Robust Sparse Analysis Recovery

Gabriel Peyré

Joint work with:
Samuel Vaiter
Charles Dossal
Jalal Fadili



CEREMADE
DAUPHINE UNIVERSITÉ PARIS

www.numerical-tours.com

Overview

- Synthesis vs. Analysis Regularization
- Risk Estimation
- Local Behavior of Sparse Regularization
- Robustness to Noise
- Numerical Illustrations

Inverse Problems

Recovering $x_0 \in \mathbb{R}^N$ from noisy observations

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

$\Phi : \mathbb{R}^N \mapsto \mathbb{R}^P$ with $P \ll N$ (missing information)

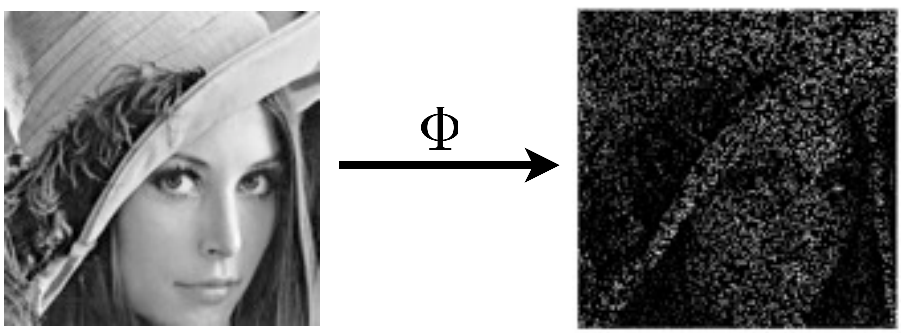
Inverse Problems

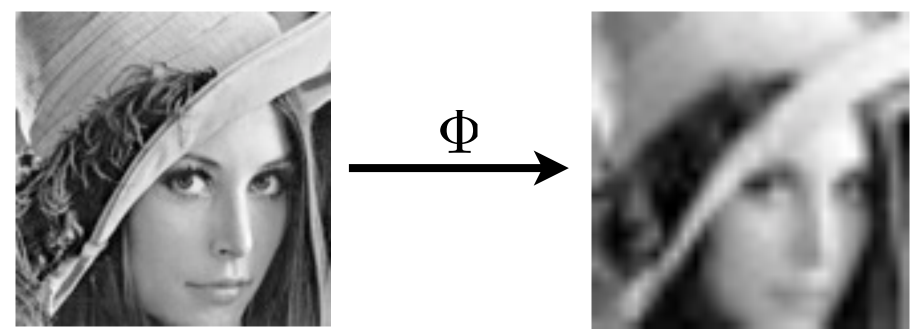
Recovering $x_0 \in \mathbb{R}^N$ from noisy observations

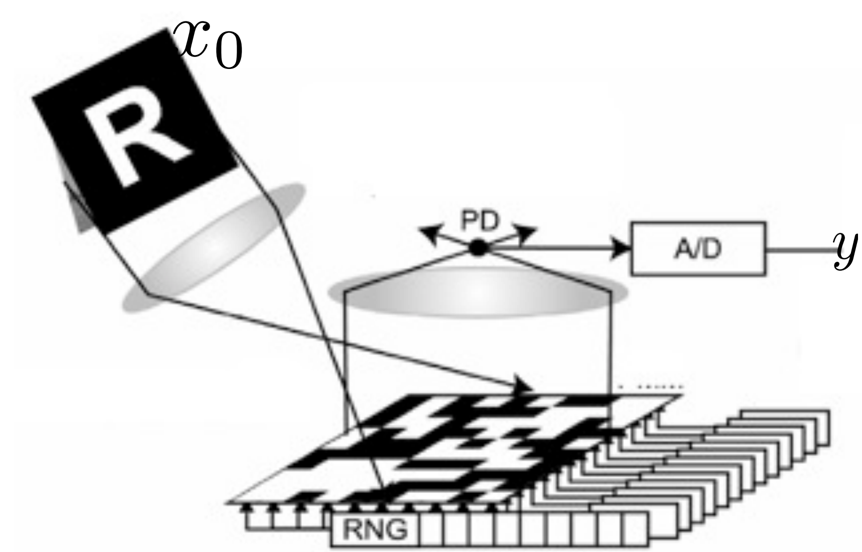
$$y = \Phi x_0 + w \in \mathbb{R}^P$$

$\Phi : \mathbb{R}^N \mapsto \mathbb{R}^P$ with $P \ll N$ (missing information)

Examples: Inpainting, super-resolution, compressed-sensing







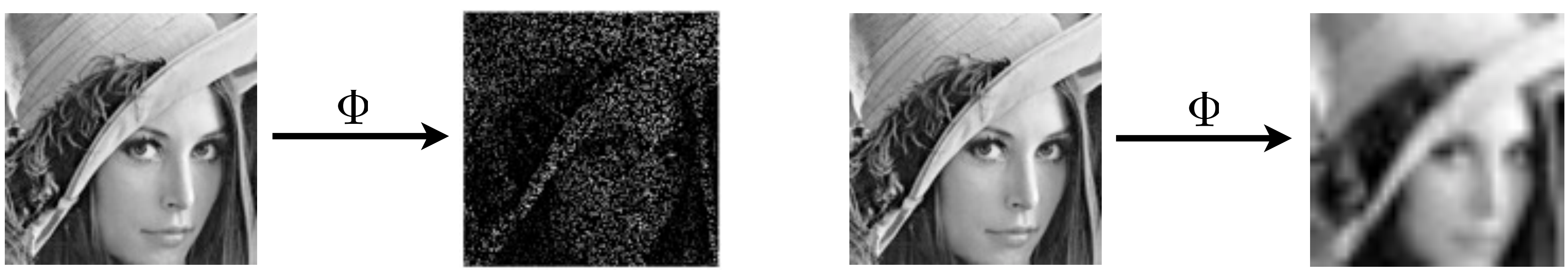
Inverse Problems

Recovering $x_0 \in \mathbb{R}^N$ from noisy observations

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

$\Phi : \mathbb{R}^N \mapsto \mathbb{R}^P$ with $P \ll N$ (missing information)

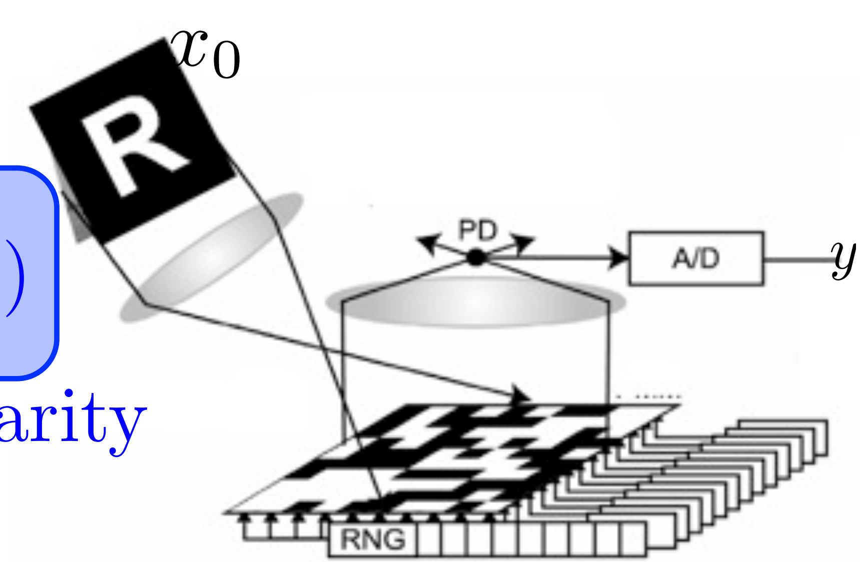
Examples: Inpainting, super-resolution, compressed-sensing



Regularized inversion:

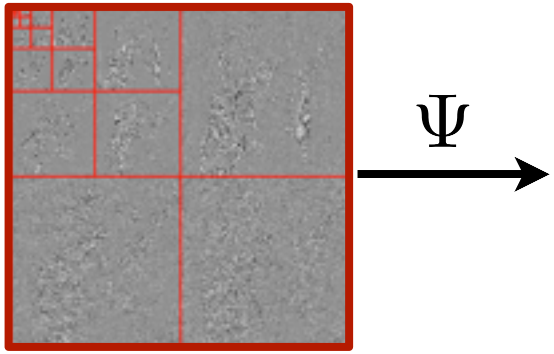
$$x_\lambda(y) \in \underset{x \in \mathbb{R}^N}{\operatorname{argmin}} \; \frac{1}{2}\|y - \Phi x\|^2 + \lambda \, J(x)$$

Data fidelity — Regularity



Sparse Regularizations

Synthesis regularization



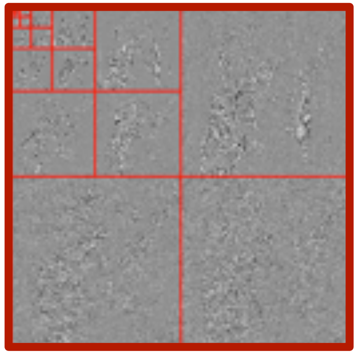
Coefficients α Image $x = \Psi\alpha$

$$\min_{\alpha \in \mathbb{R}^Q} \frac{1}{2}\|y - \Phi\Psi\alpha\|_2^2 + \lambda\|\alpha\|_1$$

Sparse Regularizations

Synthesis regularization	Analysis regularization

Synthesis regularization



$\xrightarrow{\Psi}$

Coefficients α — Image $x = \Psi\alpha$

$$\min_{\alpha \in \mathbb{R}^Q} \frac{1}{2}\|y - \Phi\Psi\alpha\|_2^2 + \lambda\|\alpha\|_1$$

Analysis regularization

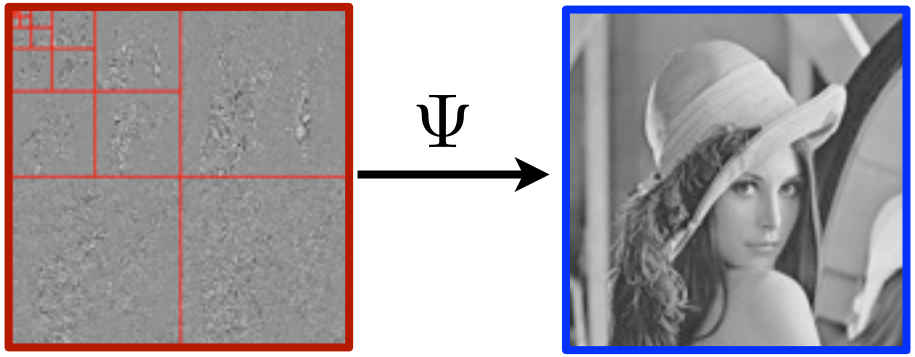
$\xrightarrow{D^*}$

Image x — Correlations $\alpha = D^* x$

$$\min_{x \in \mathbb{R}^N} \frac{1}{2}\|y - \Phi x\|_2^2 + \lambda\|D^* x\|_1$$

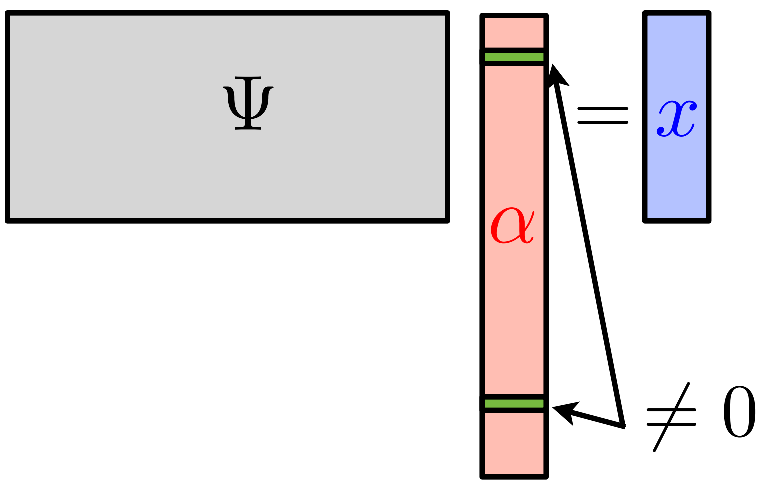
Sparse Regularizations

Synthesis regularization



Coefficients α Image $x = \Psi\alpha$

$$\min_{\alpha \in \mathbb{R}^Q} \frac{1}{2}\|y - \Phi\Psi\alpha\|_2^2 + \lambda\|\alpha\|_1$$



Analysis regularization

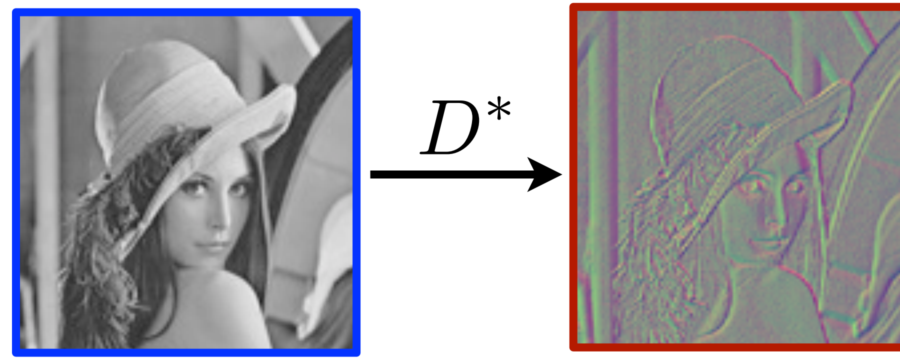
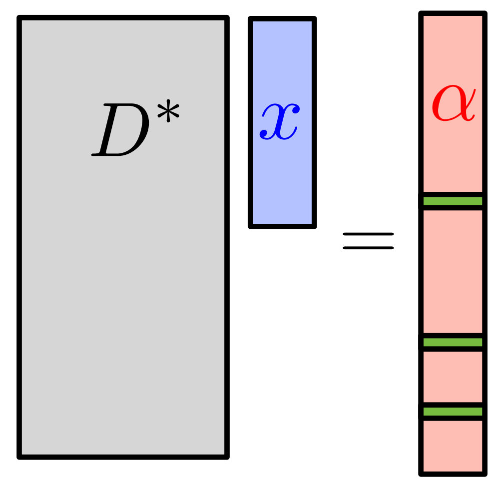


