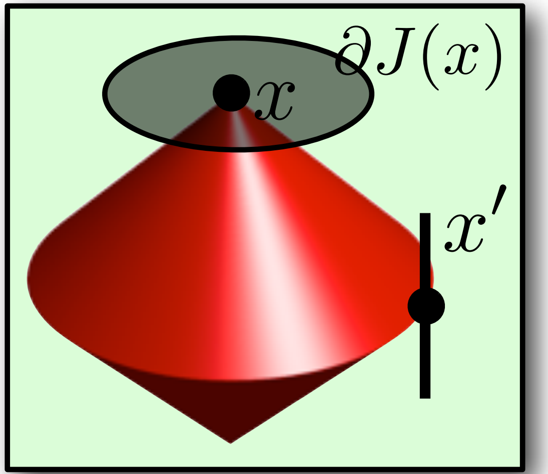
ℓ^1 *sparsity:* $J(x) = \|x\|_1$

$$e_x = \text{sign}(x) \qquad T_x = \{z \setminus \text{supp}(z) \subset \text{supp}(x)\}$$

Structured sparsity: $J(x) = \sum_b \|x_b\|$ $\qquad \mathcal{N}(a) = a/\|a\|$

$$e_x = (\mathcal{N}(x_b))_{b \in \mathcal{B}} \qquad T_x = \{z \setminus \text{supp}(z) \subset \text{supp}(x)\}$$





Examples

ℓ^1 *sparsity:* $J(x) = \|x\|_1$

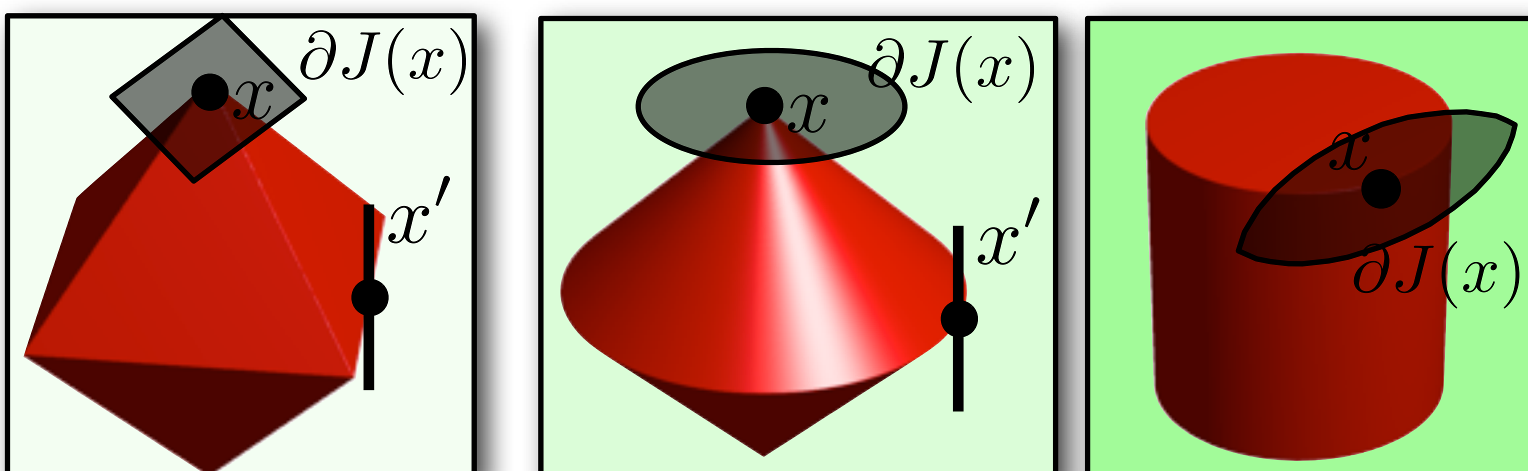
$e_x = \text{sign}(x)$ $\qquad$ $T_x = \{z \setminus \text{supp}(z) \subset \text{supp}(x)\}$

Structured sparsity: $J(x) = \sum_b \|x_b\|$ $\qquad$ $\mathcal{N}(a) = a/\|a\|$

$e_x = (\mathcal{N}(x_b))_{b \in \mathcal{B}}$ $\qquad$ $T_x = \{z \setminus \text{supp}(z) \subset \text{supp}(x)\}$

Nuclear norm: $J(x) = \|x\|_*$ $\qquad$ *SVD:* $x = U \Lambda V^*$

$e_x = UV^*$ $\qquad$ $T_x = \left\{UA + BV^* \setminus (A, B) \in (\mathbb{R}^{n \times n})^2\right\}$



Examples

ℓ^1 *sparsity:* $J(x) = \|x\|_1$

$e_x = \text{sign}(x)$ $\quad T_x = \{z \setminus \text{supp}(z) \subset \text{supp}(x)\}$

Structured sparsity: $J(x) = \sum_b \|x_b\|$ $\quad \mathcal{N}(a) = a/\|a\|$

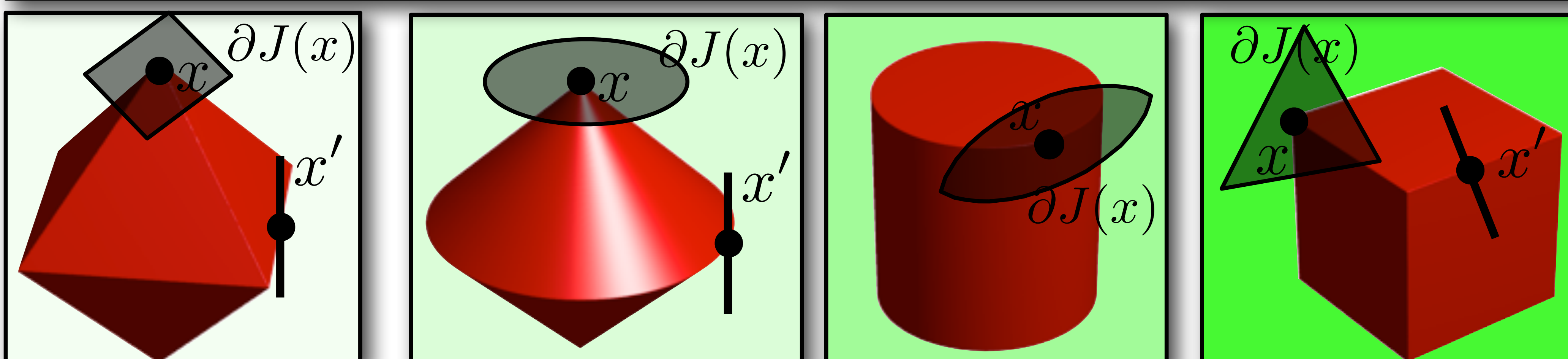
$e_x = (\mathcal{N}(x_b))_{b \in \mathcal{B}}$ $\quad T_x = \{z \setminus \text{supp}(z) \subset \text{supp}(x)\}$

Nuclear norm: $J(x) = \|x\|_*$ $\quad$ *SVD:* $x = U \Lambda V^*$

$e_x = UV^*$ $\quad T_x = \{UA + BV^* \setminus (A,B) \in (\mathbb{R}^{n \times n})^2\}$

Anti-sparsity: $J(x) = \|x\|_\infty$ $\quad I = \{i \setminus |x_i| = \|x\|_\infty\}$

$e_x = |I|^{-1} \text{sign}(x)$ $\quad T_x = \{y \setminus y_I \propto \text{sign}(x_I)\}$



Overview

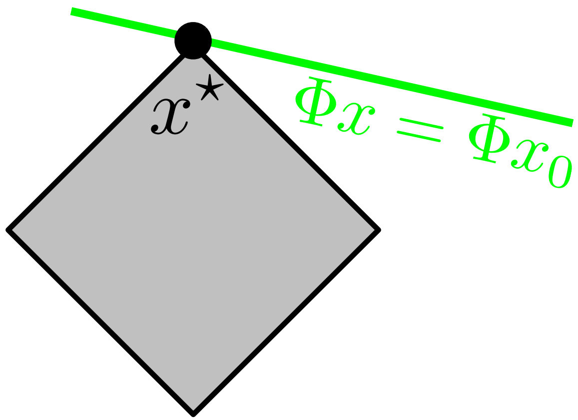
- Low-complexity Regularization with Gauges

- Performance Guarantees

- Grid-free Regularization

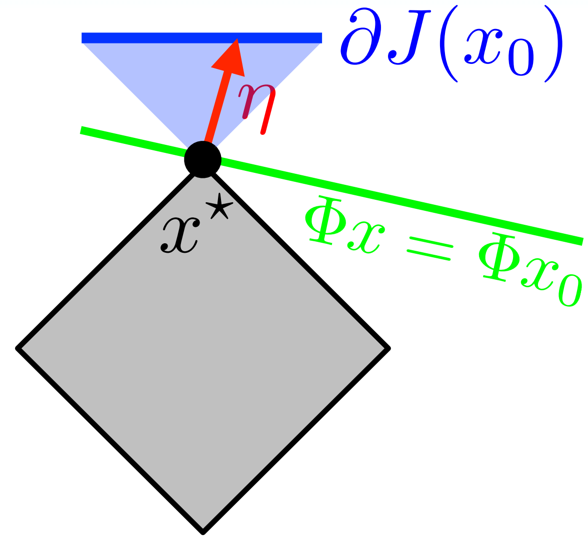
Dual Certificates

Noiseless recovery: $\min_{\Phi x = \Phi x_0} J(x) \qquad (\mathcal{P}_0)$



Dual Certificates

Noiseless recovery: $\min_{\Phi x = \Phi x_0} J(x) \qquad (\mathcal{P}_0)$

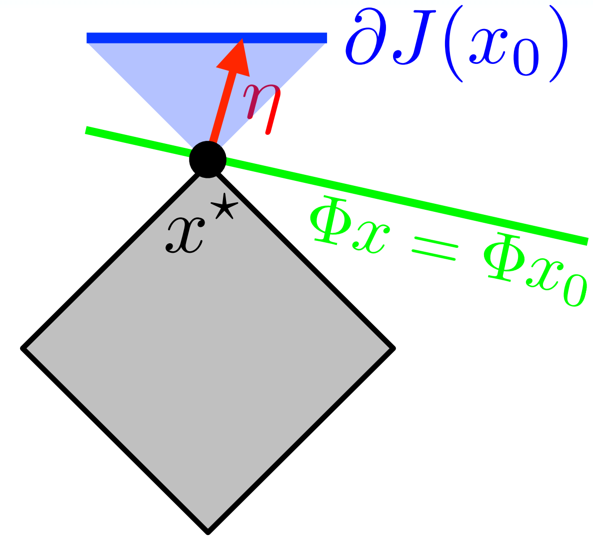


Proposition:
x_0 solution of $(\mathcal{P}_0) \iff \exists\, \eta \in \mathcal{D}(x_0)$

Dual certificates: $\mathcal{D}(x_0) = \mathrm{Im}(\Phi^*) \cap \partial J(x_0)$

Dual Certificates

Noiseless recovery: $\min_{\Phi x = \Phi x_0} J(x) \quad (\mathcal{P}_0)$



Proposition:

x_0 solution of $(\mathcal{P}_0) \iff \exists\, \eta \in \mathcal{D}(x_0)$

Dual certificates: $\mathcal{D}(x_0) = \mathrm{Im}(\Phi^*) \cap \partial J(x_0)$

Proof: $(\mathcal{P}_0) \iff \min_{\delta \in \ker(\Phi)} J(x_0 + \delta)$

$\forall\, (\eta, \delta) \in \partial J(x_0) \times \ker(\Phi), \quad J(x_0 + \delta) \geqslant J(x_0) + \langle \delta,\, \eta \rangle$

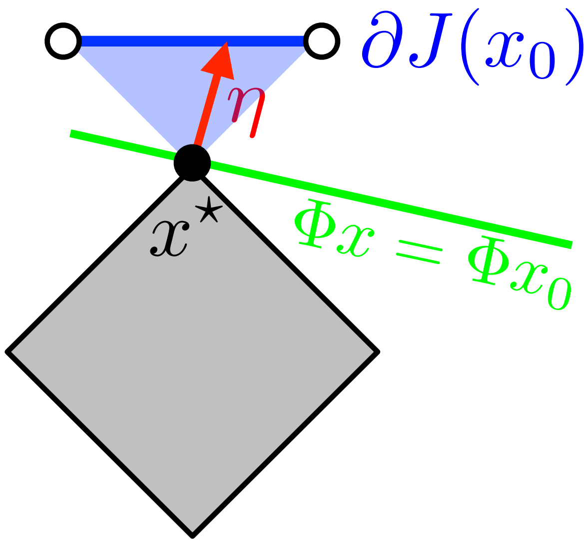
$\eta \in \mathrm{Im}(\Phi^*) \implies \langle \delta,\, \eta \rangle = 0 \implies x_0$ solution.

x_0 solution $\implies \forall\, \delta,\ \langle \delta,\, \eta \rangle \leqslant 0 \implies \eta \in \ker(\Phi)^\perp$.

Dual Certificates and L2 Stability

Tight dual certificates:

$$\bar{\mathcal{D}}(x_0) = \mathrm{Im}(\Phi^*) \cap \mathrm{ri}(\partial J(x_0))$$



$$\mathrm{ri}(E) = \text{relative interior of } E$$

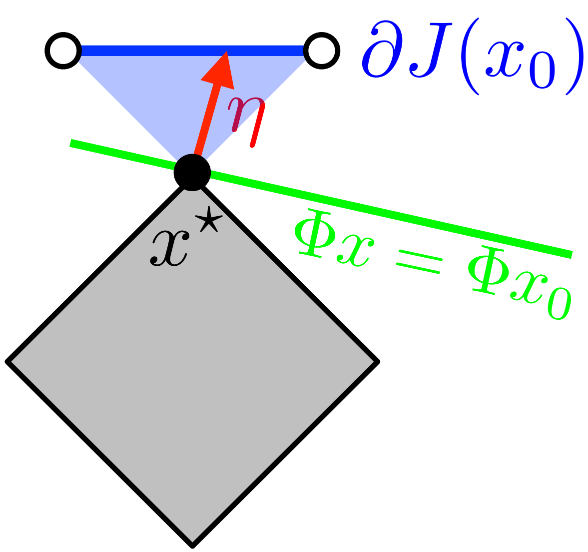
$$= \text{interior for the topology of } \mathrm{aff}(E)$$

Dual Certificates and L2 Stability

Tight dual certificates:

$$\bar{\mathcal{D}}(x_0) = \mathrm{Im}(\Phi^*) \cap \mathrm{ri}(\partial J(x_0))$$

$$\mathrm{ri}(E) = \text{relative interior of } E$$
$$= \text{interior for the topology of } \mathrm{aff}(E)$$



Theorem: [Fadili et al. 2013]

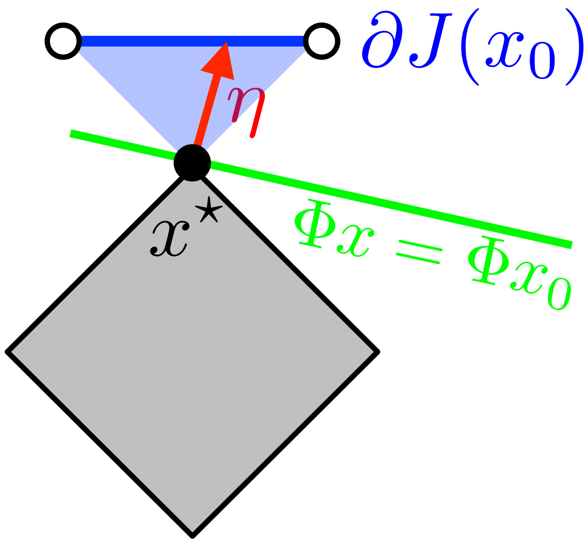
If $\exists\, \eta \in \bar{\mathcal{D}}(x_0)$, for $\lambda \sim \|w\|$ one has $\|x^\star - x_0\| = O(\|w\|)$

Dual Certificates and L2 Stability

Tight dual certificates:

$$\bar{\mathcal{D}}(x_0) = \mathrm{Im}(\Phi^*) \cap \mathrm{ri}(\partial J(x_0))$$

$$\mathrm{ri}(E) = \text{relative interior of } E$$
$$= \text{interior for the topology of } \mathrm{aff}(E)$$



Theorem: [Fadili et al. 2013]

If $\exists\, \eta \in \bar{\mathcal{D}}(x_0)$, for $\lambda \sim \|w\|$ one has $\|x^\star - x_0\| = O(\|w\|)$

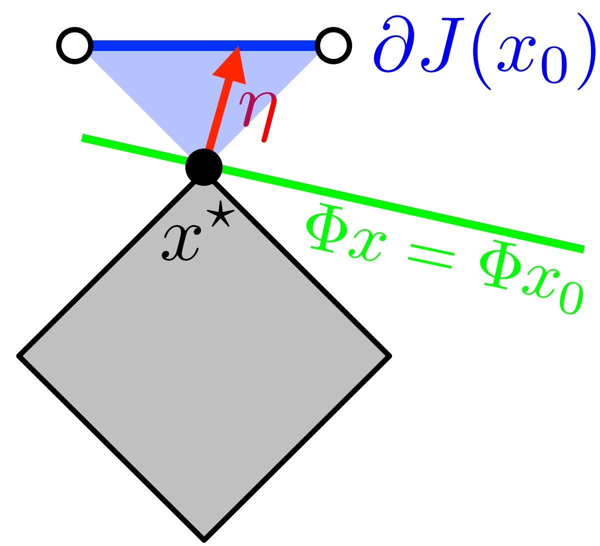
[Grassmair, Haltmeier, Scherzer 2010]: $J = \|\cdot\|_1$.

[Grassmair 2012]: $J(x^\star - x_0) = O(\|w\|)$.

Dual Certificates and L2 Stability

Tight dual certificates:

$$\bar{\mathcal{D}}(x_0) = \mathrm{Im}(\Phi^*) \cap \mathrm{ri}(\partial J(x_0))$$



$$\mathrm{ri}(E) = \text{relative interior of } E$$
$$= \text{interior for the topology of } \mathrm{aff}(E)$$

Theorem: [Fadili et al. 2013]

If $\exists\, \eta \in \bar{\mathcal{D}}(x_0)$, for $\lambda \sim \|w\|$ one has $\|x^\star - x_0\| = O(\|w\|)$

[Grassmair, Haltmeier, Scherzer 2010]: $J = \|\cdot\|_1$.

[Grassmair 2012]: $J(x^\star - x_0) = O(\|w\|)$.

$\longrightarrow$ The constants depend on N ...

Compressed Sensing Setting

Random matrix: $\Phi \in \mathbb{R}^{P \times N}, \quad \Phi_{i,j} \sim \mathcal{N}(0,1)$, i.i.d.

Compressed Sensing Setting

Random matrix: $\Phi \in \mathbb{R}^{P \times N}, \quad \Phi_{i,j} \sim \mathcal{N}(0,1)$, i.i.d.

Sparse vectors: $J = \| \cdot \|_1$.

[Rudelson, Vershynin 2006]
[Chandrasekaran et al. 2011]

Theorem: Let $s = \|x_0\|_0$. If

$$P \geqslant 2s \log(N/s)$$

Then $\exists \eta \in \bar{\mathcal{D}}(x_0)$ with high probability on Φ.

Compressed Sensing Setting

Random matrix: $\Phi \in \mathbb{R}^{P \times N}$, $\Phi_{i,j} \sim \mathcal{N}(0,1)$, i.i.d.

Sparse vectors: $J = \|\cdot\|_1$.

Theorem: Let $s = \|x_0\|_0$. If

[Rudelson, Vershynin 2006]
[Chandrasekaran et al. 2011]

$$P \geqslant 2s \log(N/s)$$

Then $\exists \eta \in \bar{\mathcal{D}}(x_0)$ with high probability on Φ.

Low-rank matrices: $J = \|\cdot\|_*$.

Theorem: Let $r = \text{rank}(x_0)$. If

[Chandrasekaran et al. 2011]

$$P \geqslant 3r(N_1 + N_2 - r)$$

$x_0 \in \mathbb{R}^{N_1 \times N_2}$

Then $\exists \eta \in \bar{\mathcal{D}}(x_0)$ with high probability on Φ.

Compressed Sensing Setting

Random matrix: $\Phi \in \mathbb{R}^{P \times N}$, $\Phi_{i,j} \sim \mathcal{N}(0,1)$, i.i.d.

Sparse vectors: $J = \|\cdot\|_1$.

[Rudelson, Vershynin 2006]
[Chandrasekaran et al. 2011]

Theorem: Let $s = \|x_0\|_0$. If

$$P \geqslant 2s \log(N/s)$$

Then $\exists \eta \in \bar{\mathcal{D}}(x_0)$ with high probability on Φ.

Low-rank matrices: $J = \|\cdot\|_*$.