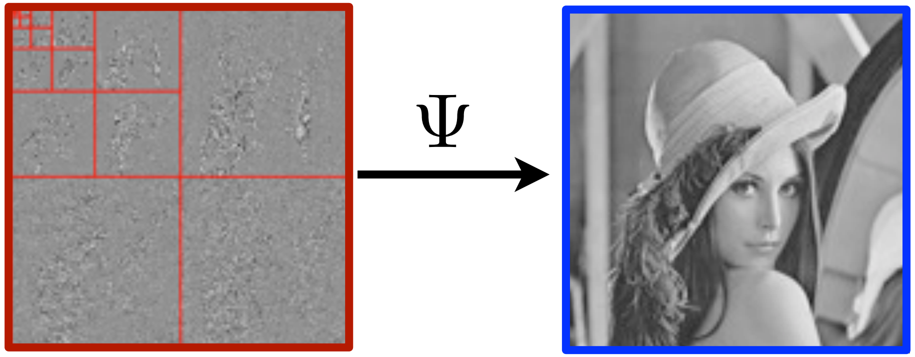
Image x Correlations $\alpha = D^* x$

$$\min_{x \in \mathbb{R}^N} \frac{1}{2}\|y - \Phi x\|_2^2 + \lambda\|D^* x\|_1$$



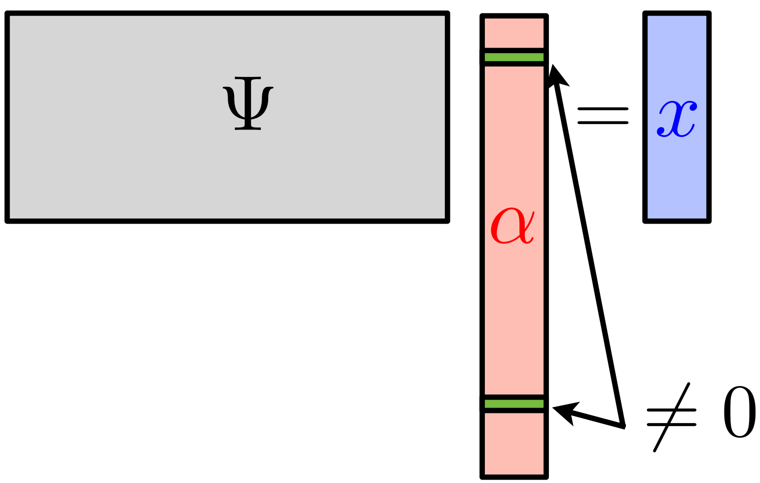
Sparse Regularizations

Synthesis regularization



Coefficients α Image $x = \Psi\alpha$

$$\min_{\alpha \in \mathbb{R}^Q} \frac{1}{2}\|y - \Phi\Psi\alpha\|_2^2 + \lambda\|\alpha\|_1$$



Analysis regularization

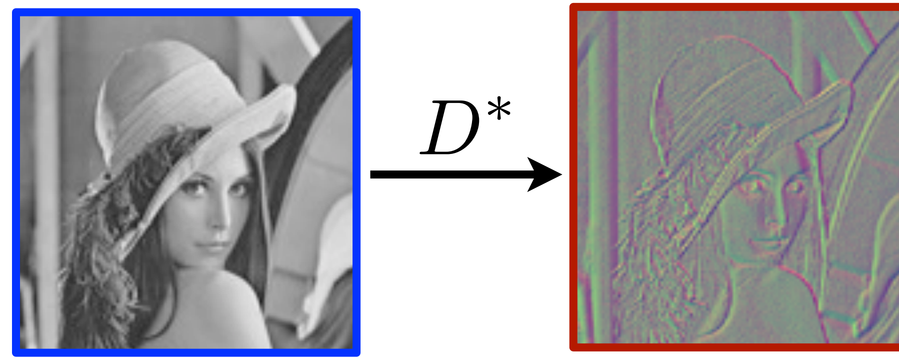
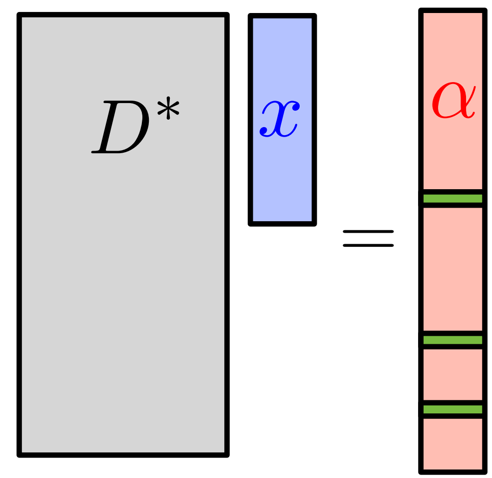


Image x Correlations $\alpha = D^* x$

$$\min_{x \in \mathbb{R}^N} \frac{1}{2}\|y - \Phi x\|_2^2 + \lambda\|D^* x\|_1$$



Unless $D = \Psi$ is orthogonal, produces different results.

Variations and Stability

Observations: $y = \Phi x_0 + w$

Recovery:

$$x_\lambda(y) \in \underset{x \in \mathbb{R}^N}{\text{argmin}} \frac{1}{2}\|\Phi x - y\|^2 + \lambda \|D^* x\|_1 \quad (\mathcal{P}_\lambda(y))$$

$\lambda \to 0^+$

$$x_{0^+}(y) \in \underset{\Phi x = y}{\text{argmin}} \|D^* x\|_1 \quad \text{(no noise)} \quad (\mathcal{P}_0(y))$$

Variations and Stability

Observations: $y = \Phi x_0 + w$

Recovery:

$$x_\lambda(y) \in \underset{x \in \mathbb{R}^N}{\text{argmin}} \frac{1}{2}\|\Phi x - y\|^2 + \lambda \|D^* x\|_1 \quad (\mathcal{P}_\lambda(y))$$

$\lambda \to 0^+$

$$x_{0^+}(y) \in \underset{\Phi x = y}{\text{argmin}} \|D^* x\|_1 \quad \text{(no noise)} \quad (\mathcal{P}_0(y))$$

Questions:

– Behavior of $x_\lambda(y)$ with respect to y and λ.

$\longrightarrow$ *Application:* risk estimation (SURE, GCV, etc.)

Variations and Stability

Observations: $y = \Phi x_0 + w$

Recovery:

$$x_\lambda(y) \in \underset{x \in \mathbb{R}^N}{\text{argmin}} \frac{1}{2}\|\Phi x - y\|^2 + \lambda \|D^* x\|_1 \quad (\mathcal{P}_\lambda(y))$$

$\lambda \to 0^+$

$$x_{0^+}(y) \in \underset{\Phi x = y}{\text{argmin}} \|D^* x\|_1 \quad \text{(no noise)} \quad (\mathcal{P}_0(y))$$

Questions:

– Behavior of $x_\lambda(y)$ with respect to y and λ.

$\longrightarrow$ *Application:* risk estimation (SURE, GCV, etc.)

– Criteria to ensure $\|x_\lambda(y) - x_0\| \leqslant C\|w\|$

(with "reasonable" C)

Variations and Stability

Observations: $y = \Phi x_0 + w$

Recovery:

$$x_\lambda(y) \in \underset{x \in \mathbb{R}^N}{\text{argmin}} \frac{1}{2}\|\Phi x - y\|^2 + \lambda \|D^* x\|_1 \quad (\mathcal{P}_\lambda(y))$$

$\lambda \to 0^+$

$$x_{0^+}(y) \in \underset{\Phi x = y}{\text{argmin}} \|D^* x\|_1 \quad \text{(no noise)} \quad (\mathcal{P}_0(y))$$

Questions:

– Behavior of $x_\lambda(y)$ with respect to y and λ.

$\longrightarrow$ *Application:* risk estimation (SURE, GCV, etc.)

– Criteria to ensure $\|x_\lambda(y) - x_0\| \leqslant C\|w\|$

(with "reasonable" C)

Synthesis case ($D = \text{Id}$): works of Fuchs and Tropp.

Analysis case: [Nam et al. 2011] for $w = 0$.

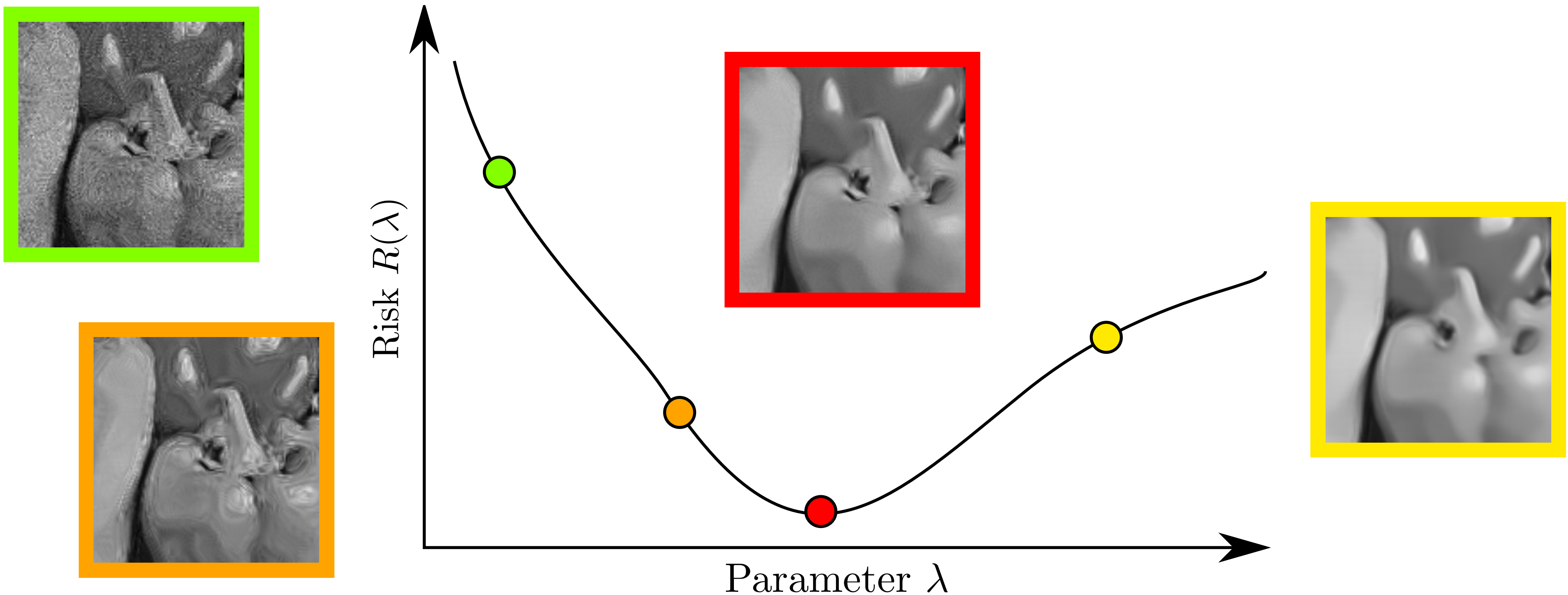
Overview

- Synthesis vs. Analysis Regularization
- **Risk Estimation**
- Local Behavior of Sparse Regularization
- Robustness to Noise
- Numerical Illustrations

Risk Minimization

Average risk: $R(\lambda) = \mathbb{E}_w(\|x_\lambda(y) - x_0\|^2)$

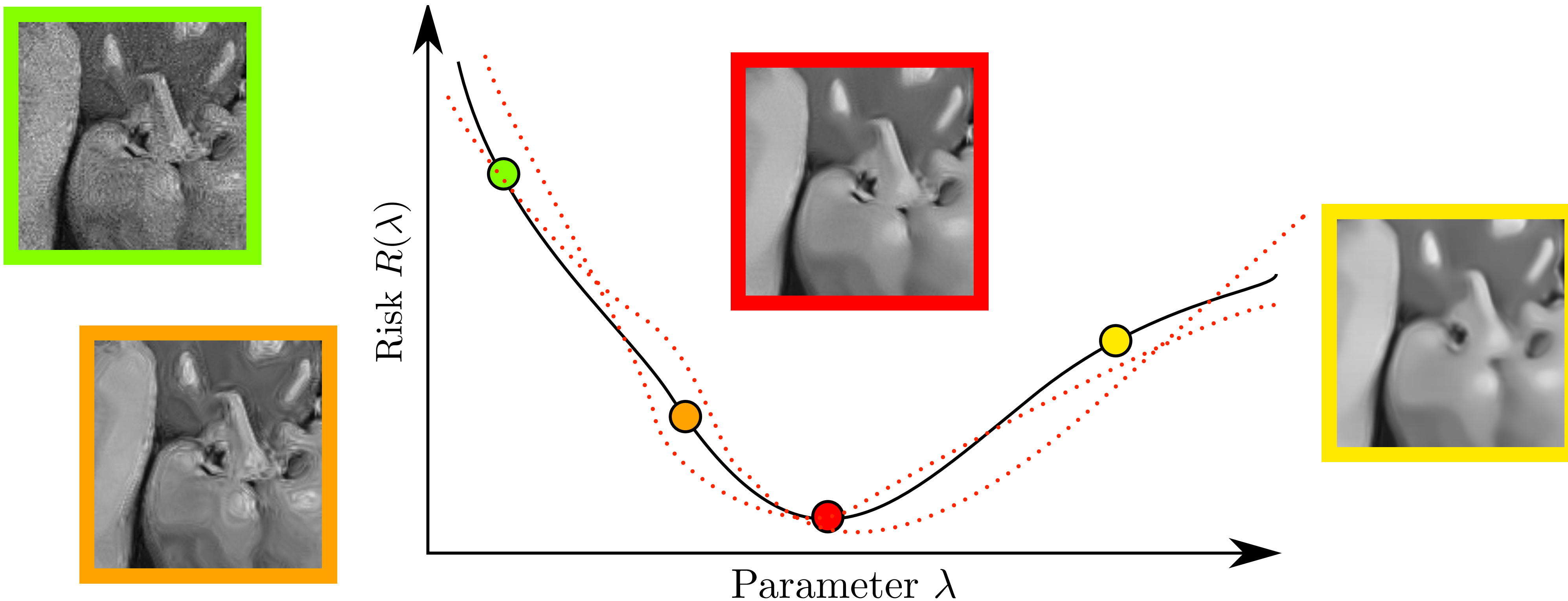
$\lambda^\star(y) = \underset{\lambda}{\text{argmin}}\ R(\lambda)$ *Plugin-estimator:* $x_{\lambda^\star(y)}(y)$



Risk Minimization

Average risk: $R(\lambda) = \mathbb{E}_w(\|x_\lambda(y) - x_0\|^2)$

$$\lambda^\star(y) = \underset{\lambda}{\text{argmin}}\ R(\lambda) \qquad \textit{Plugin-estimator: } x_{\lambda^\star(y)}(y)$$

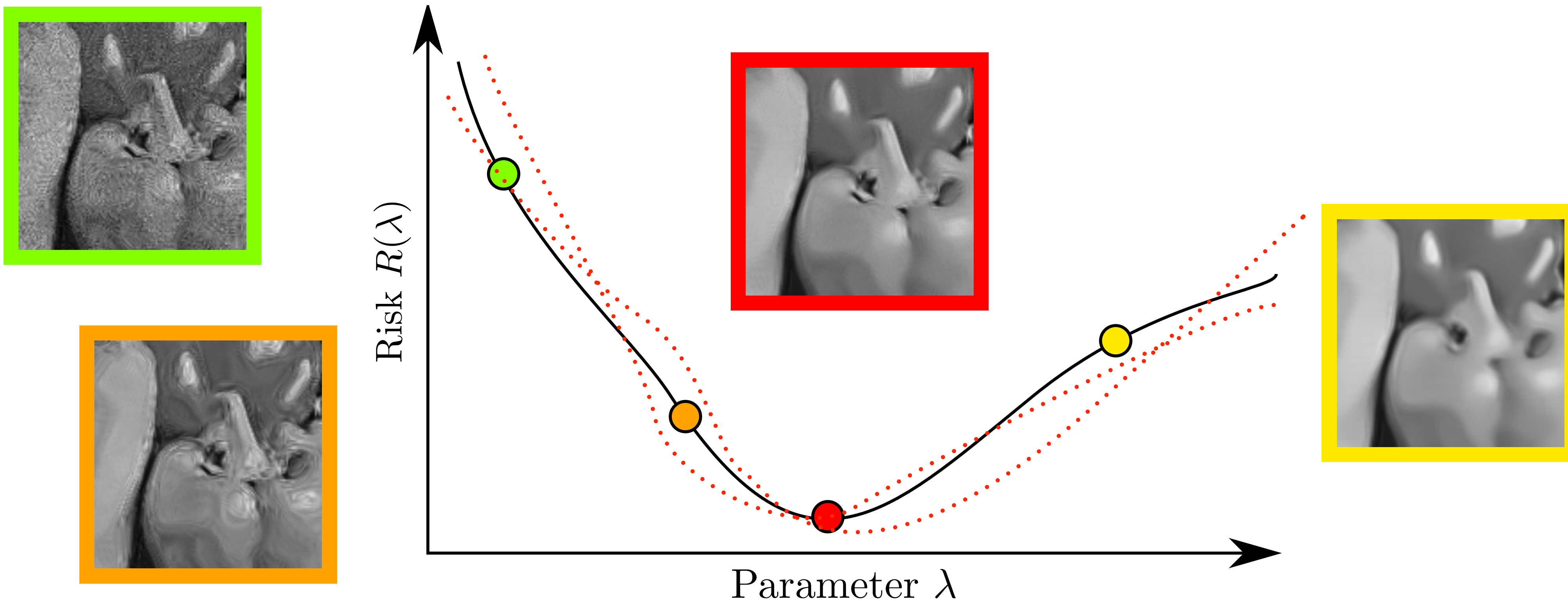


But: | $\mathbb{E}_w$ is not accessible $\rightarrow$ use one observation. ··········

Risk Minimization

Average risk: $R(\lambda) = \mathbb{E}_w(\|x_\lambda(y) - x_0\|^2)$

$\lambda^\star(y) = \underset{\lambda}{\operatorname{argmin}}\ R(\lambda)$ *Plugin-estimator:* $x_{\lambda^\star(y)}(y)$



But: | $\mathbb{E}_w$ is not accessible $\rightarrow$ use one observation. ....
x_0 is not accessible $\rightarrow$ needs risk estimators.