[Chandrasekaran et al. 2011]

Theorem: Let $r = \text{rank}(x_0)$. If

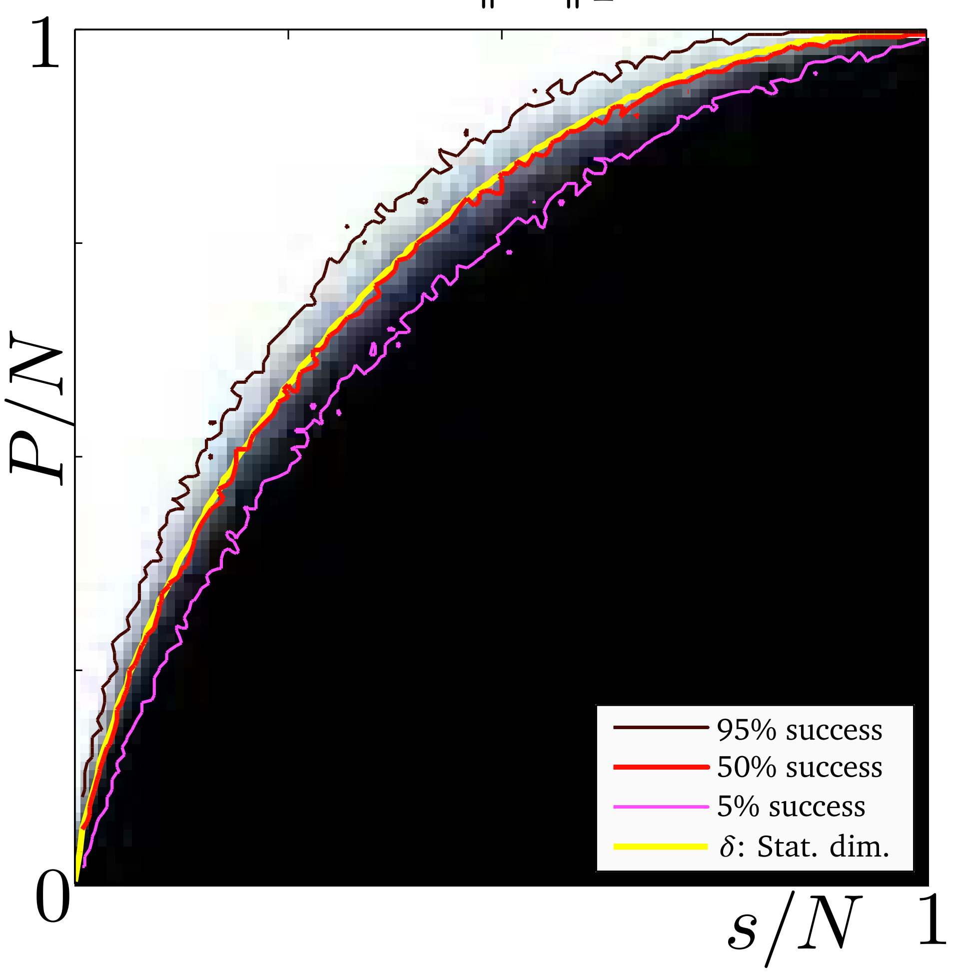
$$P \geqslant 3r(N_1 + N_2 - r) \qquad x_0 \in \mathbb{R}^{N_1 \times N_2}$$

Then $\exists \eta \in \bar{\mathcal{D}}(x_0)$ with high probability on Φ.

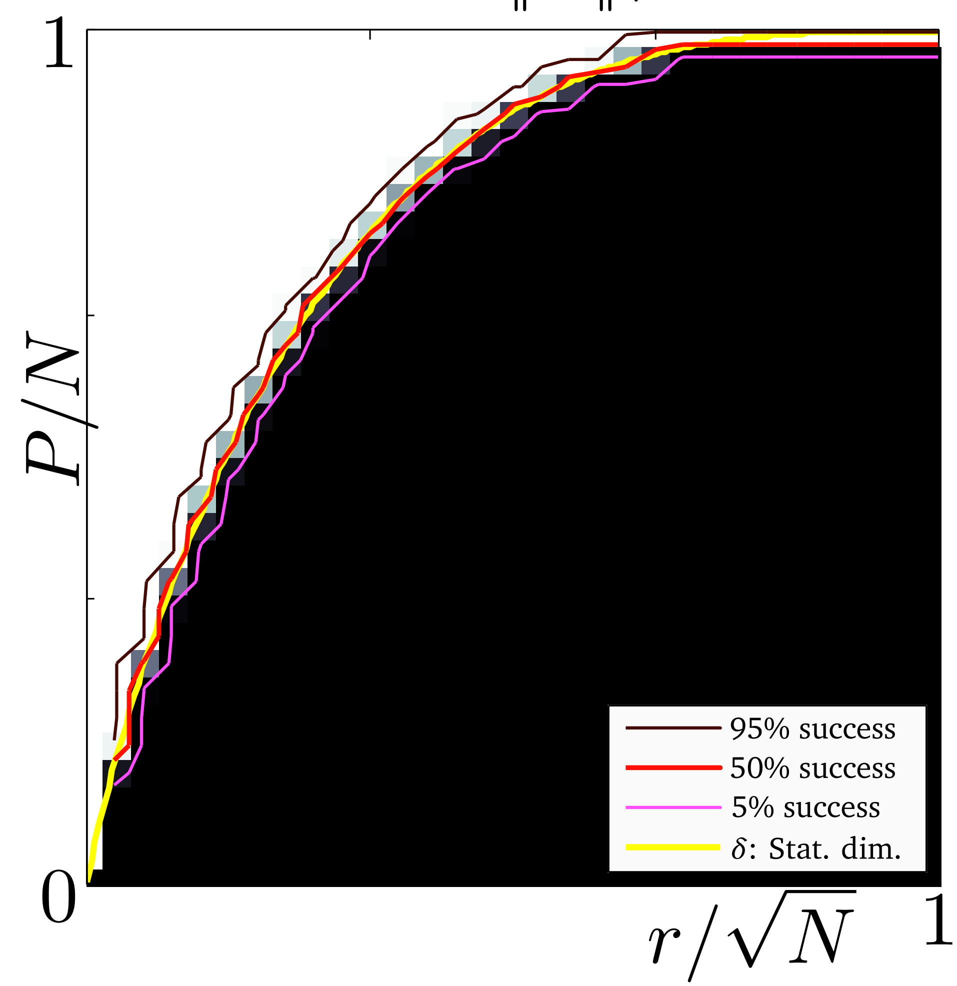
$\longrightarrow$ Similar results for $\|\cdot\|_{1,2}$, $\|\cdot\|_\infty$.

Phase Transitions

$J = \|\cdot\|_1$



$J = \|\cdot\|_*$



From [Amelunxen et al. 20013]

Minimal-norm Certificate

$$\eta \in \mathcal{D}(x_0) \quad \Longrightarrow \quad \begin{cases} \eta = \Phi^* q \\ \text{Proj}_T(\eta) = e \end{cases}$$

$T = T_{x_0}$
$e = e_{x_0}$

Minimal-norm Certificate

$$\eta \in \mathcal{D}(x_0) \quad \Longrightarrow \quad \begin{cases} \eta = \Phi^* q \\ \text{Proj}_T(\eta) = e \end{cases}$$

$T = T_{x_0}$
$e = e_{x_0}$

Minimal-norm pre-certificate: $\eta_0 = \underset{\eta = \Phi^* q, \eta_T = e}{\text{argmin}} \|q\|$

Minimal-norm Certificate

$$\eta \in \mathcal{D}(x_0) \quad \Longrightarrow \quad \begin{cases} \eta = \Phi^* q \\ \text{Proj}_T(\eta) = e \end{cases}$$

$T = T_{x_0}$
$e = e_{x_0}$

Minimal-norm pre-certificate: $\eta_0 = \underset{\eta = \Phi^* q, \eta_T = e}{\text{argmin}} \; \|q\|$

Proposition: One has $\eta_0 = (\Phi_T^+ \Phi)^* e$

$$\Phi_T = \Phi \circ \text{Proj}_T$$

Minimal-norm Certificate

$$\eta \in \mathcal{D}(x_0) \quad \Longrightarrow \quad \begin{cases} \eta = \Phi^* q \\ \mathrm{Proj}_T(\eta) = e \end{cases}$$

$T = T_{x_0}$
$e = e_{x_0}$

Minimal-norm pre-certificate: $\eta_0 = \underset{\eta=\Phi^* q, \eta_T = e}{\operatorname{argmin}} \|q\|$

Proposition: One has $\eta_0 = (\Phi_T^+ \Phi)^* e$

$\Phi_T = \Phi \circ \mathrm{Proj}_T$

Theorem: If $\eta_0 \in \bar{\mathcal{D}}(x_0)$ and $\lambda \sim \|w\|$,

the unique solution $x^\star$ of $\mathcal{P}_\lambda(y)$ for $y = \Phi x_0 + w$ satisfies

$T_{x^\star} = T_{x_0}$ and $\|x^\star - x_0\| = O(\|w\|)$ [Vaiter et al. 2013]

Minimal-norm Certificate

$$\eta \in \mathcal{D}(x_0) \quad \Longrightarrow \quad \begin{cases} \eta = \Phi^* q \\ \text{Proj}_T(\eta) = e \end{cases} \qquad \begin{matrix} T = T_{x_0} \\ e = e_{x_0} \end{matrix}$$

Minimal-norm pre-certificate: $\eta_0 = \underset{\eta = \Phi^* q, \eta_T = e}{\text{argmin}} \|q\|$

Proposition: One has $\eta_0 = (\Phi_T^+ \Phi)^* e$

$$\Phi_T = \Phi \circ \text{Proj}_T$$

Theorem: If $\eta_0 \in \bar{\mathcal{D}}(x_0)$ and $\lambda \sim \|w\|$,

the unique solution $x^\star$ of $\mathcal{P}_\lambda(y)$ for $y = \Phi x_0 + w$ satisfies

$$T_{x^\star} = T_{x_0} \quad \text{and} \quad \|x^\star - x_0\| = O(\|w\|)$$ [Vaiter et al. 2013]

[Fuchs 2004]: $J = \|\cdot\|_1$. [Vaiter et al. 2011]: $J = \|D^* \cdot\|_1$.

[Bach 2008]: $J = \|\cdot\|_{1,2}$ and $J = \|\cdot\|_*$.

Compressed Sensing Setting

Random matrix: $\Phi \in \mathbb{R}^{P \times N}, \quad \Phi_{i,j} \sim \mathcal{N}(0,1)$, i.i.d.

Sparse vectors: $J = \|\cdot\|_1$.

[Wainwright 2009]
[Dossal et al. 2011]

Theorem: Let $s = \|x_0\|_0$. If

$$P > 2s \log(N)$$

Then $\eta_0 \in \bar{\mathcal{D}}(x_0)$ with high probability on Φ.

Compressed Sensing Setting

Random matrix: $\Phi \in \mathbb{R}^{P \times N}$, $\Phi_{i,j} \sim \mathcal{N}(0,1)$, i.i.d.

Sparse vectors: $J = \|\cdot\|_1$.

Theorem: Let $s = \|x_0\|_0$. If

[Wainwright 2009]
[Dossal et al. 2011]

$$P > 2s\log(N)$$

Then $\eta_0 \in \bar{\mathcal{D}}(x_0)$ with high probability on Φ.

Phase transitions:

L^2 stability	vs.	Model stability
$P \sim 2s\log(N/s)$		$P \sim 2s\log(N)$

Compressed Sensing Setting

Random matrix: $\Phi \in \mathbb{R}^{P \times N}$, $\Phi_{i,j} \sim \mathcal{N}(0,1)$, i.i.d.

Sparse vectors: $J = \|\cdot\|_1$.

Theorem: Let $s = \|x_0\|_0$. If

[Wainwright 2009]
[Dossal et al. 2011]

$$P > 2s\log(N)$$

Then $\eta_0 \in \bar{\mathcal{D}}(x_0)$ with high probability on Φ.

Phase transitions:

L^2 stability
$P \sim 2s\log(N/s)$

vs.

Model stability
$P \sim 2s\log(N)$

$\longrightarrow$ Similar results for $\|\cdot\|_{1,2}$, $\|\cdot\|_*$, $\|\cdot\|_\infty$.

Compressed Sensing Setting

Random matrix: $\Phi \in \mathbb{R}^{P \times N}$, $\Phi_{i,j} \sim \mathcal{N}(0,1)$, i.i.d.

Sparse vectors: $J = \|\cdot\|_1$.

Theorem: Let $s = \|x_0\|_0$. If

[Wainwright 2009]
[Dossal et al. 2011]

$$P > 2s \log(N)$$

Then $\eta_0 \in \bar{\mathcal{D}}(x_0)$ with high probability on Φ.

Phase transitions:

L^2 stability
$P \sim 2s \log(N/s)$

vs.

Model stability
$P \sim 2s \log(N)$

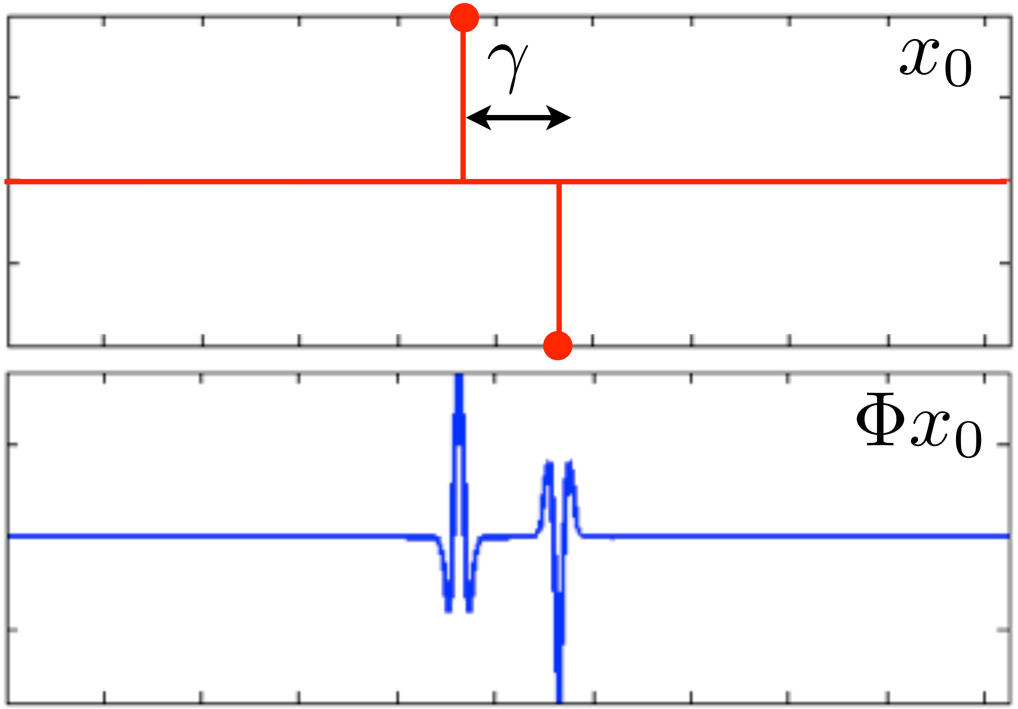
$\longrightarrow$ Similar results for $\|\cdot\|_{1,2}, \|\cdot\|_*, \|\cdot\|_\infty$.

$\longrightarrow$ Not using RIP technics (non-uniform result on x_0).

1-D Sparse Spikes Deconvolution

$$\Phi x = \sum_i x_i \varphi(\cdot - \Delta i)$$

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



Increasing Δ:

$\rightarrow$ reduces correlation.

$\rightarrow$ reduces resolution.

1-D Sparse Spikes Deconvolution

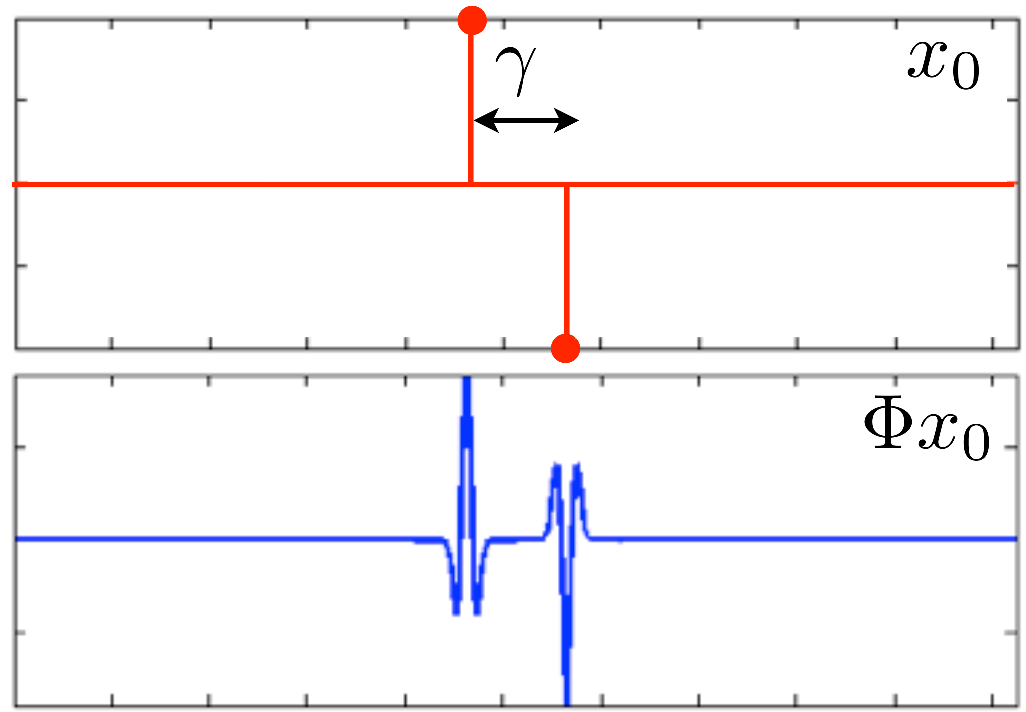
$$\Phi x = \sum_i x_i \varphi(\cdot - \Delta i)$$

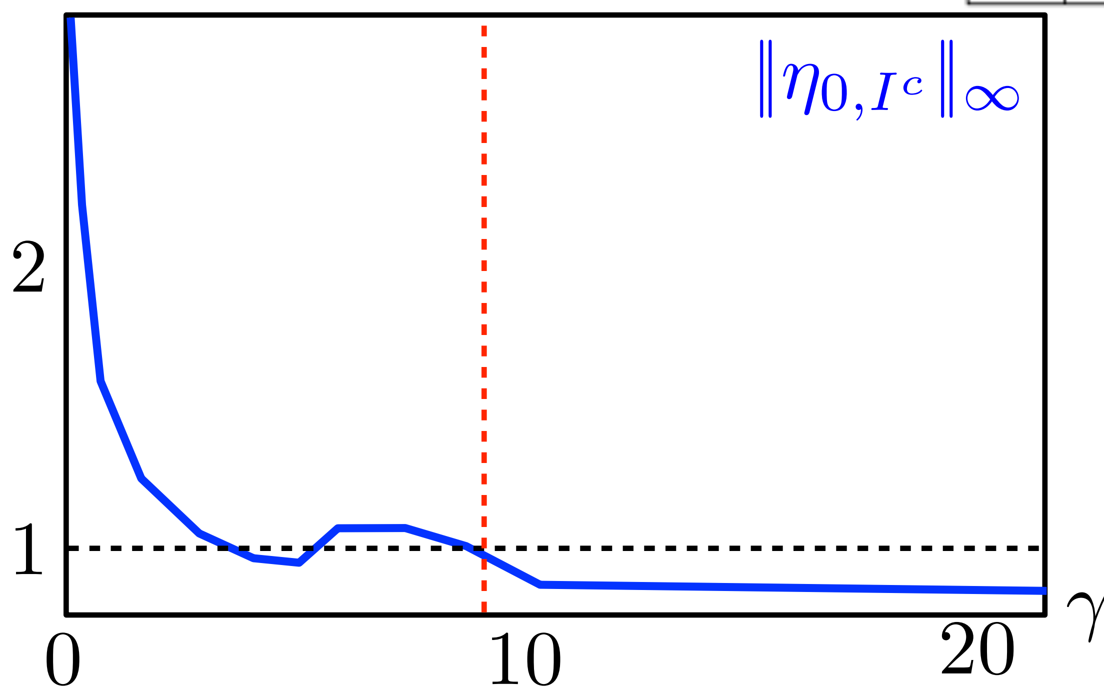
$$J(x) = \|x\|_1$$

Increasing Δ:

$\longrightarrow$ reduces correlation.

$\longrightarrow$ reduces resolution.





$$I = \{ j \setminus x_0(j) \neq 0 \}$$

$$\|\eta_{0,I^c}\|_\infty < 1$$

$$\Longleftrightarrow$$

$$\eta_0 \in \bar{\mathcal{D}}(x_0)$$

$$\Longleftrightarrow$$

support recovery.