Prediction Risk Estimation

Prediction: $\mu_\lambda(y) = \Phi x_\lambda(y)$

Sensitivity analysis: if μ_λ is weakly differentiable

$$\mu_\lambda(y+\delta) = \mu_\lambda(y) + \partial\mu_\lambda(y) \cdot \delta + O(\|\delta\|^2)$$

Prediction Risk Estimation

Prediction: $\mu_\lambda(y) = \Phi x_\lambda(y)$

Sensitivity analysis: if μ_λ is weakly differentiable

$$\mu_\lambda(y+\delta) = \mu_\lambda(y) + \partial\mu_\lambda(y) \cdot \delta + O(\|\delta\|^2)$$

Stein Unbiased Risk Estimator:

$$\text{SURE}_\lambda(y) = \|y - \mu_\lambda(y)\|^2 - \sigma^2 P + 2\sigma^2 \text{df}_\lambda(y)$$

$$\text{df}_\lambda(y) = \text{tr}(\partial\mu_\lambda(y)) = \text{div}(\mu_\lambda)(y)$$

Prediction Risk Estimation

Prediction: $\mu_\lambda(y) = \Phi x_\lambda(y)$

Sensitivity analysis: if μ_λ is weakly differentiable

$$\mu_\lambda(y+\delta) = \mu_\lambda(y) + \partial\mu_\lambda(y) \cdot \delta + O(\|\delta\|^2)$$

Stein Unbiased Risk Estimator:

$$\mathrm{SURE}_\lambda(y) = \|y - \mu_\lambda(y)\|^2 - \sigma^2 P + 2\sigma^2 \mathrm{df}_\lambda(y)$$

$$\mathrm{df}_\lambda(y) = \mathrm{tr}(\partial\mu_\lambda(y)) = \mathrm{div}(\mu_\lambda)(y)$$

Theorem: [*Stein*, 1981] $\mathbb{E}_w(\mathrm{SURE}_\lambda(y)) = \mathbb{E}_w(\|\Phi x_0 - \mu_\lambda(y)\|^2)$

Prediction Risk Estimation

Prediction: $\mu_\lambda(y) = \Phi x_\lambda(y)$

Sensitivity analysis: if μ_λ is weakly differentiable

$$\mu_\lambda(y+\delta) = \mu_\lambda(y) + \partial\mu_\lambda(y)\cdot\delta + O(\|\delta\|^2)$$

Stein Unbiased Risk Estimator:

$$\mathrm{SURE}_\lambda(y) = \|y - \mu_\lambda(y)\|^2 - \sigma^2 P + 2\sigma^2 \mathrm{df}_\lambda(y)$$

$$\mathrm{df}_\lambda(y) = \mathrm{tr}(\partial\mu_\lambda(y)) = \mathrm{div}(\mu_\lambda)(y)$$

Theorem: [*Stein*, 1981] $\mathbb{E}_w(\mathrm{SURE}_\lambda(y)) = \mathbb{E}_w(\|\Phi x_0 - \mu_\lambda(y)\|^2)$

Other estimators: GCV, BIC, AIC, . . .

SURE:
- → Requires σ (not always available)
- → Unbiased and good practical performances

Generalized SURE

Problem: $\|\Phi x_0 - \Phi x_\lambda(y)\|$ poor indicator of $\|x_0 - x_\lambda(y)\|$.

Generalized SURE: take into account risk on $\ker(\Phi)^\perp$

Generalized SURE

Problem: $\|\Phi x_0 - \Phi x_\lambda(y)\|$ poor indicator of $\|x_0 - x_\lambda(y)\|$.

Generalized SURE: take into account risk on $\ker(\Phi)^\perp$

$$\mathrm{GSURE}_\lambda(y) = \|\hat{x}(y) - \Phi x_\lambda(y)\|^2 - \sigma^2 \mathrm{tr}((\Phi\Phi^*)^+) + 2\sigma^2 \mathrm{gdf}_\lambda(y)$$

Generalized df: $\mathrm{gdf}_\lambda(y) = \mathrm{tr}\left((\Phi\Phi^*)^+ \partial\mu_\lambda(y) \right)$

$$\mathrm{Proj}_{\ker(\Phi)^\perp} = \Pi = \Phi^*(\Phi\Phi^*)^+\Phi$$

ML estimator: $\hat{x}(y) = \Phi^*(\Phi\Phi^*)^+ y$.

Generalized SURE

Problem: $\|\Phi x_0 - \Phi x_\lambda(y)\|$ poor indicator of $\|x_0 - x_\lambda(y)\|$.

Generalized SURE: take into account risk on $\ker(\Phi)^\perp$

$$\text{GSURE}_\lambda(y) = \|\hat{x}(y) - \Phi x_\lambda(y)\|^2 - \sigma^2 \text{tr}((\Phi\Phi^*)^+) + 2\sigma^2 \text{gdf}_\lambda(y)$$

Generalized df: $\text{gdf}_\lambda(y) = \text{tr}\left((\Phi\Phi^*)^+ \partial\mu_\lambda(y)\right)$

$$\text{Proj}_{\ker(\Phi)^\perp} = \Pi = \Phi^*(\Phi\Phi^*)^+\Phi$$

ML estimator: $\hat{x}(y) = \Phi^*(\Phi\Phi^*)^+ y$.

Theorem: [Eldar 09, Pesquet al. 09, Vonesh et al. 08]

$$\mathbb{E}_w(\text{GSURE}_\lambda(y)) = \mathbb{E}_w(\|\Pi(x_0 - x_\lambda(y))\|^2)$$

Generalized SURE

Problem: $\|\Phi x_0 - \Phi x_\lambda(y)\|$ poor indicator of $\|x_0 - x_\lambda(y)\|$.

Generalized SURE: take into account risk on $\ker(\Phi)^\perp$

$$\text{GSURE}_\lambda(y) = \|\hat{x}(y) - \Phi x_\lambda(y)\|^2 - \sigma^2 \text{tr}((\Phi\Phi^*)^+) + 2\sigma^2 \text{gdf}_\lambda(y)$$

Generalized df: $\text{gdf}_\lambda(y) = \text{tr}\left((\Phi\Phi^*)^+ \partial\mu_\lambda(y)\right)$

$$\text{Proj}_{\ker(\Phi)^\perp} = \Pi = \Phi^*(\Phi\Phi^*)^+\Phi$$

ML estimator: $\hat{x}(y) = \Phi^*(\Phi\Phi^*)^+ y$.

Theorem: [Eldar 09, Pesquet al. 09, Vonesh et al. 08]

$$\mathbb{E}_w(\text{GSURE}_\lambda(y)) = \mathbb{E}_w(\|\Pi(x_0 - x_\lambda(y))\|^2)$$

$\longrightarrow$ How to compute $\partial\mu_\lambda(y)$?

Overview

- Synthesis vs. Analysis Regularization
- Risk Estimation
- **Local Behavior of Sparse Regularization**
- Robustness to Noise
- Numerical Illustrations

Variations and Stability

Observations: $\quad y = \Phi x_0 + w$

Recovery:

$$x_\lambda(y) \in \underset{x \in \mathbb{R}^N}{\operatorname{argmin}} \frac{1}{2}\|\Phi x - y\|^2 + \lambda \|D^* x\|_1 \qquad (\mathcal{P}_\lambda(y))$$

Variations and Stability

Observations: $y = \Phi x_0 + w$

Recovery:

$$x_\lambda(y) \in \underset{x \in \mathbb{R}^N}{\text{argmin}} \frac{1}{2}\|\Phi x - y\|^2 + \lambda \|D^* x\|_1 \qquad (\mathcal{P}_\lambda(y))$$

Remark: $x_\lambda(y)$ not always unique **but**

$\mu_\lambda(y) = \Phi x_\lambda(y)$ always unique.

Variations and Stability

Observations: $y = \Phi x_0 + w$

Recovery:

$$x_\lambda(y) \in \underset{x \in \mathbb{R}^N}{\text{argmin}} \frac{1}{2} \|\Phi x - y\|^2 + \lambda \|D^* x\|_1 \qquad (\mathcal{P}_\lambda(y))$$

Remark: $x_\lambda(y)$ not always unique **but**

$\mu_\lambda(y) = \Phi x_\lambda(y)$ always unique.

Questions:

– When is $y \to \mu_\lambda(y)$ differentiable ?

– Formula for $\partial \mu_\lambda(y)$.

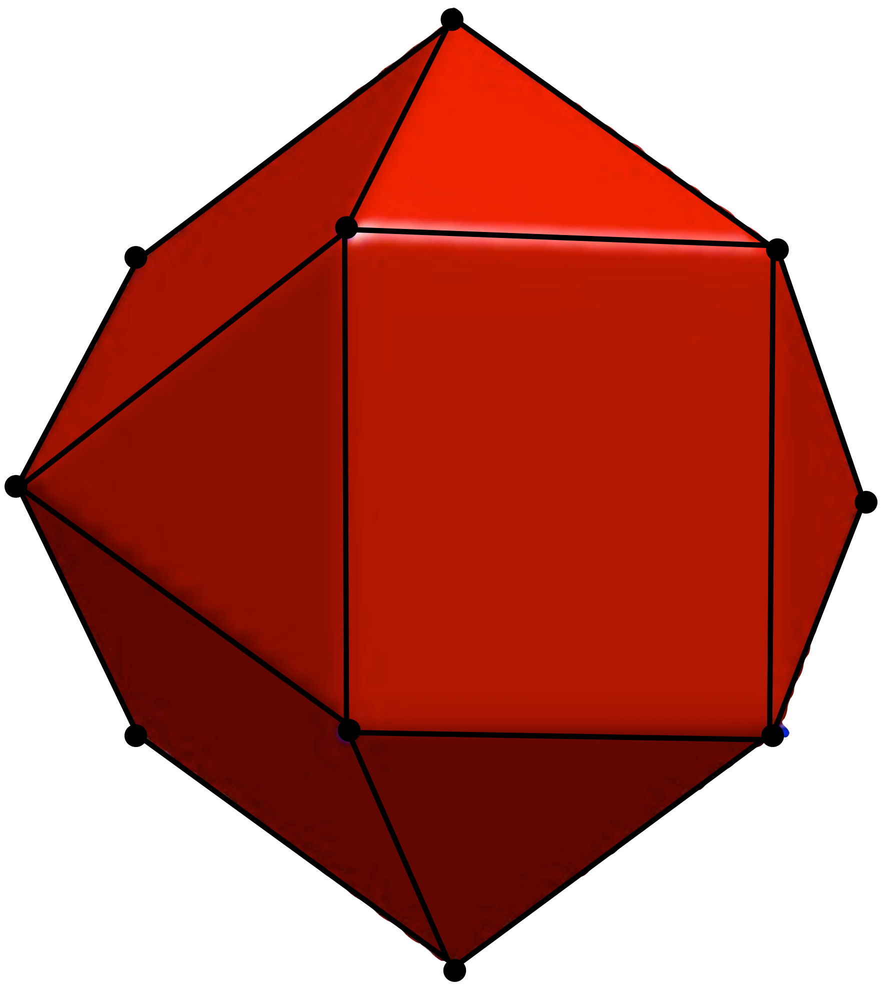
TV-1D Polytope

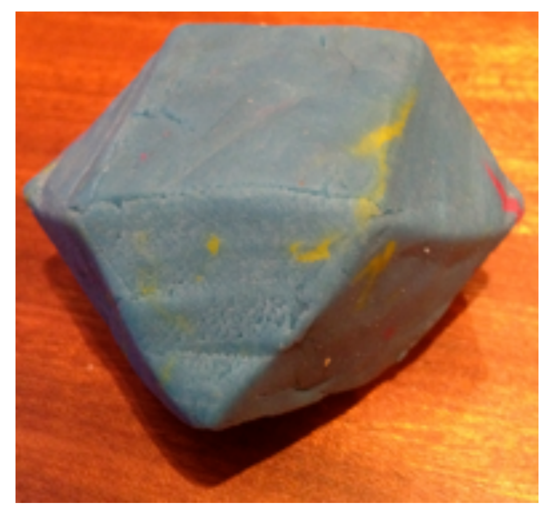
TV-1D ball: $\mathcal{B} = \{x \setminus \|D^* x\|_1 \leqslant 1\}$