Overview

- Low-complexity Regularization with Gauges

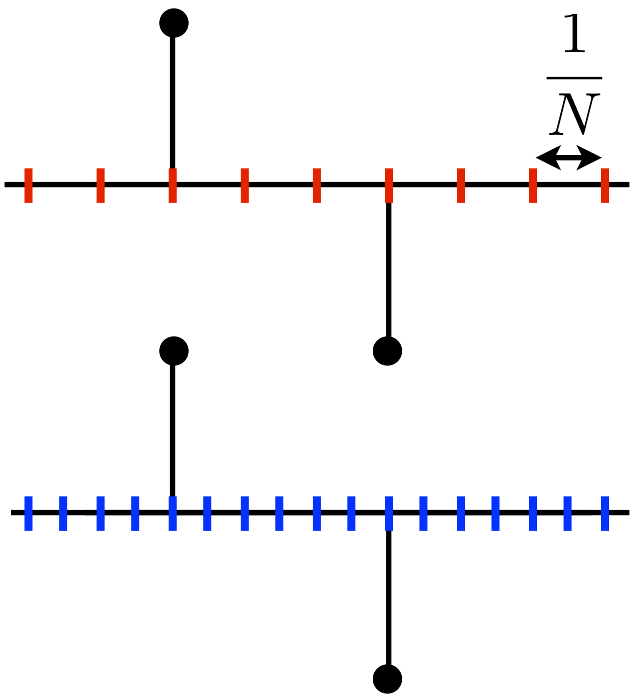
- Performance Guarantees

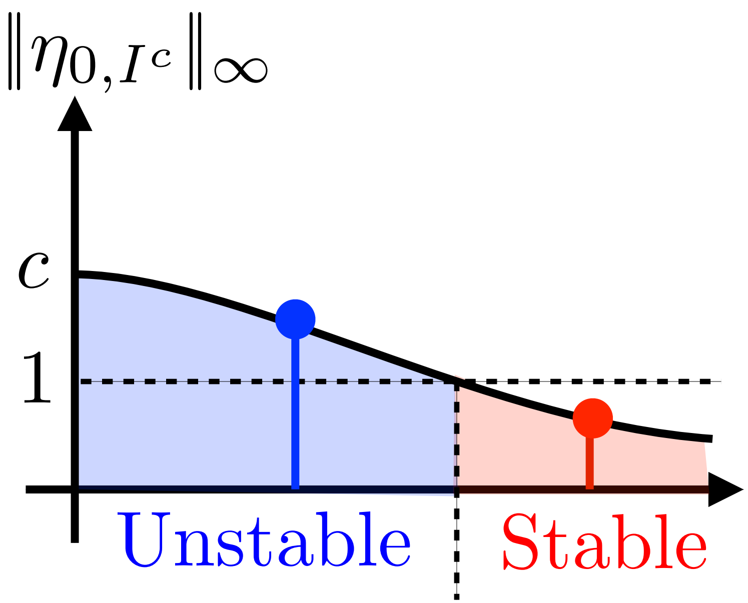
- **Grid-free Regularization**

Support Instability and Measures

When $N \to +\infty$, support is not stable:

$$\|\eta_{0,I^c}\|_\infty \underset{N \to +\infty}{\longrightarrow} c > 1.$$

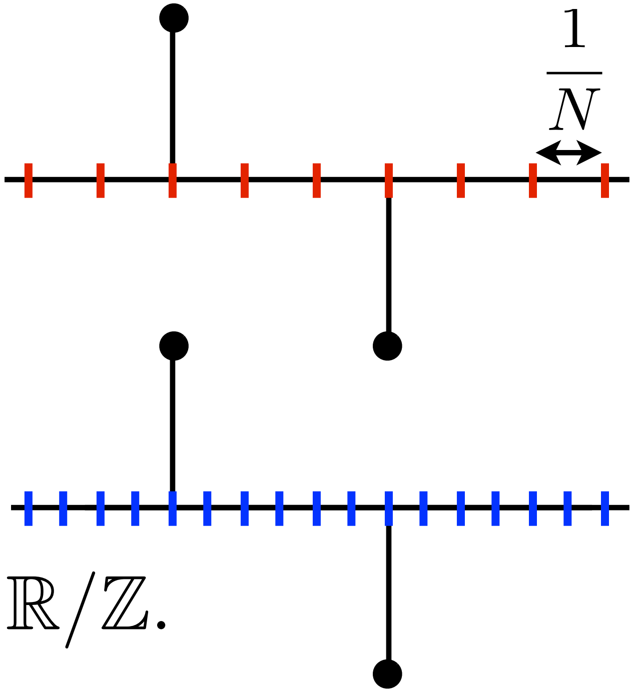




Support Instability and Measures

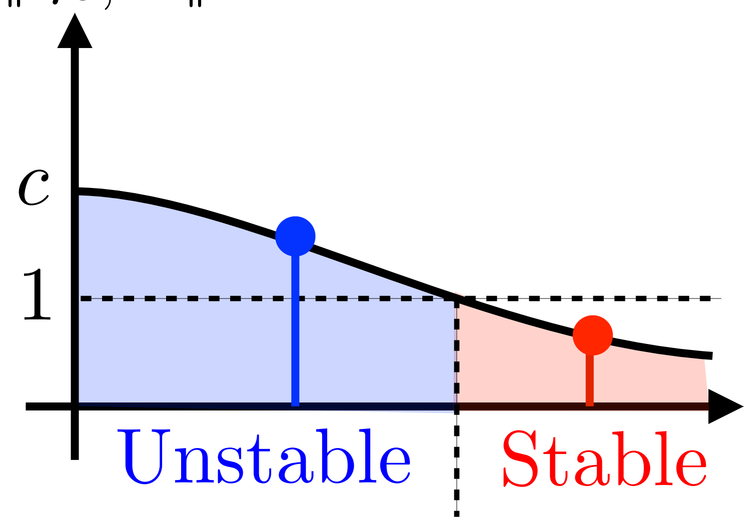
When $N \to +\infty$, support is not stable:

$$\|\eta_{0,I^c}\|_\infty \underset{N \to +\infty}{\longrightarrow} c > 1.$$



Intuition: spikes wants to move laterally.

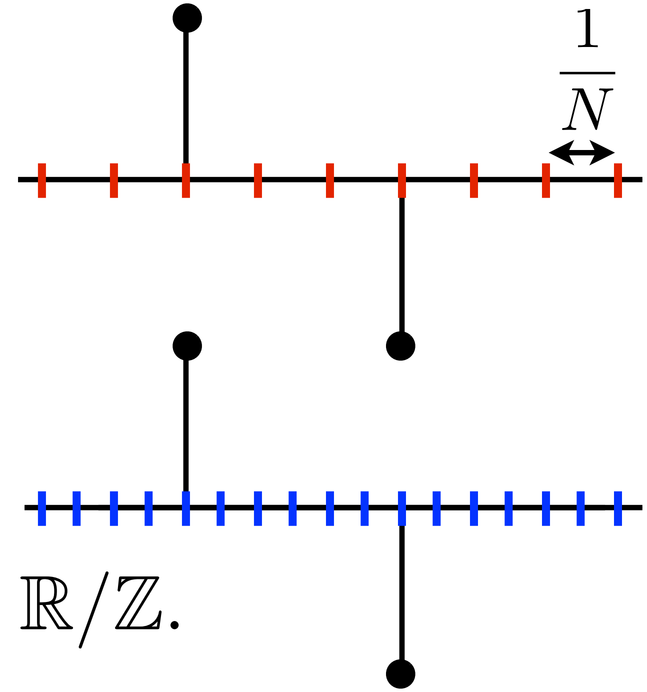
$\longrightarrow$ Use Radon measures $m \in \mathcal{M}(\mathbb{T})$, $\mathbb{T} = \mathbb{R}/\mathbb{Z}$.



Support Instability and Measures

When $N \to +\infty$, support is not stable:

$$\|\eta_{0,I^c}\|_\infty \underset{N \to +\infty}{\longrightarrow} c > 1.$$

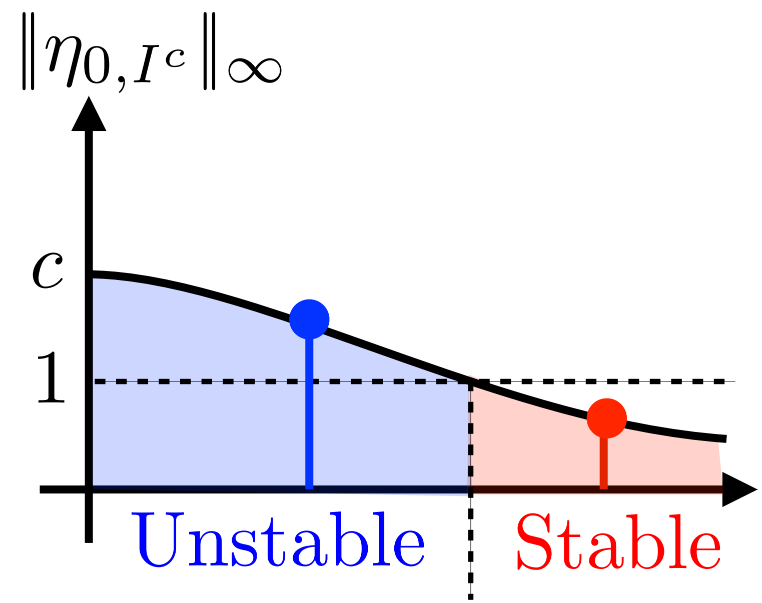


Intuition: spikes wants to move laterally.

$\longrightarrow$ Use Radon measures $m \in \mathcal{M}(\mathbb{T})$, $\mathbb{T} = \mathbb{R}/\mathbb{Z}$.

Extension of ℓ^1: total variation

$$\|m\|_{\text{TV}} = \sup_{\|g\|_\infty \leqslant 1} \int_{\mathbb{T}} g(x)\,\mathrm{d}m(x)$$



Discrete measure: $m_{x,a} = \sum_i a_i \delta_{x_i}$.

One has $\|m_{x,a}\|_{\text{TV}} = \|a\|_1$

Sparse Measure Regularization

Measurements: $y = \Phi(m_0) + w$ where $\begin{cases} m_0 \in \mathcal{M}(\mathbb{T}), \\ \Phi : \mathcal{M}(\mathbb{T}) \to \mathrm{L}^2(\mathbb{T}), \\ w \in \mathrm{L}^2(\mathbb{T}). \end{cases}$

Sparse Measure Regularization

Measurements: $y = \Phi(m_0) + w$ where $\begin{cases} m_0 \in \mathcal{M}(\mathbb{T}), \\ \Phi : \mathcal{M}(\mathbb{T}) \to \mathrm{L}^2(\mathbb{T}), \\ w \in \mathrm{L}^2(\mathbb{T}). \end{cases}$

Acquisition operator:

$$\Phi(m)(x) = \int_{\mathbb{T}} \varphi(x, x') \mathrm{d}m(x') \quad \text{where} \quad \varphi \in C^2(\mathbb{T} \times \mathbb{T})$$

Sparse Measure Regularization

Measurements: $y = \Phi(m_0) + w$ where $\begin{cases} m_0 \in \mathcal{M}(\mathbb{T}), \\ \Phi : \mathcal{M}(\mathbb{T}) \to \mathrm{L}^2(\mathbb{T}), \\ w \in \mathrm{L}^2(\mathbb{T}). \end{cases}$

Acquisition operator:

$$\Phi(m)(x) = \int_{\mathbb{T}} \varphi(x, x') \mathrm{d}m(x') \quad \text{where} \quad \varphi \in C^2(\mathbb{T} \times \mathbb{T})$$

Total-variation over measures regularization:

$$\min_{m \in \mathcal{M}(\mathbb{T})} \frac{1}{2} \|\Phi(m) - y\|^2 + \lambda \|m\|_{\mathrm{TV}}$$

Sparse Measure Regularization

Measurements: $y = \Phi(m_0) + w$ where $\begin{cases} m_0 \in \mathcal{M}(\mathbb{T}), \\ \Phi : \mathcal{M}(\mathbb{T}) \to \mathrm{L}^2(\mathbb{T}), \\ w \in \mathrm{L}^2(\mathbb{T}). \end{cases}$

Acquisition operator:

$$\Phi(m)(x) = \int_{\mathbb{T}} \varphi(x, x')\mathrm{d}m(x') \quad \text{where} \quad \varphi \in C^2(\mathbb{T} \times \mathbb{T})$$

Total-variation over measures regularization:

$$\min_{m \in \mathcal{M}(\mathbb{T})} \frac{1}{2}\|\Phi(m) - y\|^2 + \lambda\|m\|_{\mathrm{TV}}$$

$\longrightarrow$ Infinite dimensional convex program.