Displayed in $\{x \setminus \langle x,\, 1\rangle = 0\} \sim \mathbb{R}^3$

$$D^* = \begin{pmatrix} 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \\ -1 & 0 & 0 & 1 \end{pmatrix}$$

$$x_\lambda(y) \in \underset{y=\Phi x}{\operatorname{argmin}} \|D^* x\|_1$$





Copyright © Mael

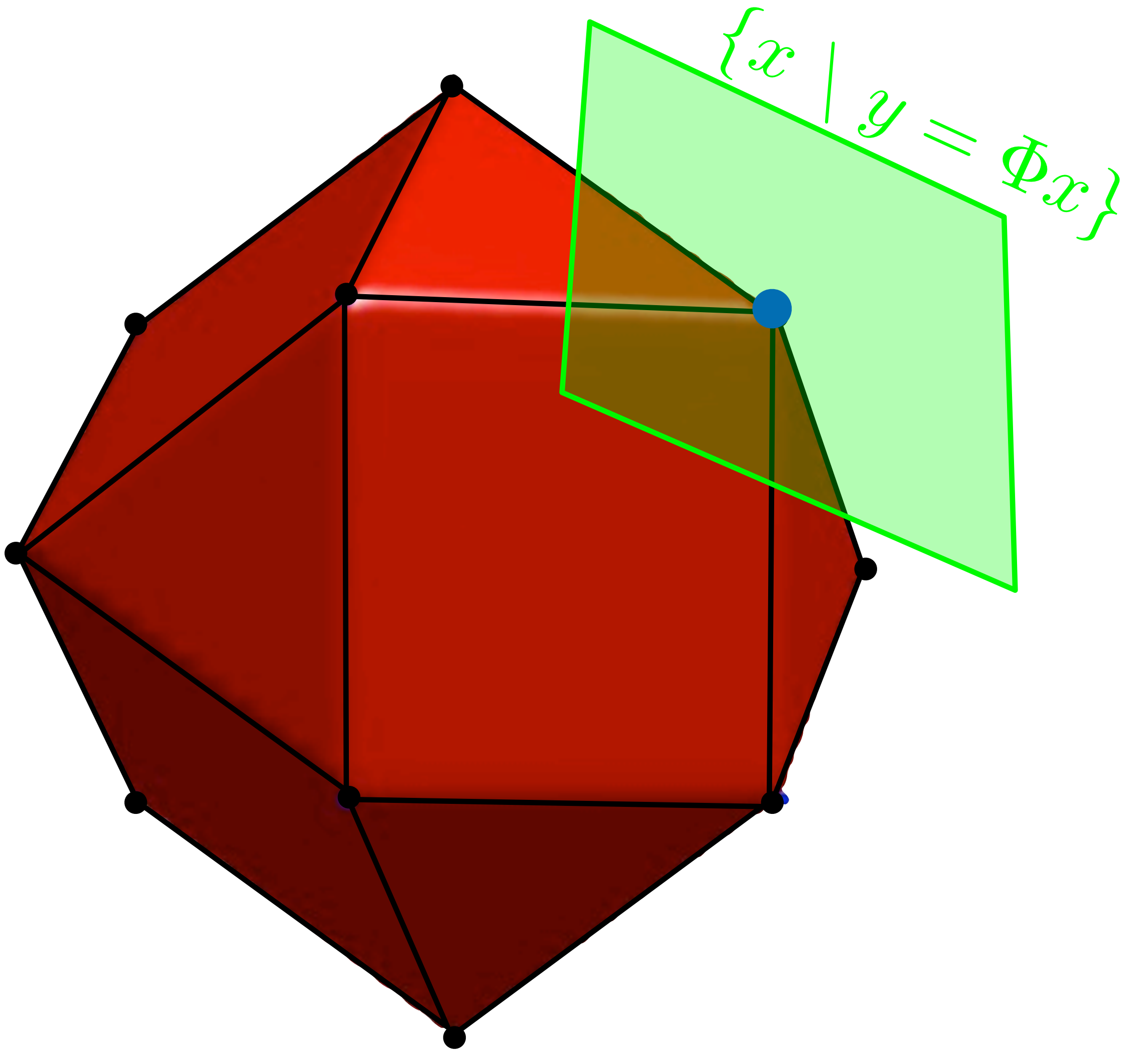
TV-1D Polytope

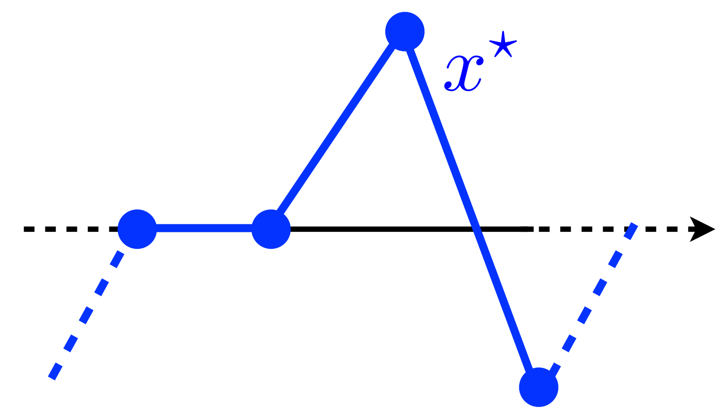
TV-1D ball: $\mathcal{B} = \{x \setminus \|D^* x\|_1 \leqslant 1\}$

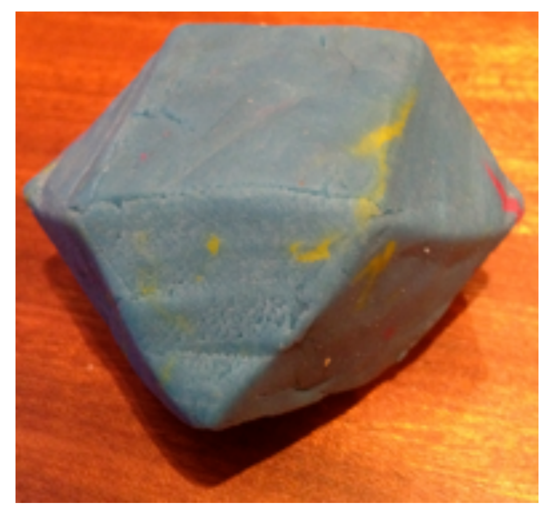
Displayed in $\{x \setminus \langle x,\, 1\rangle = 0\} \sim \mathbb{R}^3$

$$D^* = \begin{pmatrix} 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \\ -1 & 0 & 0 & 1 \end{pmatrix}$$

$$x_\lambda(y) \in \underset{y=\Phi x}{\operatorname{argmin}} \|D^* x\|_1$$







Copyright © Mael

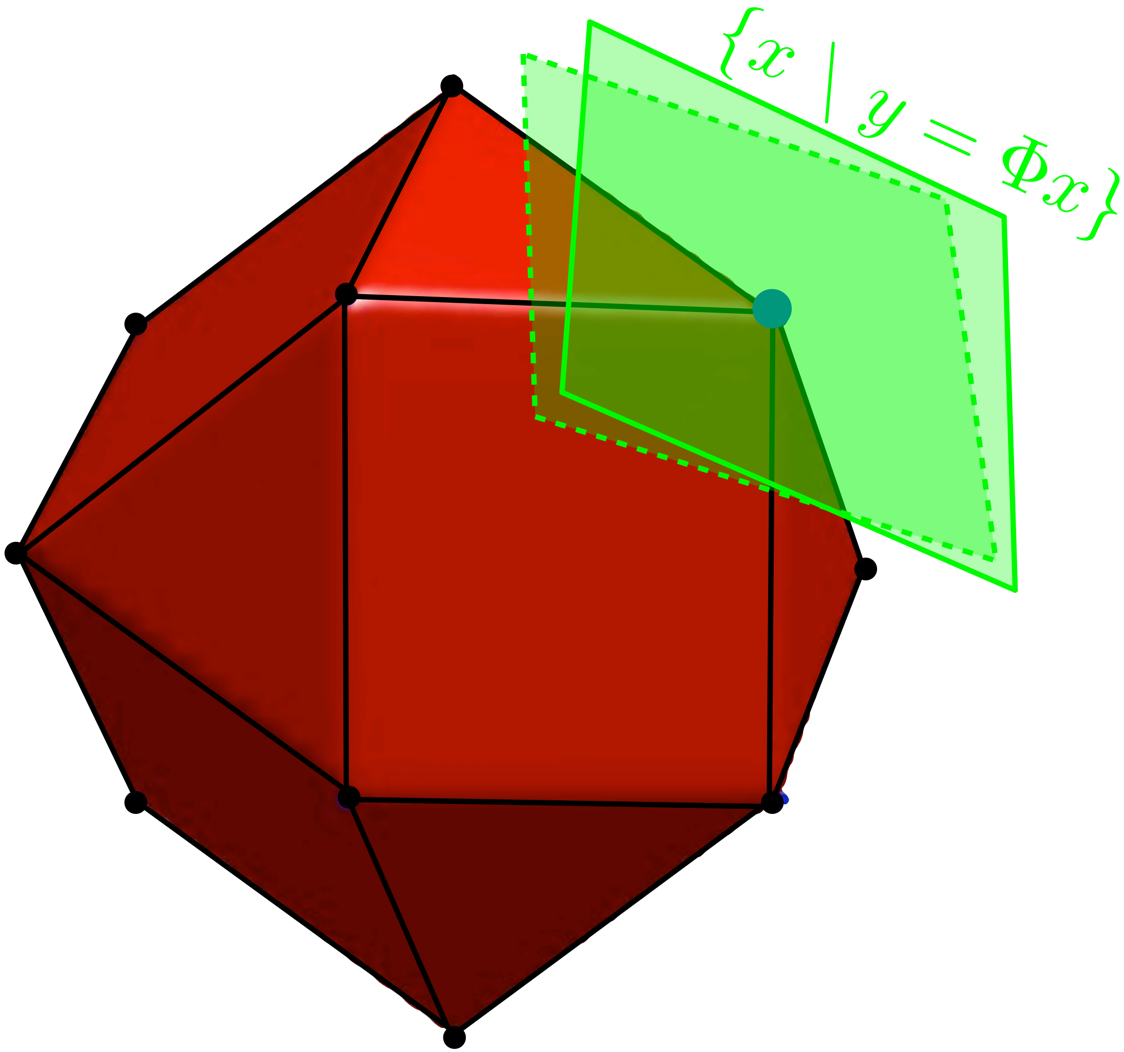
TV-1D Polytope

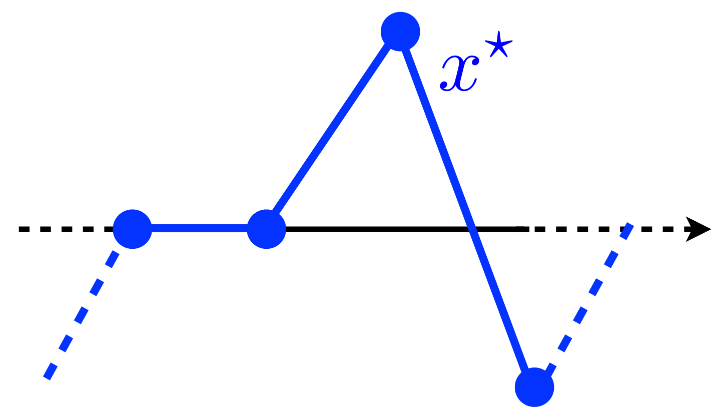
TV-1D ball: $\mathcal{B} = \{x \setminus \|D^* x\|_1 \leqslant 1\}$

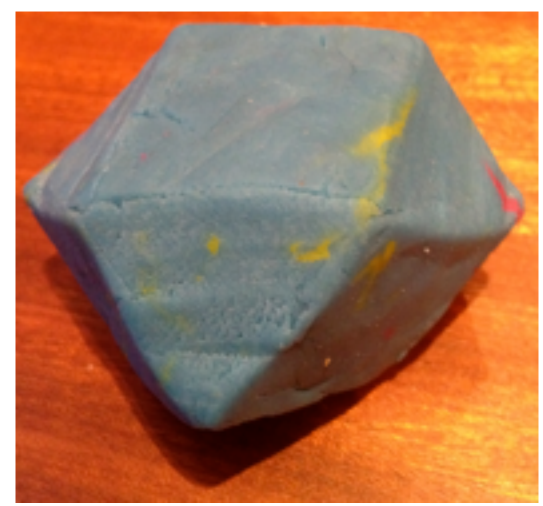
Displayed in $\{x \setminus \langle x,\, 1\rangle = 0\} \sim \mathbb{R}^3$

$$D^* = \begin{pmatrix} 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \\ -1 & 0 & 0 & 1 \end{pmatrix}$$

$$x_\lambda(y) \in \underset{y=\Phi x}{\text{argmin}} \|D^* x\|_1$$







Copyright © Mael

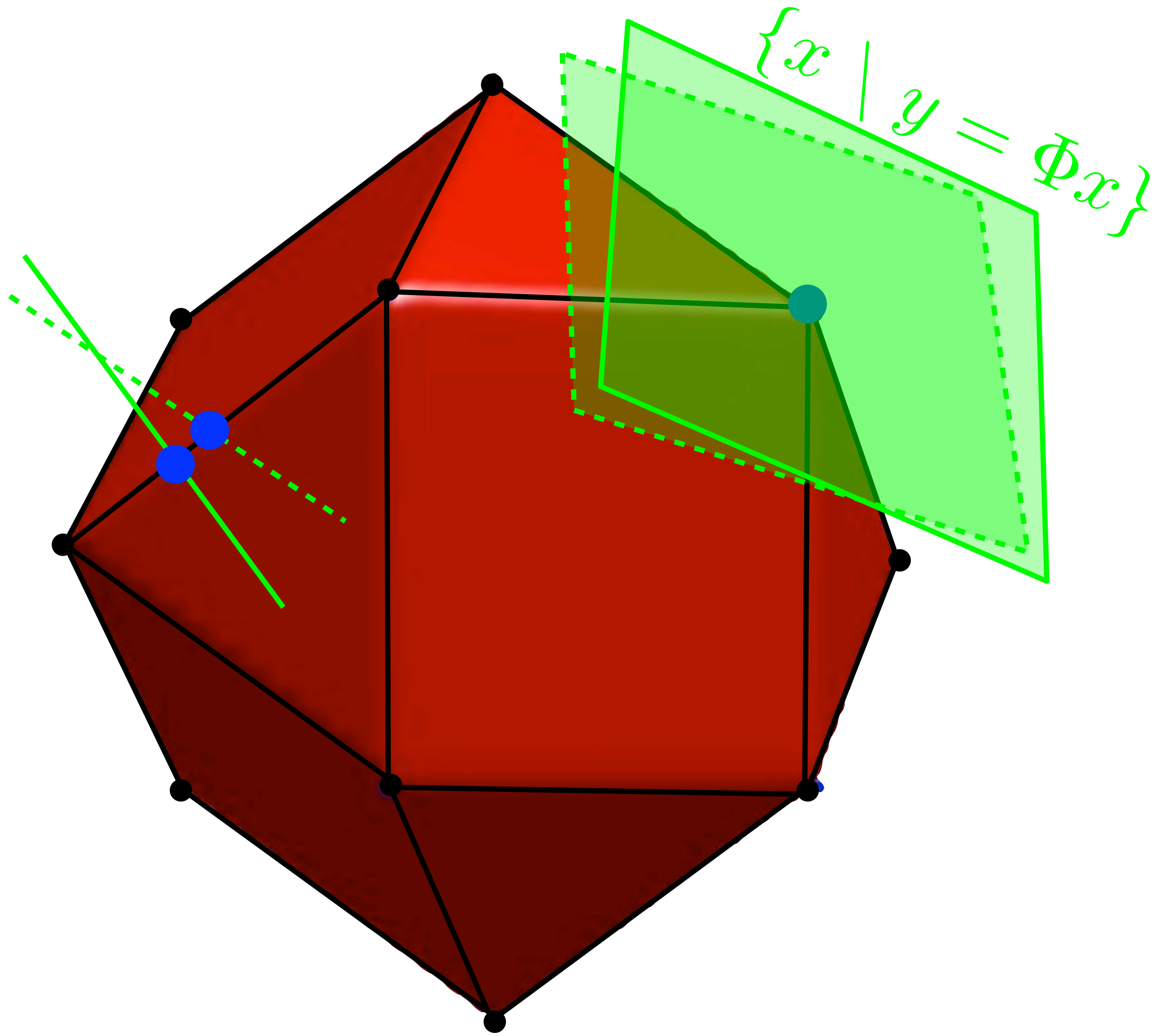
TV-1D Polytope

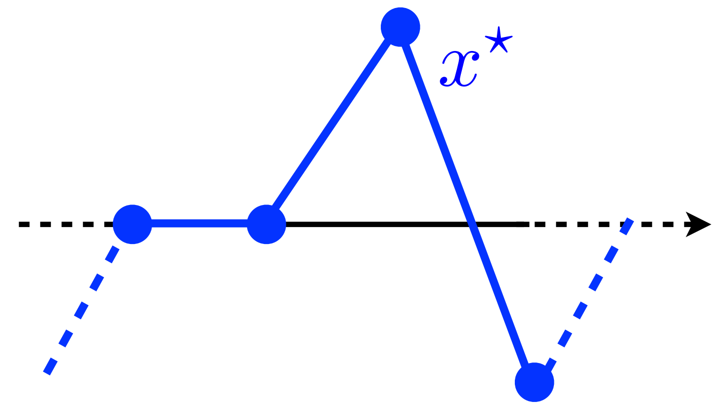
TV-1D ball: $\mathcal{B} = \{x \backslash \|D^* x\|_1 \leqslant 1\}$