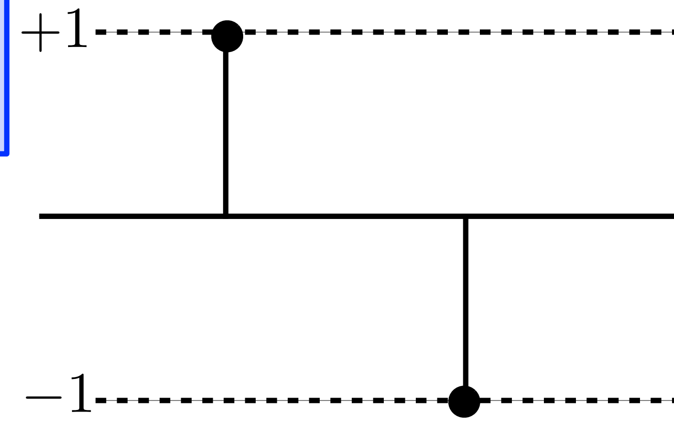
$\longrightarrow$ If $\dim(\mathrm{Im}(\Phi)) < +\infty$, dual is finite dimensional.

$\longrightarrow$ If Φ is a filtering, re-cast dual as SDP program.

Fuchs vs. Vanishing Pre-Certificates

Measures:

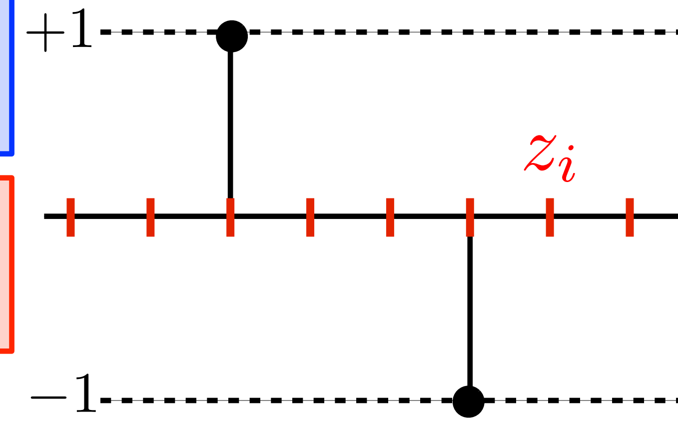
$$\min_{m \in \mathcal{M}} \frac{1}{2}\|\Phi m - y\|^2 + \lambda \|m\|_{\text{TV}}$$



Fuchs vs. Vanishing Pre-Certificates

Measures: $$\min_{m \in \mathcal{M}} \frac{1}{2}\|\Phi m - y\|^2 + \lambda \|m\|_{\mathrm{TV}}$$

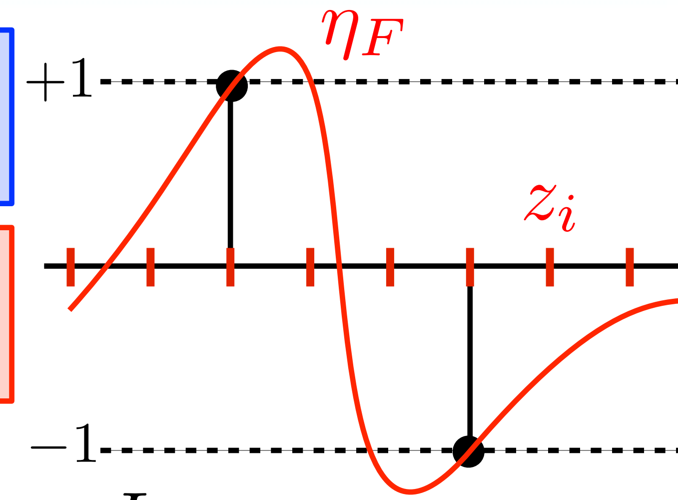
On a grid z: $$\min_{a \in \mathbb{R}^N} \frac{1}{2}\|\Phi_z a - y\|^2 + \lambda \|a\|_1$$



Fuchs vs. Vanishing Pre-Certificates

Measures: $\min_{m \in \mathcal{M}} \frac{1}{2}\|\Phi m - y\|^2 + \lambda \|m\|_{\mathrm{TV}}$

On a grid z: $\min_{a \in \mathbb{R}^N} \frac{1}{2}\|\Phi_z a - y\|^2 + \lambda \|a\|_1$



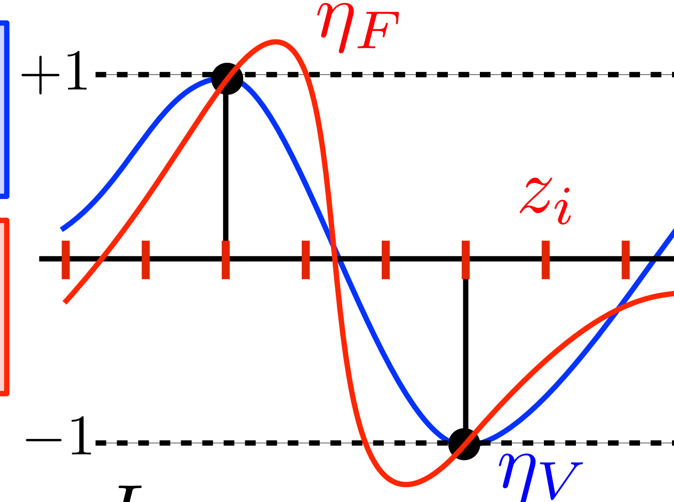
For $m_0 = m_{z,a_0}$, $\mathrm{supp}(m_0) = x_0$, $\mathrm{supp}(a_0) = I$:

$$\eta_F = \Phi^* \Phi_I^{*,+} \mathrm{sign}(a_{0,I})$$

Fuchs vs. Vanishing Pre-Certificates

Measures: $\min_{m \in \mathcal{M}} \frac{1}{2}\|\Phi m - y\|^2 + \lambda \|m\|_{\text{TV}}$

On a grid z: $\min_{a \in \mathbb{R}^N} \frac{1}{2}\|\Phi_z a - y\|^2 + \lambda \|a\|_1$



For $m_0 = m_{z,a_0}$, $\text{supp}(m_0) = x_0$, $\text{supp}(a_0) = I$:

$$\eta_F = \Phi^* \Phi_I^{*,+} \text{sign}(a_{0,I})$$

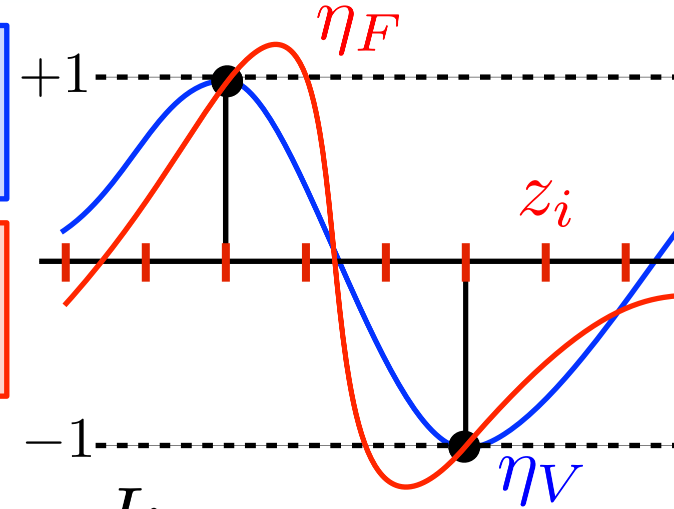
$$\eta_V = \Phi^* \Gamma_{x_0}^{+,*} \left(\text{sign}(a_0), 0\right)^*$$

where $\Gamma_x(a,b) = \sum_i a_i \varphi(\cdot, x_i) + b_i \varphi'(\cdot, x_i)$

Fuchs vs. Vanishing Pre-Certificates

Measures: $\min_{m \in \mathcal{M}} \frac{1}{2}\|\Phi m - y\|^2 + \lambda \|m\|_{\text{TV}}$

On a grid z: $\min_{a \in \mathbb{R}^N} \frac{1}{2}\|\Phi_z a - y\|^2 + \lambda \|a\|_1$



For $m_0 = m_{z,a_0}$, $\text{supp}(m_0) = x_0$, $\text{supp}(a_0) = I$:

$$\eta_F = \Phi^* \Phi_I^{*,+} \text{sign}(a_{0,I})$$

$$\eta_V = \Phi^* \Gamma_{x_0}^{+,*} \left(\text{sign}(a_0), 0\right)^*$$

where $\Gamma_x(a,b) = \sum_i a_i \varphi(\cdot, x_i) + b_i \varphi'(\cdot, x_i)$

Theorem: [Fuchs 2004]

If $\forall\, j \notin I, |\eta_F(x_j)| < 1$,

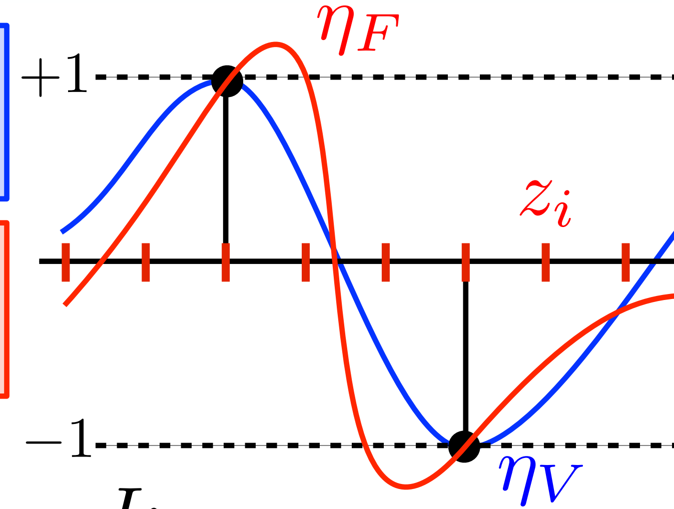
then $\text{supp}(a_\lambda) = \text{supp}(a_0)$

(holds for $\|w\|$ small enough and $\lambda \sim \|w\|$)

Fuchs vs. Vanishing Pre-Certificates

Measures: $\min_{m \in \mathcal{M}} \frac{1}{2}\|\Phi m - y\|^2 + \lambda \|m\|_{\text{TV}}$

On a grid z: $\min_{a \in \mathbb{R}^N} \frac{1}{2}\|\Phi_z a - y\|^2 + \lambda \|a\|_1$



For $m_0 = m_{z,a_0}$, $\text{supp}(m_0) = x_0$, $\text{supp}(a_0) = I$:

$$\eta_F = \Phi^* \Phi_I^{*,+} \, \text{sign}(a_{0,I})$$

$$\eta_V = \Phi^* \Gamma_{x_0}^{+,*} \left(\text{sign}(a_0), 0\right)^*$$

where $\Gamma_x(a,b) = \sum_i a_i \varphi(\cdot, x_i) + b_i \varphi'(\cdot, x_i)$

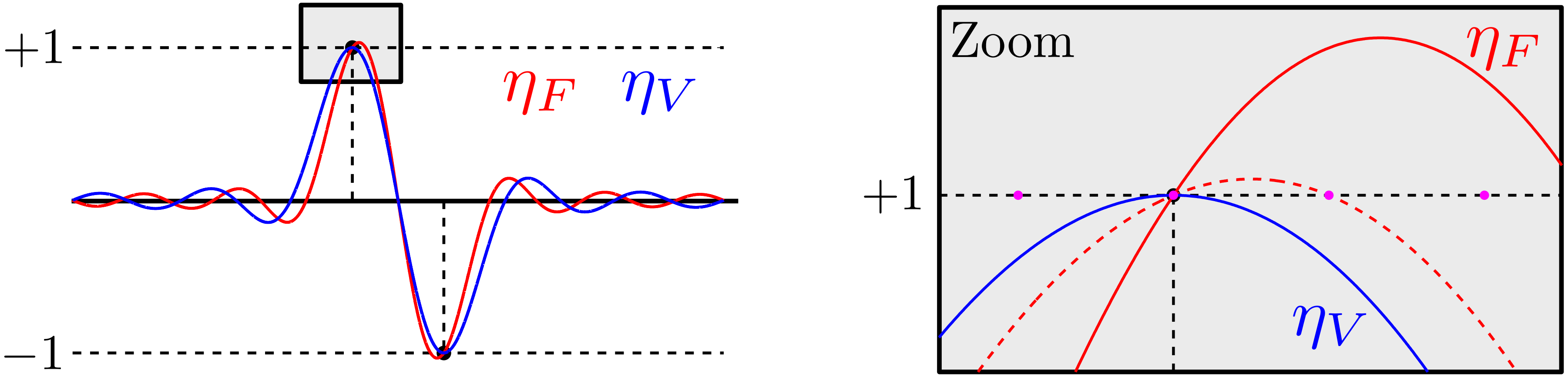
Theorem: [Fuchs 2004]
If $\forall\, j \notin I, |\eta_F(x_j)| < 1$,
then $\text{supp}(a_\lambda) = \text{supp}(a_0)$

Theorem: [Duval-Peyré 2013]
If $\forall\, t \notin x_0, |\eta_V(t)| < 1$,
then $m_\lambda = m_{x_\lambda, a_\lambda}$ with
$\|x_\lambda - x_0\|_\infty = O(\|w\|)$

(holds for $\|w\|$ small enough and $\lambda \sim \|w\|$)

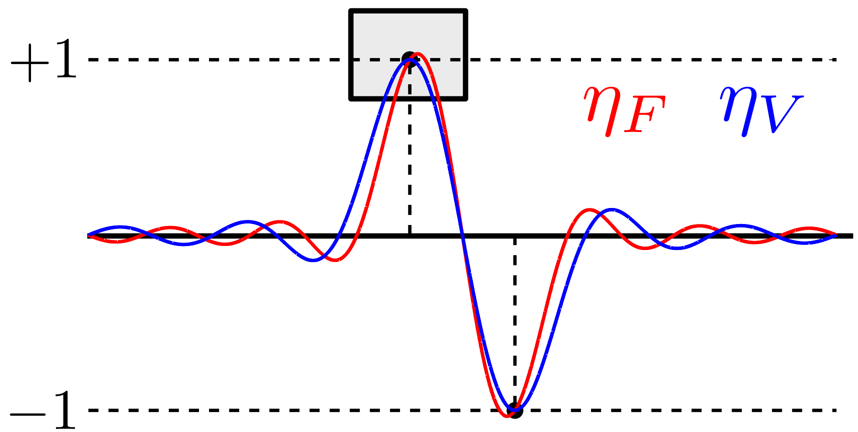
Numerical Illustration

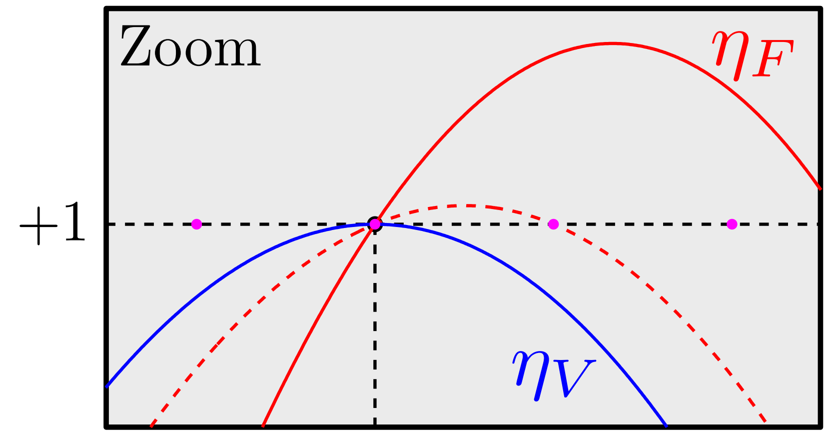
Ideal low-pass filter: $\varphi(x, x') = \frac{\sin((2f_c+1)\pi(x-x'))}{\sin(\pi(x-x'))}$, $f_c = 6$.

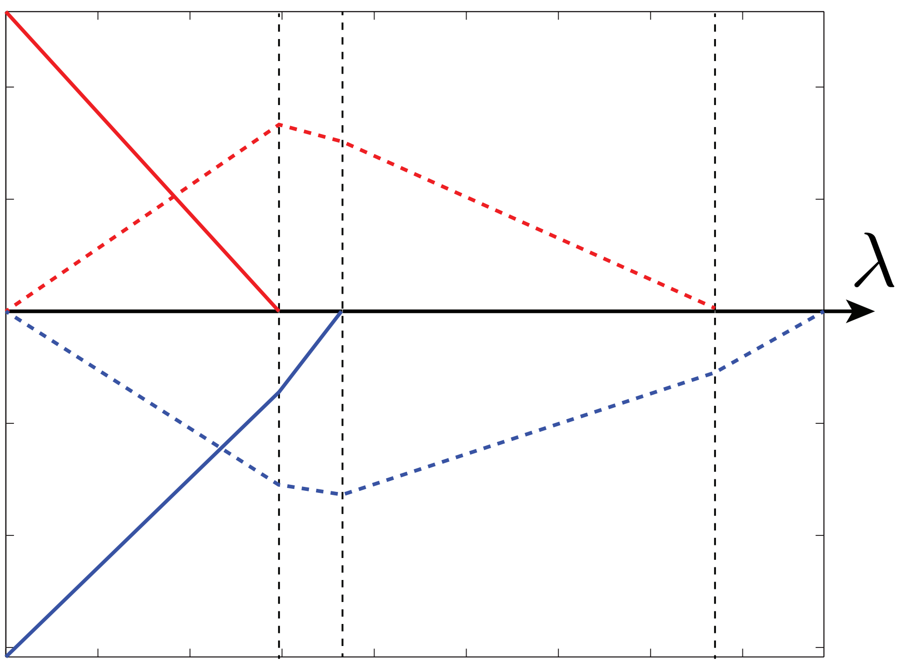


Numerical Illustration

Ideal low-pass filter: $\varphi(x,x') = \frac{\sin((2f_c+1)\pi(x-x'))}{\sin(\pi(x-x'))}$, $f_c = 6$.