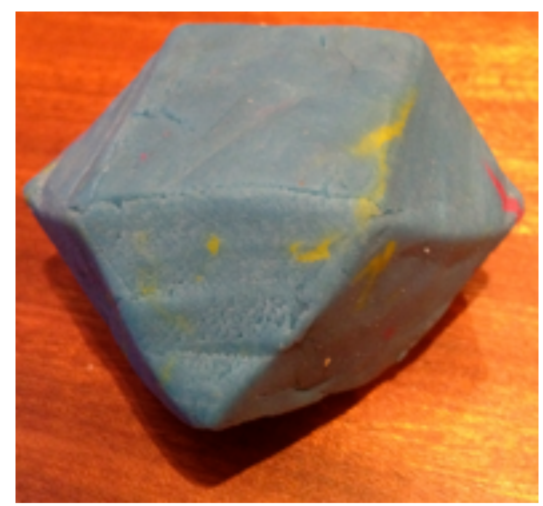
Displayed in $\{x \backslash \langle x,\, 1\rangle = 0\} \sim \mathbb{R}^3$

$$D^* = \begin{pmatrix} 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \\ -1 & 0 & 0 & 1 \end{pmatrix}$$

$$x_\lambda(y) \in \underset{y=\Phi x}{\operatorname{argmin}} \|D^* x\|_1$$







Copyright © Mael

Union of Subspaces Model

$$x_\lambda(y) \in \underset{x \in \mathbb{R}^N}{\operatorname{argmin}} \ \frac{1}{2}\|\Phi x - y\|^2 + \lambda \|D^* x\|_1 \qquad (\mathcal{P}_\lambda(y))$$

Support of the solution:

$$I = \{i \setminus (D^* x_\lambda(y))_i \neq 0\}$$

$$J = I^c$$

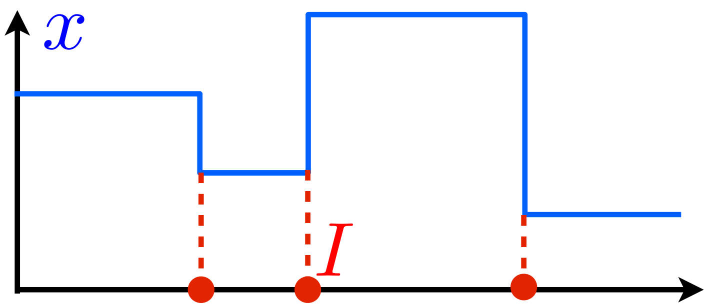
Union of Subspaces Model

$$x_\lambda(y) \in \underset{x \in \mathbb{R}^N}{\text{argmin}} \; \frac{1}{2}\|\Phi x - y\|^2 + \lambda \|D^* x\|_1 \qquad (\mathcal{P}_\lambda(y))$$

Support of the solution:

$$I = \{i \setminus (D^* x_\lambda(y))_i \neq 0\}$$

$$J = I^c$$



1-D total variation: $D^* x = (x_i - x_{i-1})_i$

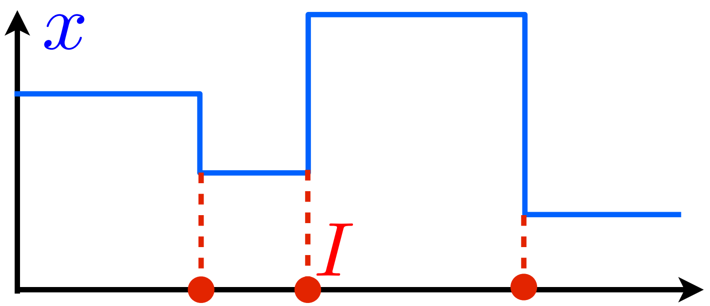
Union of Subspaces Model

$$x_\lambda(y) \in \underset{x \in \mathbb{R}^N}{\text{argmin}} \; \frac{1}{2}\|\Phi x - y\|^2 + \lambda \|D^* x\|_1 \qquad (\mathcal{P}_\lambda(y))$$

Support of the solution:

$$I = \{i \setminus (D^* x_\lambda(y))_i \neq 0\}$$

$$J = I^c$$



1-D total variation: $D^* x = (x_i - x_{i-1})_i$

Sub-space model: $\mathcal{G}_J = \ker(D_J^*) = \text{Im}(D_J)^\perp$

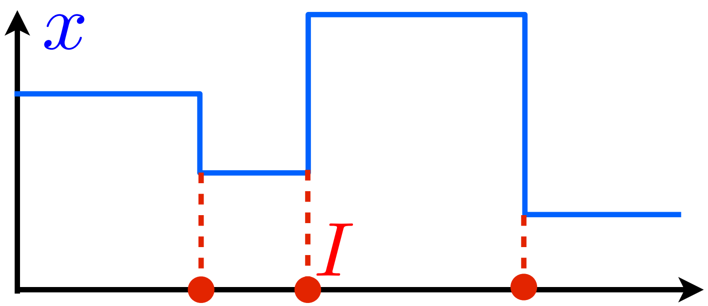
Union of Subspaces Model

$$x_\lambda(y) \in \underset{x \in \mathbb{R}^N}{\text{argmin}} \frac{1}{2}\|\Phi x - y\|^2 + \lambda \|D^* x\|_1 \qquad (\mathcal{P}_\lambda(y))$$

Support of the solution:

$$I = \{i \setminus (D^* x_\lambda(y))_i \neq 0\}$$

$$J = I^c$$



1-D total variation: $D^* x = (x_i - x_{i-1})_i$

Sub-space model: $\mathcal{G}_J = \ker(D_J^*) = \text{Im}(D_J)^\perp$

Local well-posedness: $\ker(\Phi) \cap \mathcal{G}_J = \{0\} \quad (H_J)$

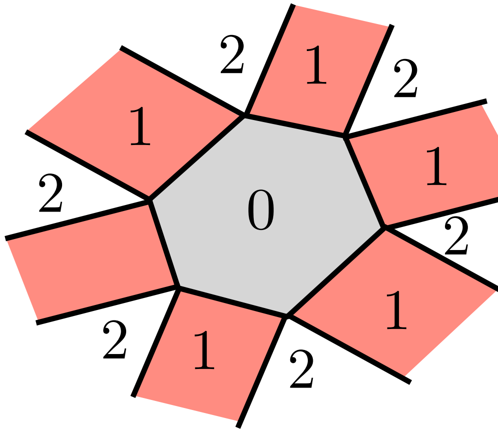
Lemma: There exists a solution $x^\star$ such that (H_J) holds.

Local Sign Stability

$$x_\lambda(y) \in \operatorname{argmin}_x \frac{1}{2}\|\Phi x - y\|^2 + \lambda \|D^* x\|_1 \qquad (\mathcal{P}_\lambda(y))$$

Lemma: $\operatorname{sign}(D^* x_\lambda(y))$ is constant around $(y, \lambda) \notin \mathcal{H}$.

To be understood: there exists a solution with same sign.



$\dim(\mathcal{G}_J)$

Local Sign Stability

$$x_\lambda(y) \in \operatorname{argmin}_x \frac{1}{2}\|\Phi x - y\|^2 + \lambda\|D^* x\|_1 \qquad (\mathcal{P}_\lambda(y))$$

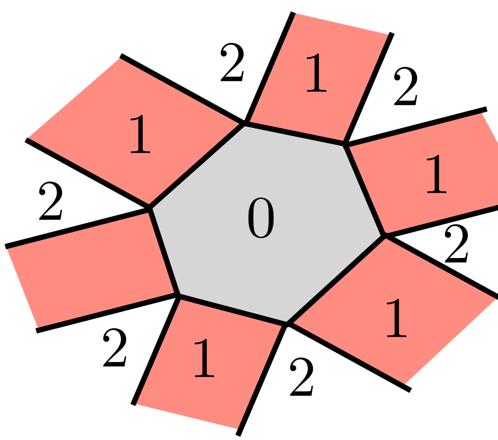
Lemma: $\text{sign}(D^* x_\lambda(y))$ is constant around $(y, \lambda) \notin \mathcal{H}$.

To be understood: there exists a solution with same sign.

Linearized problem:

$$= \text{sign}(D_I^* x_\lambda(y))$$

$$\hat{x}_{\bar\lambda}(\bar y) = \underset{x \in \mathcal{G}_J}{\operatorname{argmin}} \frac{1}{2}\|\Phi x - \bar y\|^2 + \bar\lambda \langle D_I^* x, s_I \rangle$$



Local Sign Stability

$$x_\lambda(y) \in \text{argmin}_x \frac{1}{2}\|\Phi x - y\|^2 + \lambda\|D^* x\|_1 \qquad (\mathcal{P}_\lambda(y))$$

Lemma: $\text{sign}(D^* x_\lambda(y))$ is constant around $(y, \lambda) \notin \mathcal{H}$.

To be understood: there exists a solution with same sign.

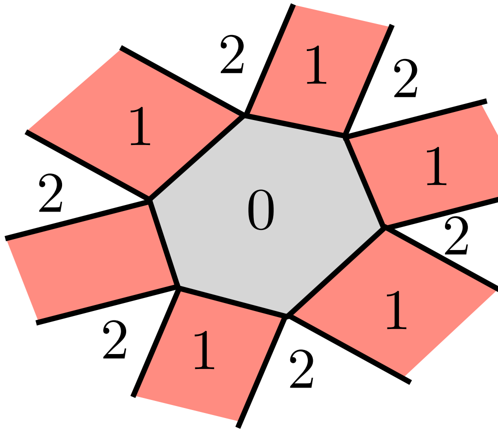
Linearized problem:

$$= \text{sign}(D_I^* x_\lambda(y))$$

$$\hat{x}_{\bar{\lambda}}(\bar{y}) = \underset{x \in \mathcal{G}_J}{\text{argmin}} \frac{1}{2}\|\Phi x - \bar{y}\|^2 + \bar{\lambda}\langle D_I^* x, s_I \rangle$$

$$= A^{[J]}\left(\Phi^* \bar{y} - \bar{\lambda} D_I s_I\right)$$

$$A^{[J]} z = \underset{x \in \mathcal{G}_J}{\text{argmin}} \frac{1}{2}\|\Phi x\|^2 - \langle x, z \rangle$$



$\dim(\mathcal{G}_J)$

Local Sign Stability

$$x_\lambda(y) \in \text{argmin}_x \frac{1}{2}\|\Phi x - y\|^2 + \lambda \|D^* x\|_1 \qquad (\mathcal{P}_\lambda(y))$$

Lemma: $\text{sign}(D^* x_\lambda(y))$ is constant around $(y, \lambda) \notin \mathcal{H}$.

To be understood: there exists a solution with same sign.

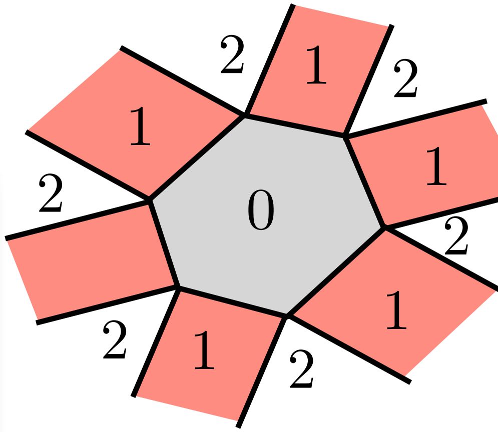
Linearized problem: $= \text{sign}(D_I^* x_\lambda(y))$

$$\hat{x}_{\bar\lambda}(\bar y) = \underset{x \in \mathcal{G}_J}{\text{argmin}} \frac{1}{2}\|\Phi x - \bar y\|^2 + \bar\lambda \langle D_I^* x, s_I \rangle$$

$$= A^{[J]} \left(\Phi^* \bar y - \bar\lambda D_I s_I \right)$$

$$A^{[J]} z = \underset{x \in \mathcal{G}_J}{\text{argmin}} \frac{1}{2}\|\Phi x\|^2 - \langle x, z \rangle$$

Theorem: If $(y, \lambda) \notin \mathcal{H}$, for $(\bar y, \bar\lambda)$ close to (y, λ), $\hat{x}_{\bar\lambda}(\bar y)$ is a solution of $\mathcal{P}_{\bar\lambda}(\bar y)$.

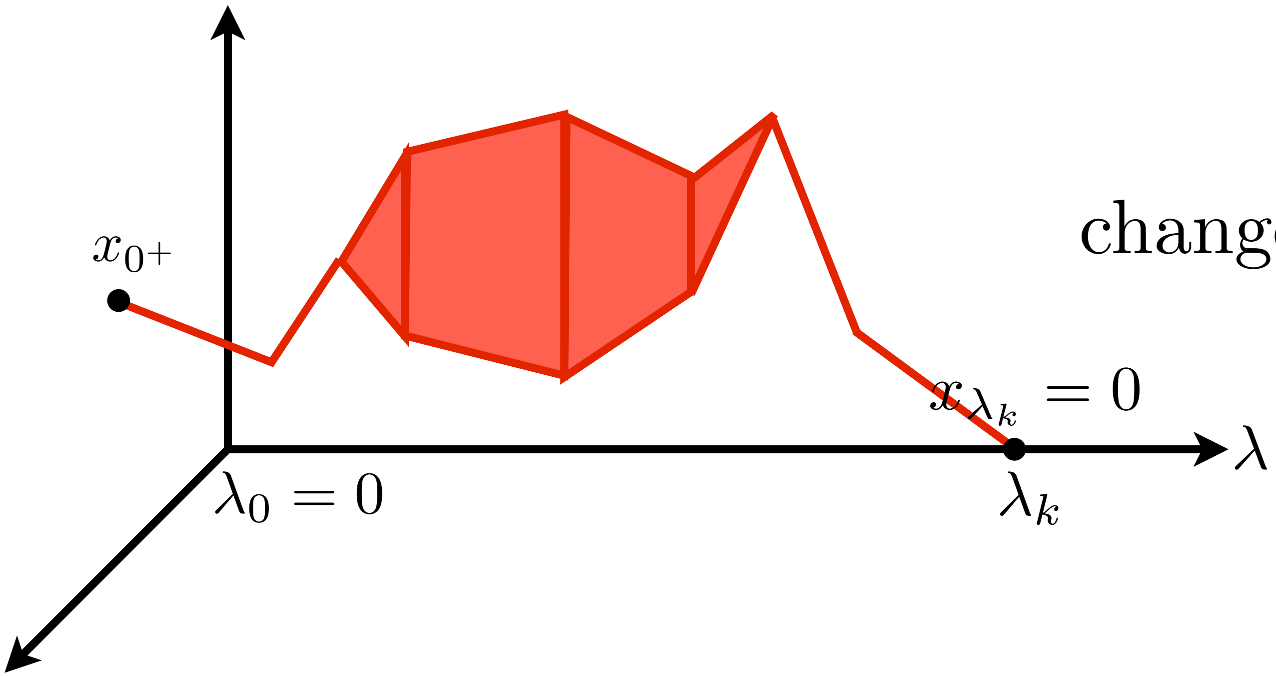


Local Affine Maps

Local parameterization: $\hat{x}_{\bar{\lambda}}(\bar{y}) = A^{[J]}\Phi^* \bar{y} - \bar{\lambda} A^{[J]} D_I \; s_I$

Under uniqueness assumption:

$y \mapsto x_\lambda(y)$
$\lambda \mapsto x_\lambda(y)$ | are piecewise affine functions.



breaking points

$\Longleftrightarrow$

change of support of $D^* x_\lambda(y)$

Application to GSURE

For $y \notin \mathcal{H}$, one has locally: $\mu_\lambda(y) = \Phi A^{[J]} \Phi^* y + \text{cst}.$

Corollary: Let $I = \text{supp}(D^* x_\lambda(y))$ such that H_J holds.

$$\text{df}_\lambda(y) = \text{div}\,(\mu_\lambda)\,(y) = \dim(\mathcal{G}_J)$$

$$\text{df}_\lambda(y) = \|x_\lambda(y)\|_0 \qquad \text{for } D = \text{Id (synthesis)}$$

$$\text{gdf}_\lambda(y) = \text{tr}(\Pi A^{[J]})$$

are unbiased estimators of df and gdf.