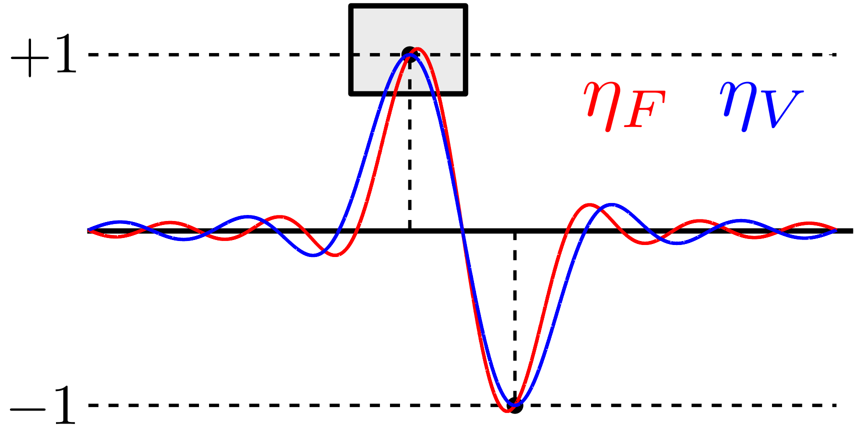


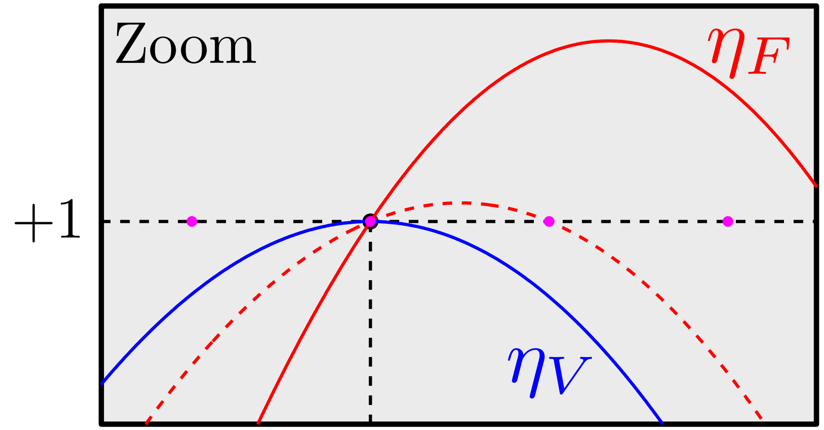


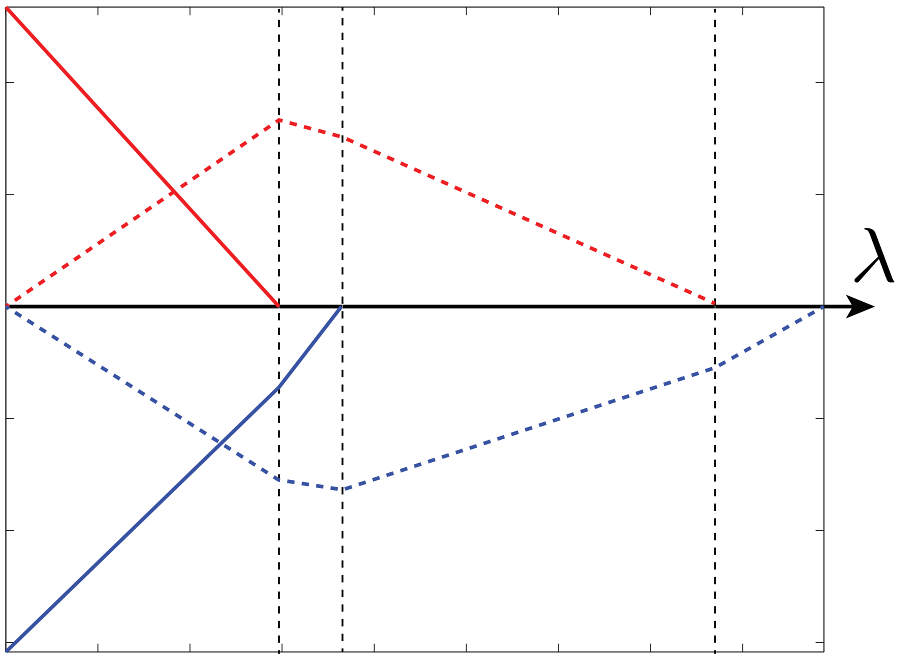
Solution path $\lambda \mapsto a_\lambda$

Numerical Illustration

Ideal low-pass filter: $\varphi(x,x') = \frac{\sin((2f_c+1)\pi(x-x'))}{\sin(\pi(x-x'))}$, $f_c = 6$.







Solution path $\lambda \mapsto a_\lambda$

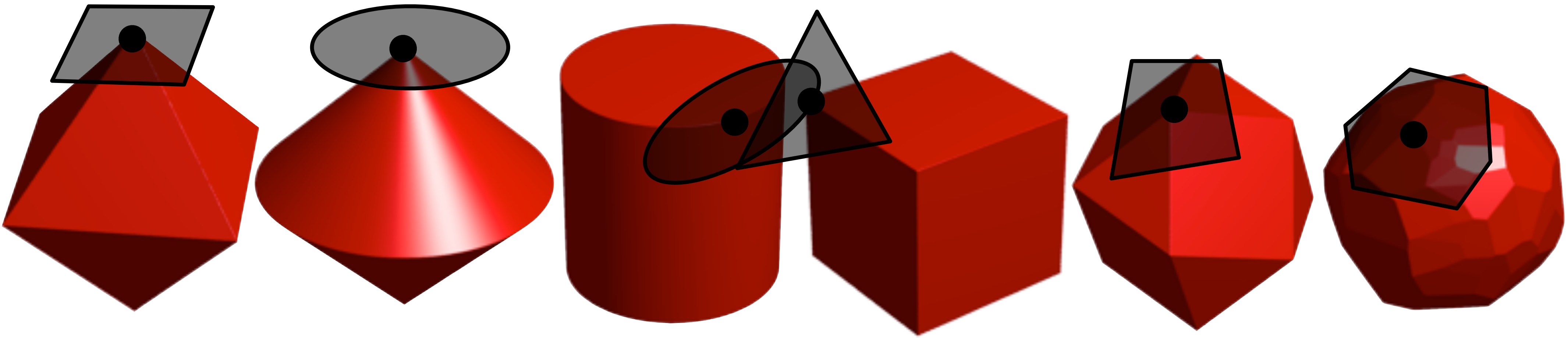
Discrete → continuous:

Theorem: [Duval-Peyré 2013]
If η_V is valid, then a_λ is supported on pairs of neighbors around $\mathrm{supp}(m_0)$.

(holds for $\lambda \sim \|w\|$ small enough.

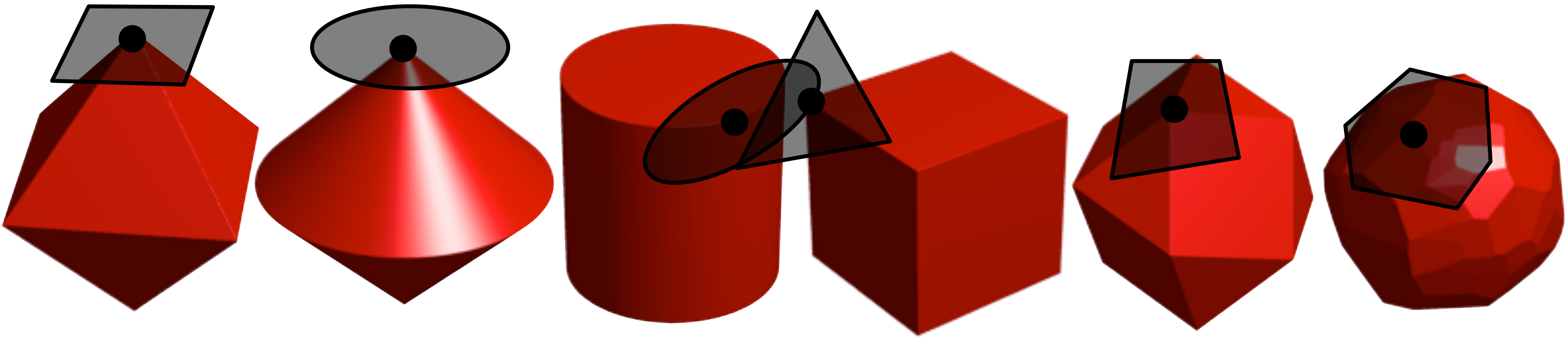
Conclusion

Gauges: encode linear models as singular points.



Conclusion

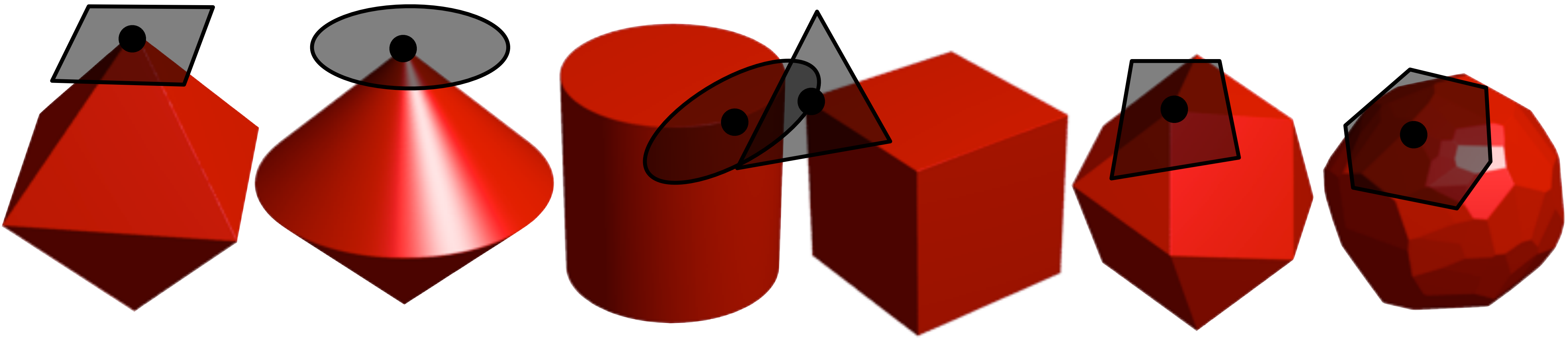
Gauges: encode linear models as singular points.



Performance measures → L^2 error / model → different CS guarantees

Conclusion

Gauges: encode linear models as singular points.

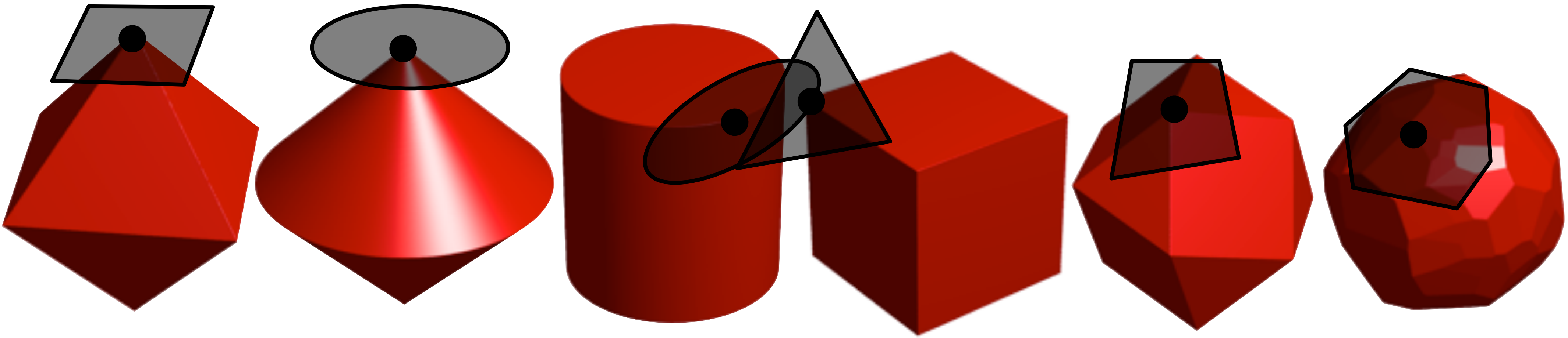


Performance measures → L^2 error / model → different CS guarantees

model → Specific certificate: $\eta_0, \eta_F, \eta_V, \ldots$

Conclusion

Gauges: encode linear models as singular points.



Performance measures → L^2 error / **model** → different CS guarantees

model → Specific certificate: $\eta_0, \eta_F, \eta_V, \ldots$

Open problems:

– CS performance with arbitrary gauges.

– Approximate model recovery $T_{x^\star} \approx T_{x_0}$.
(e.g. grid-free recovery)