Application to GSURE

For $y \notin \mathcal{H}$, one has locally: $\mu_\lambda(y) = \Phi A^{[J]} \Phi^* y + \text{cst}.$

Corollary: Let $I = \text{supp}(D^* x_\lambda(y))$ such that H_J holds.

$$\text{df}_\lambda(y) = \text{div}\,(\mu_\lambda)\,(y) = \dim(\mathcal{G}_J)$$

$$\text{df}_\lambda(y) = \|x_\lambda(y)\|_0 \qquad \text{for } D = \text{Id (synthesis)}$$

$$\text{gdf}_\lambda(y) = \text{tr}(\Pi A^{[J]})$$

are unbiased estimators of df and gdf.

Trick: $\text{tr}(A) = \mathbb{E}_z(\langle Az,\, z\rangle),\ z \sim \mathcal{N}(0, \text{Id}_P).$

Proposition: $\text{gdf}_\lambda(y) = \mathbb{E}_z(\langle \nu(z),\, \Phi^+ z\rangle), \quad z \sim \mathcal{N}(0, \text{Id}_P)$

where $\nu(z)$ solves $\begin{pmatrix} \Phi^*\Phi & D_J \\ D_J^* & 0 \end{pmatrix} \begin{pmatrix} \nu(z) \\ \tilde{\nu} \end{pmatrix} = \begin{pmatrix} \Phi^* z \\ 0 \end{pmatrix}$

In practice: $\text{gdf}_\lambda(y) \approx \frac{1}{K} \sum_{k=1}^{K} \langle \nu(z_k),\, \Phi^+ z_k \rangle,\ z_k \sim \mathcal{N}(0, \text{Id}_P).$

CS with Analysis TI Wavelets

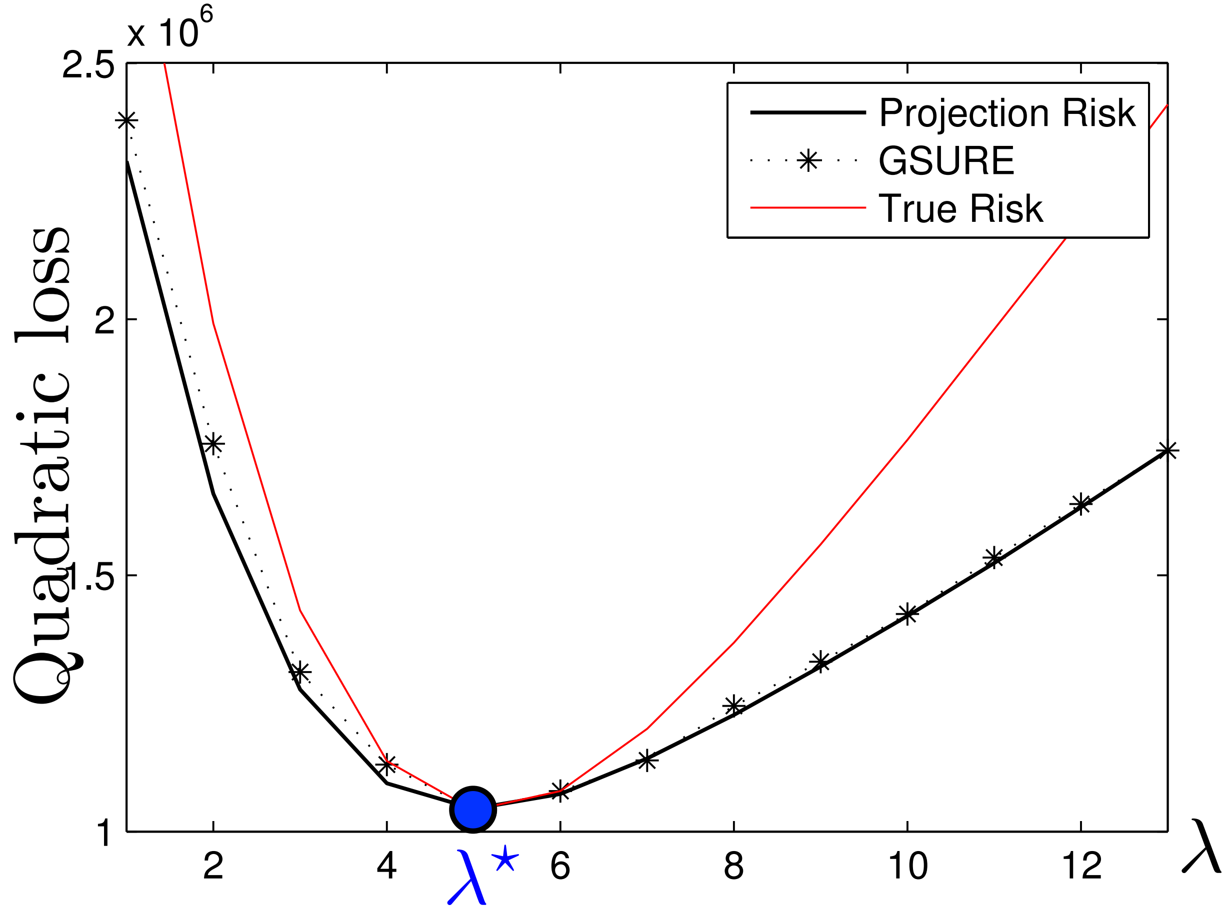
$\Phi^+ y$

$x_{\lambda^\star}(y)$

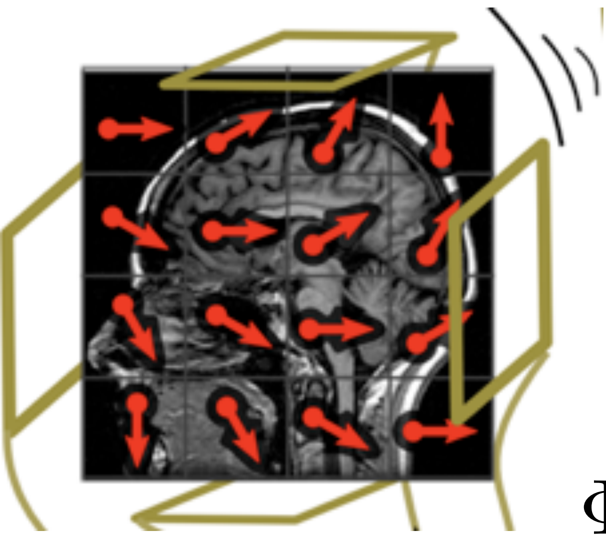
$\Phi \in \mathbb{R}^{P \times N}$ realization of a random vector.

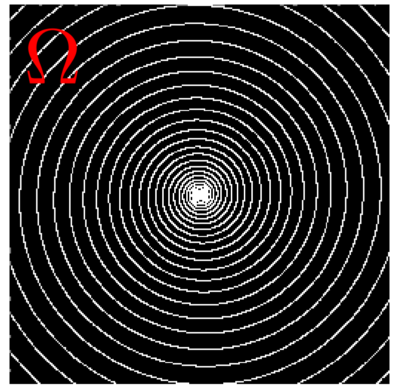
$$P = N/4$$

D: wavelet TI redundant tight frame.

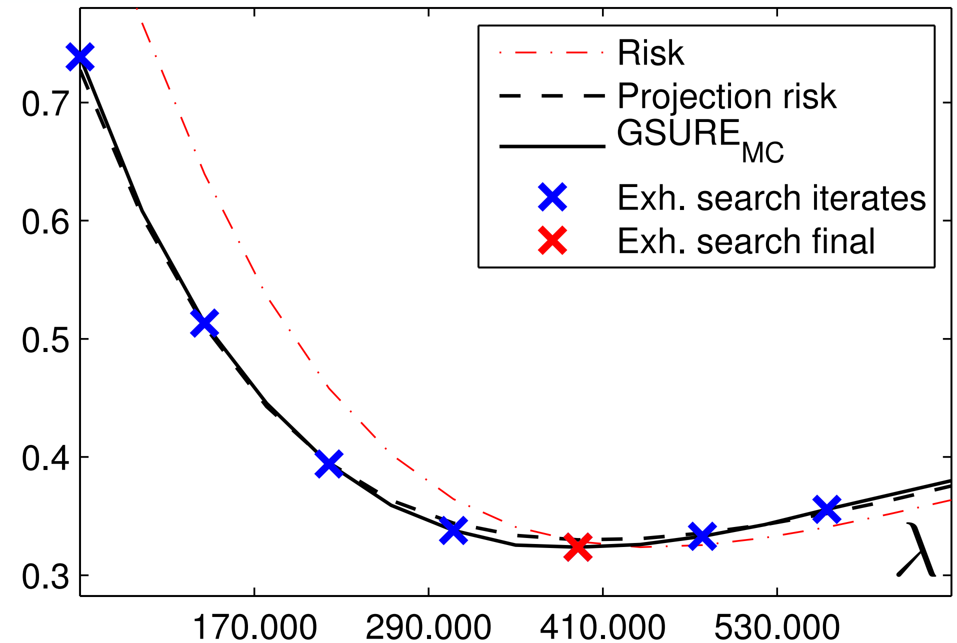


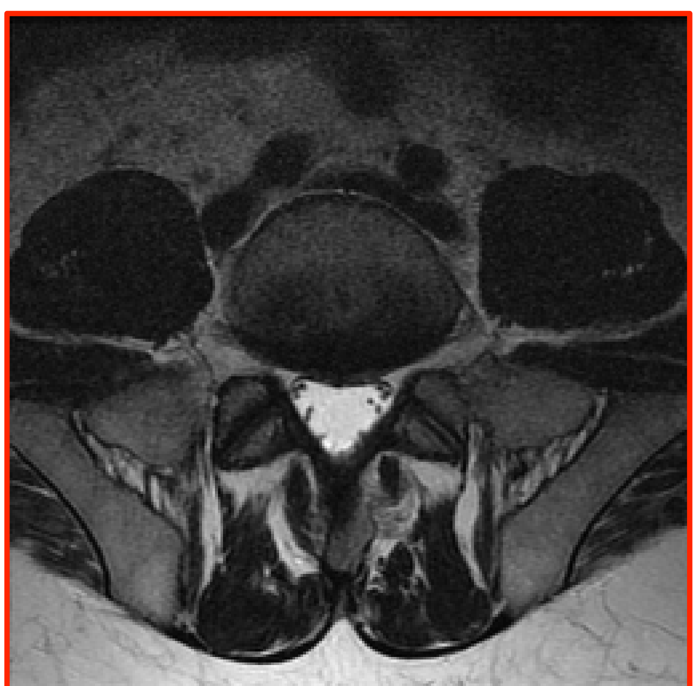
MRI with Anisotropic TV





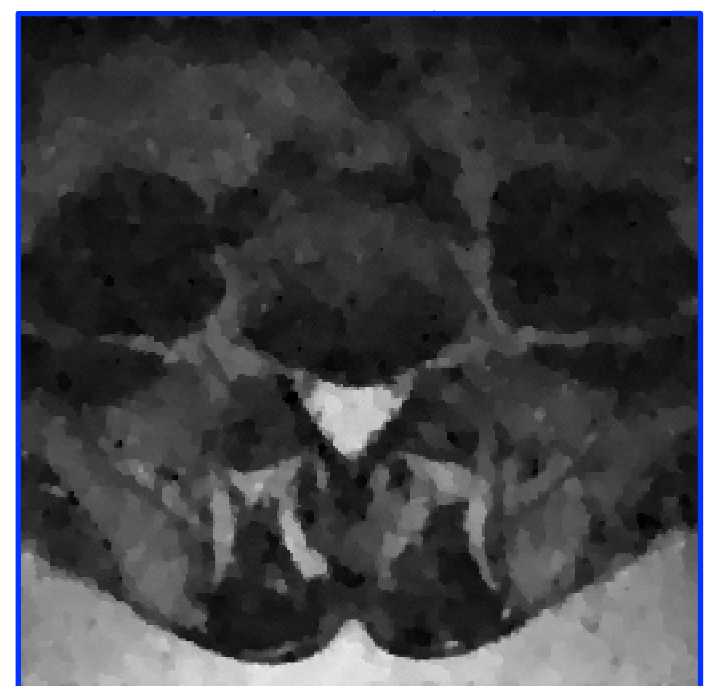
$$\Phi f = (\hat{f}(\omega))_{\omega \in \Omega}$$





x_0

$\Phi^+ y$



$x_{\lambda^\star}(y)$

Overview

- Synthesis vs. Analysis Regularization
- Risk Estimation
- Local Behavior of Sparse Regularization
- **Robustness to Noise**
- Numerical Illustrations

Identifiability Criterion

Identifiability criterion of a sign: we suppose (H_J) holds

$$\mathrm{IC}(s) = \min_{u \in \mathrm{Ker}\, D_J} \|\Omega s_I - u\|_\infty \qquad \text{(convex} \rightarrow \text{computable)}$$

where $\Omega = D_J^+(\mathrm{Id} - \Phi^* \Phi A^{[J]}) D_I$

Identifiability Criterion

Identifiability criterion of a sign: we suppose (H_J) holds

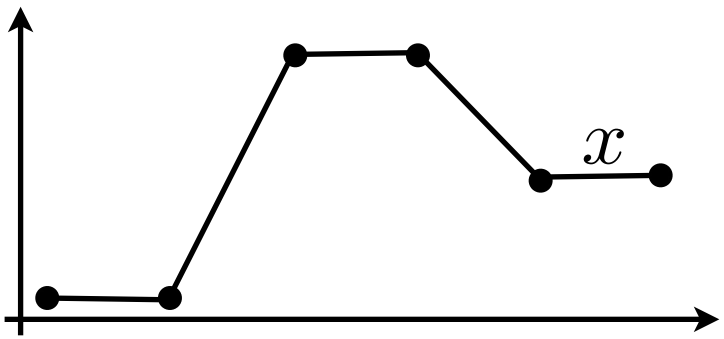
$$\mathrm{IC}(s) = \min_{u \in \mathrm{Ker}\, D_J} \|\Omega s_I - u\|_\infty \qquad \text{(convex} \rightarrow \text{computable)}$$

where $\Omega = D_J^+(\mathrm{Id} - \Phi^*\Phi A^{[J]})D_I$

Discrete 1-D derivative:

$$D^* x = (x_i - x_{i-1})_i$$

$$\Phi = \mathrm{Id} \ \text{(denoising)}$$



Identifiability Criterion

Identifiability criterion of a sign: we suppose (H_J) holds

$$\mathrm{IC}(s) = \min_{u \in \mathrm{Ker}\, D_J} \|\Omega s_I - u\|_\infty$$

(convex → computable)

where $\Omega = D_J^+(\mathrm{Id} - \Phi^*\Phi A^{[J]}) D_I$

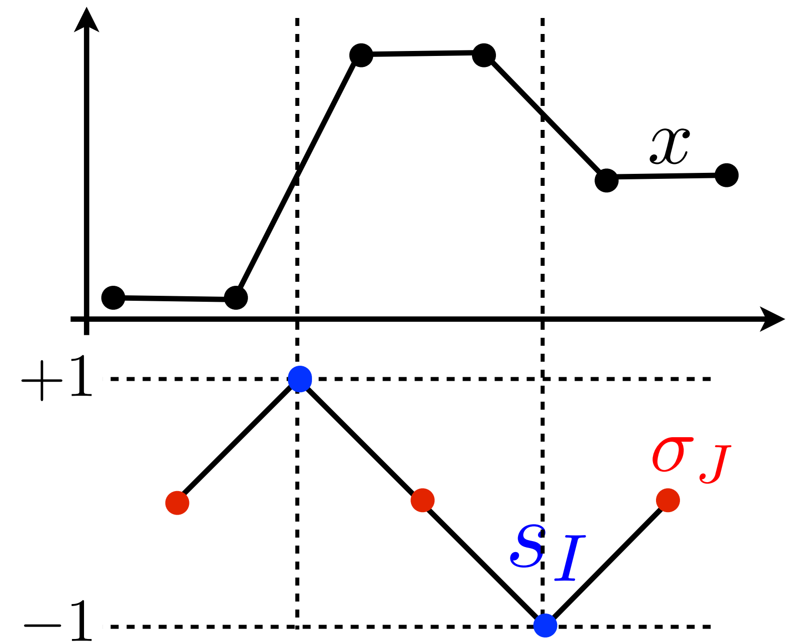
Discrete 1-D derivative:

$$D^* x = (x_i - x_{i-1})_i$$

$$\Phi = \mathrm{Id} \text{ (denoising)}$$

$$\mathrm{IC}(s) = \|\sigma_J\|_\infty$$

$$\begin{cases} s_I = \mathrm{sign}(D_I^* x) \\ \sigma_J = \Omega s_I \end{cases}$$



$\mathrm{IC}(\mathrm{sign}(D^* x)) < 1$

Robustness to Small Noise

$$\mathrm{IC}(s) = \min_{u \in \mathrm{Ker}\, D_J} \|\Omega s_I - u\|_\infty \qquad \Omega = D_J^+(\mathrm{Id} - \Phi^* \Phi A^{[J]}) D_I$$

Robustness to Small Noise

$$\text{IC}(s) = \min_{u \in \text{Ker } D_J} \|\Omega s_I - u\|_\infty \qquad \Omega = D_J^+(\text{Id} - \Phi^* \Phi A^{[J]}) D_I$$

Theorem: If $\text{IC}(\text{sign}(D^* x_0)) < 1$ $\qquad T = \min_{i \in I} |(D^* x_0)_i|$

If $\|w\|/T$ is small enough and $\lambda \sim \|w\|$, then

$$x^\star = x_0 + A^{[J]} \Phi^* w - \lambda A^{[J]} D_I s_I,$$

is the unique solution of $\mathcal{P}_\lambda(y)$.

Robustness to Small Noise

$$\mathrm{IC}(s) = \min_{u \in \mathrm{Ker}\, D_J} \|\Omega s_I - u\|_\infty \qquad \Omega = D_J^+(\mathrm{Id} - \Phi^* \Phi A^{[J]}) D_I$$

Theorem: If $\mathrm{IC}(\mathrm{sign}(D^* x_0)) < 1$ $\qquad T = \min_{i \in I} |(D^* x_0)_i|$

If $\|w\|/T$ is small enough and $\lambda \sim \|w\|$, then

$$x^\star = x_0 + A^{[J]} \Phi^* w - \lambda A^{[J]} D_I s_I,$$

is the unique solution of $\mathcal{P}_\lambda(y)$.

Linear convergence rate: $\|x^\star - x_0\| = O(\|w\|)$

Robustness to Small Noise

$$\text{IC}(s) = \min_{u \in \text{Ker } D_J} \|\Omega s_I - u\|_\infty \qquad \Omega = D_J^+(\text{Id} - \Phi^* \Phi A^{[J]}) D_I$$

Theorem: If $\text{IC}(\text{sign}(D^* x_0)) < 1$ $\qquad T = \min_{i \in I} |(D^* x_0)_i|$

If $\|w\|/T$ is small enough and $\lambda \sim \|w\|$, then

$$x^\star = x_0 + A^{[J]} \Phi^* w - \lambda A^{[J]} D_I s_I,$$

is the unique solution of $\mathcal{P}_\lambda(y)$.

Linear convergence rate: $\|x^\star - x_0\| = O(\|w\|)$

Theorem: If $\text{IC}(\text{sign}(D^* x_0)) < 1$

x_0 is the unique solution of $\mathcal{P}_0(\Phi x_0)$.

Robustness to Small Noise

$$\mathrm{IC}(s) = \min_{u \in \mathrm{Ker}\, D_J} \|\Omega s_I - u\|_\infty \qquad \Omega = D_J^+(\mathrm{Id} - \Phi^* \Phi A^{[J]}) D_I$$

Theorem: If $\mathrm{IC}(\mathrm{sign}(D^* x_0)) < 1$ $\qquad T = \min_{i \in I} |(D^* x_0)_i|$

If $\|w\|/T$ is small enough and $\lambda \sim \|w\|$, then

$$x^\star = x_0 + A^{[J]} \Phi^* w - \lambda A^{[J]} D_I s_I,$$

is the unique solution of $\mathcal{P}_\lambda(y)$.

Linear convergence rate: $\|x^\star - x_0\| = O(\|w\|)$

Theorem: If $\mathrm{IC}(\mathrm{sign}(D^* x_0)) < 1$

x_0 is the unique solution of $\mathcal{P}_0(\Phi x_0)$.

$\longrightarrow$ When $D = \mathrm{Id}$, results of J.J. Fuchs.

IC is Sharp for Sign Stability

Theorem: Suppose $\text{IC}(\text{sign}(D^* x_0)) > 1$,

if $$\lambda > \frac{\|\Pi^{[J]} w\|_\infty}{\text{IC}(\text{sign}(D^* x_0)) - 1}$$

where $$\Pi^{[J]} = D_J^+ \Phi^* (\Phi A^{[J]} \Phi^* - \text{Id})$$

then for any solution $x^\star$ of $\mathcal{P}_\lambda(y)$

$$\text{sign}(D^* x_0) \neq \text{sign}(D^* x^\star)$$

Corrolary: Suppose $\text{IC}(\text{sign}(D^* x_0)) > 1$,

then for all $\lambda \geqslant 0$ and for any solution $x^\star$ of $\mathcal{P}_\lambda(\Phi x_0)$,

$$\text{sign}(D^* x_0) \neq \text{sign}(D^* x^\star)$$

IC is Sharp for Sign Stability

Theorem: Suppose $\text{IC}(\text{sign}(D^* x_0)) > 1$,

if $$\lambda > \frac{\|\Pi^{[J]} w\|_\infty}{\text{IC}(\text{sign}(D^* x_0)) - 1}$$

where $$\Pi^{[J]} = D_J^+ \Phi^* (\Phi A^{[J]} \Phi^* - \text{Id})$$

then for any solution $x^\star$ of $\mathcal{P}_\lambda(y)$

$$\text{sign}(D^* x_0) \neq \text{sign}(D^* x^\star)$$

Corrolary: Suppose $\text{IC}(\text{sign}(D^* x_0)) > 1$,

then for all $\lambda \geqslant 0$ and for any solution $x^\star$ of $\mathcal{P}_\lambda(\Phi x_0)$,

$$\text{sign}(D^* x_0) \neq \text{sign}(D^* x^\star)$$

If $\text{IC}(\text{sign}(D^* x_0)) = 1$: both stability / no-stability depending on the value of w.

Robustness to Bounded Noise

$$\textit{Robustness criterion: } \mathrm{RC}(I) = \max_{\|p_I\|_\infty \leqslant 1} \min_{u \in \ker(D_J)} \|\Omega p_I - u\|_\infty$$

$$= \mathrm{IC}(p)$$

Robustness to Bounded Noise

Robustness criterion: $\mathrm{RC}(I) = \max_{\|p_I\|_\infty \leqslant 1} \min_{u \in \ker(D_J)} \|\Omega p_I - u\|_\infty$

$= \mathrm{IC}(p)$

Theorem: If $\mathrm{RC}(I) < 1$ for $I = \mathrm{Supp}(D^* x_0)$, setting

$$\lambda = \rho \|w\|_2 \frac{c_J}{1 - \mathrm{RC}(I)} \quad \text{with} \quad \rho > 1$$

$\mathcal{P}_\lambda(y)$ has a unique solution $x^\star \in \mathcal{G}_J$ and

$$\|x_0 - x^\star\|_2 \leqslant C_J \|w\|_2$$

Constants: $c_J = \|D_J^+ \Phi^* (\Phi A^{[J]} \Phi^* - \mathrm{Id})\|_{2,\infty}$

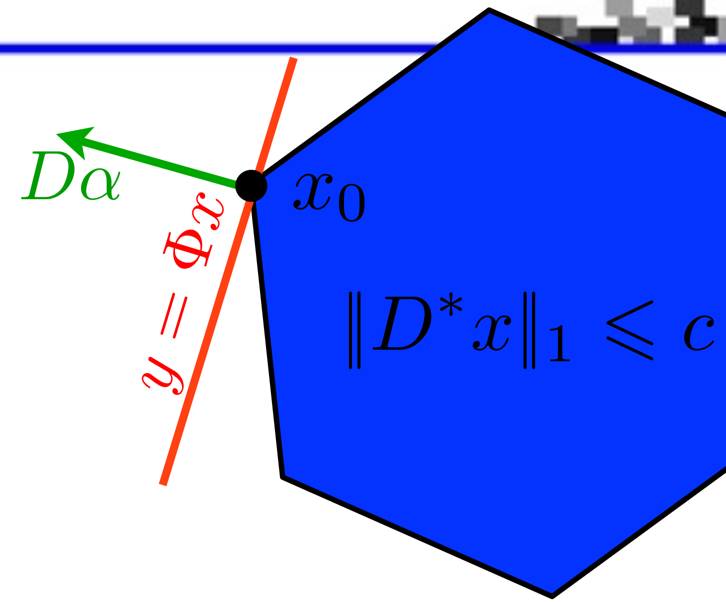
$$C_J = \|A^{[J]}\|_{2,2} \left(\|\Phi\|_{2,2} + \frac{\rho c_J}{1 - \mathrm{RC}(I)} \|D_I\|_{2,\infty} \right)$$

$\longrightarrow$ When $D = \mathrm{Id}$, results of Tropp (ERC)

Source Condition

Noiseless CNS: $x_0 \in \underset{\Phi x = \Phi x_0}{\text{argmin}} \|D^* x\|_1$

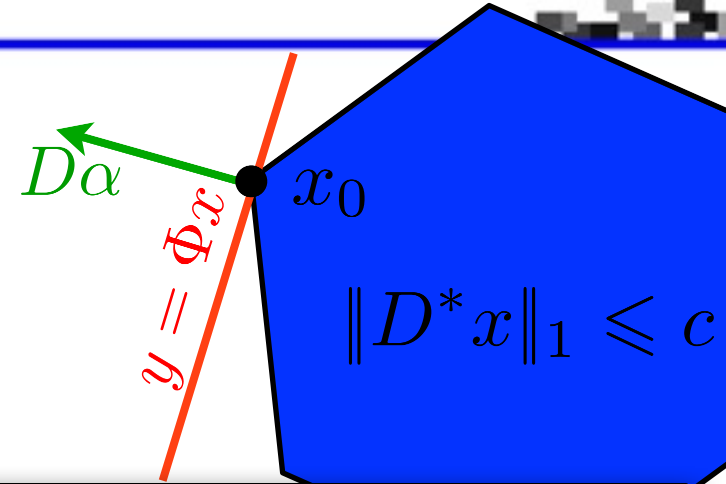
$$(\text{SC}_{x_0}) \iff \exists \alpha \in \partial \|D^* x_0\|_1, \; D\alpha \in \ker(\Phi)^\perp$$



Source Condition

Noiseless CNS: $x_0 \in \underset{\Phi x = \Phi x_0}{\text{argmin}} \, \|D^* x\|_1$

$(\text{SC}_{x_0}) \iff \exists \alpha \in \partial \|D^* x_0\|_1, \; D\alpha \in \ker(\Phi)^\perp$



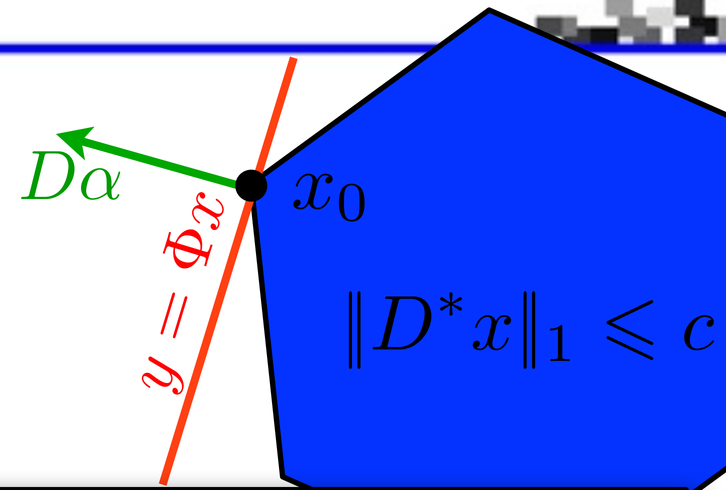
Theorem: If (SC_{x_0}), (H_J) and $\|\alpha_J\|_\infty < 1$, then

$$\|D^*(x^\star - x_0)\| = O\left(\frac{\|w\|}{1 - \|\alpha_J\|_\infty}\right)$$

[Grassmair, Inverse Prob., 2011]

Source Condition

Noiseless CNS: $x_0 \in \underset{\Phi x = \Phi x_0}{\operatorname{argmin}} \|D^* x\|_1$



$(\text{SC}_{x_0}) \iff \exists \alpha \in \partial \|D^* x_0\|_1,\ D\alpha \in \ker(\Phi)^\perp$

Theorem: If (SC_{x_0}), (H_J) and $\|\alpha_J\|_\infty < 1$, then

$$\|D^*(x^\star - x_0)\| = O\left(\frac{\|w\|}{1 - \|\alpha_J\|_\infty}\right)$$

[Grassmair, Inverse Prob., 2011]

Proposition: Let $s = \text{sign}(D^* x_0)$ and $\begin{cases} \alpha_J = \Omega s_I - u \\ \alpha_I = s_I \end{cases}$

Then $\text{IC}(s) < 1 \implies (\text{SC}_{x_0})$ and $\|\alpha_J\|_\infty = \text{IC}(s)$.

$$\text{IC}(s) = \min \|\Omega s_I - u\|_\infty \quad \text{subj.to} \quad u \in \text{Ker}\, D_J$$

Overview

- Synthesis vs. Analysis Regularization
- Risk Estimation
- Local Behavior of Sparse Regularization
- Robustness to Noise
- **Numerical Illustrations**

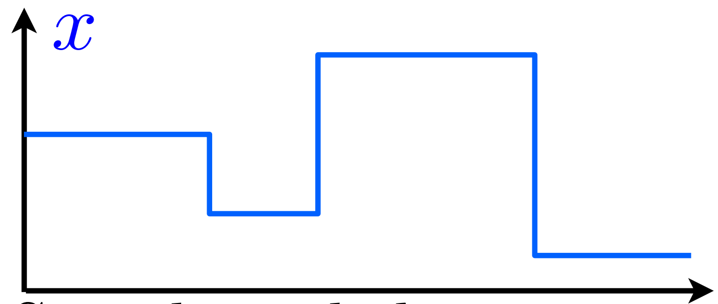
Example: TV Denoising in 1-D

Discrete 1-D derivative:

$$D^* x = (x_i - x_{i-1})_i$$

Denoising $\Phi = \mathrm{Id}$.

$\{\mathcal{G}_J \setminus \dim \mathcal{G}_J = k\}$



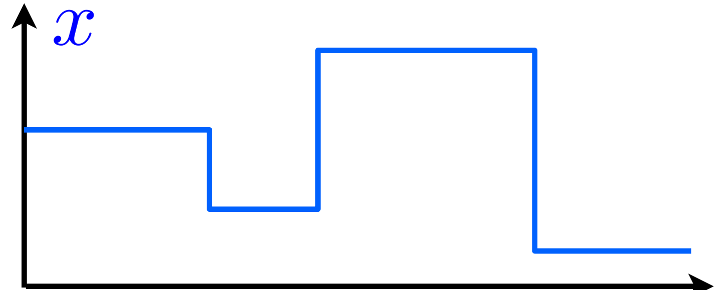
Signals with $k-1$ steps

Example: TV Denoising in 1-D

Discrete 1-D derivative:

$$D^* x = (x_i - x_{i-1})_i$$

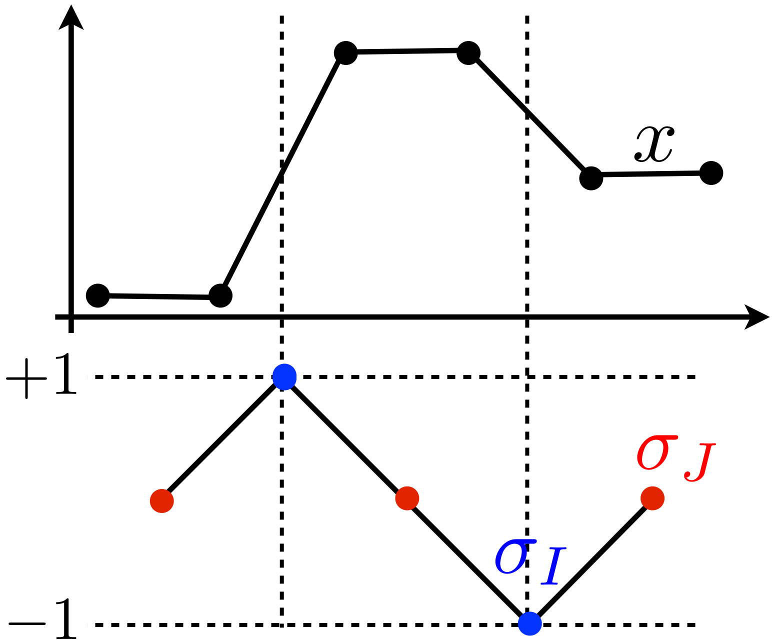
$$\{\mathcal{G}_J \setminus \dim \mathcal{G}_J = k\}$$



Signals with $k-1$ steps

Denoising $\Phi = \mathrm{Id}$.

$$\mathrm{IC}(s) = \|\sigma_J\|_\infty \quad \begin{cases} \sigma_I = \mathrm{sign}(D_I^* x) \\ \sigma_J = \Omega\, \mathrm{sign}(D_I^* x) \end{cases}$$



$\mathrm{IC}(\mathrm{sign}(D^* x)) < 1$

Support stability.

Example: TV Denoising in 1-D

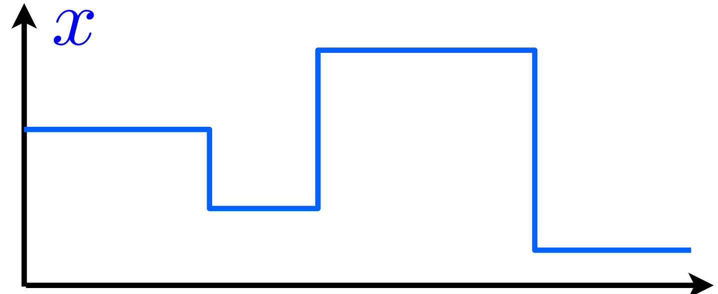
Discrete 1-D derivative:

$$D^* x = (x_i - x_{i-1})_i$$

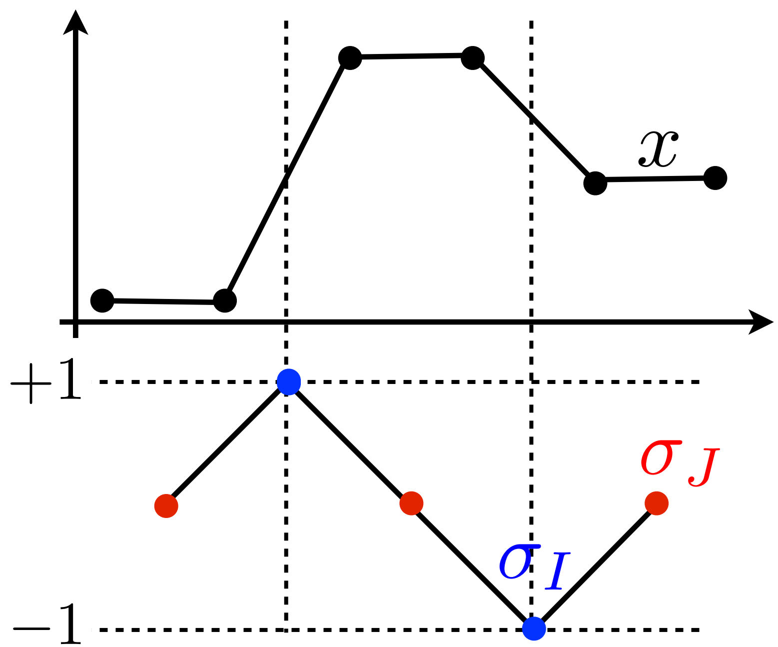
Denoising $\Phi = \mathrm{Id}$.

$$\mathrm{IC}(s) = \|\sigma_J\|_\infty \quad \begin{cases} \sigma_I = \mathrm{sign}(D_I^* x) \\ \sigma_J = \Omega\, \mathrm{sign}(D_I^* x) \end{cases}$$

$\{\mathcal{G}_J \setminus \dim \mathcal{G}_J = k\}$

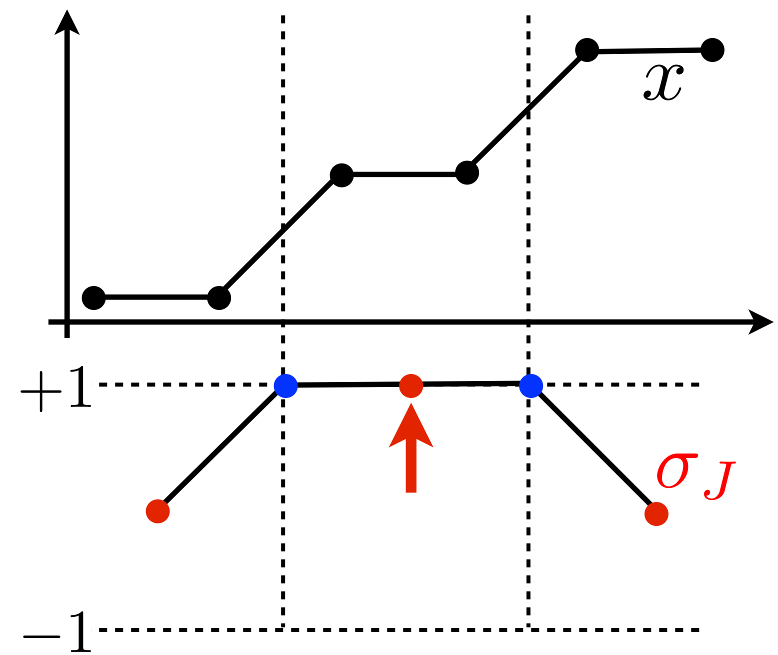


Signals with $k-1$ steps



$\mathrm{IC}(\mathrm{sign}(D^* x)) < 1$

Support stability.

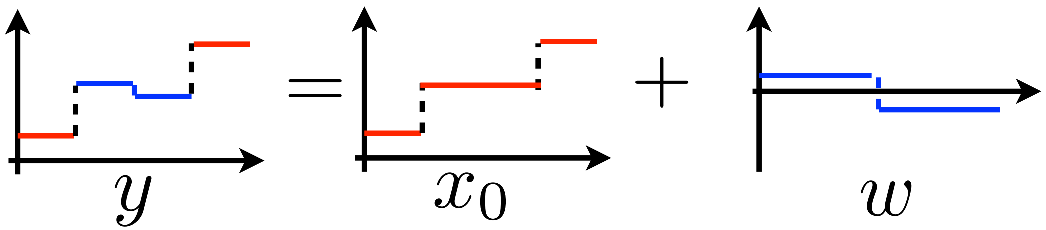


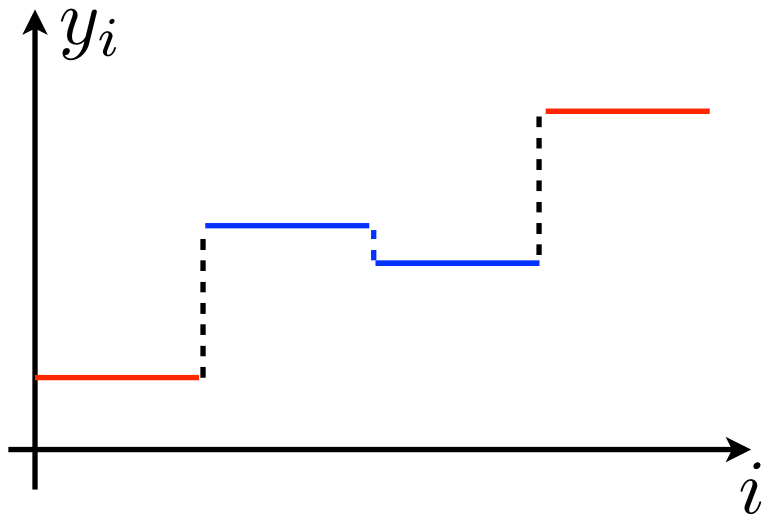
$\mathrm{IC}(\mathrm{sign}(D^* x)) = 1$

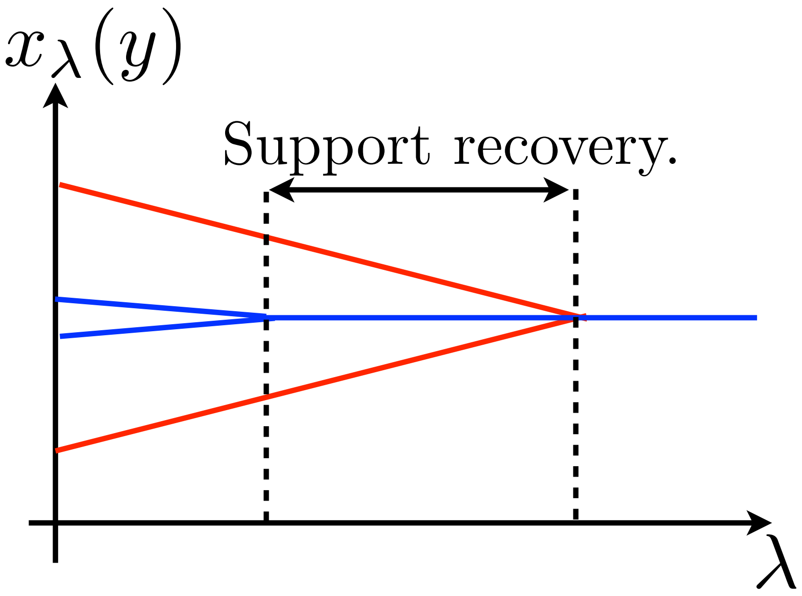
ℓ^2 stability only

Staircasing Example

Case IC$= 1$:

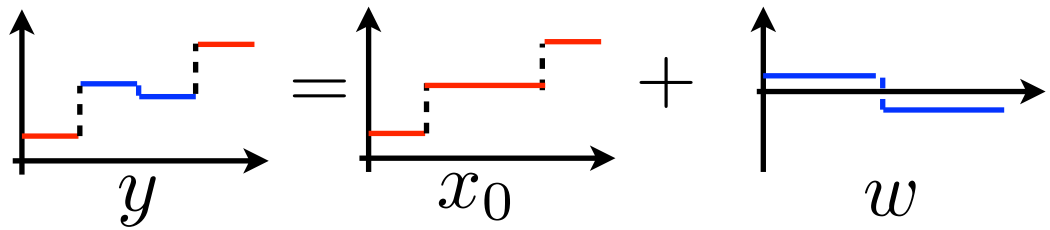


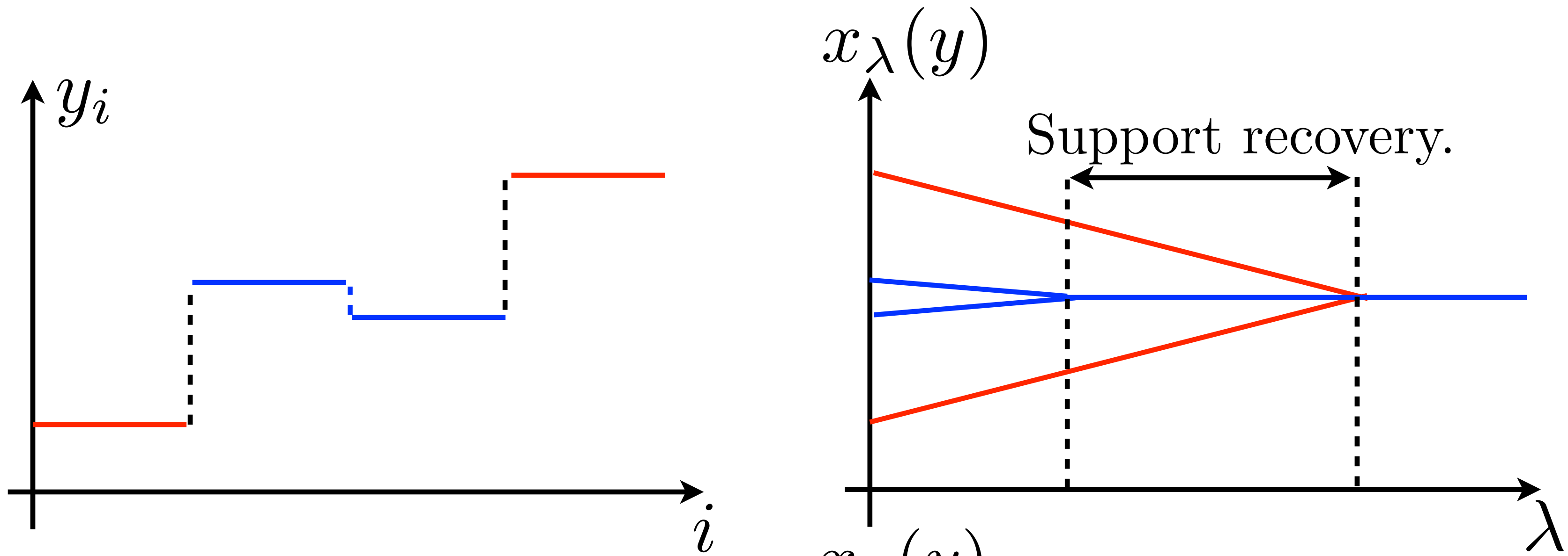


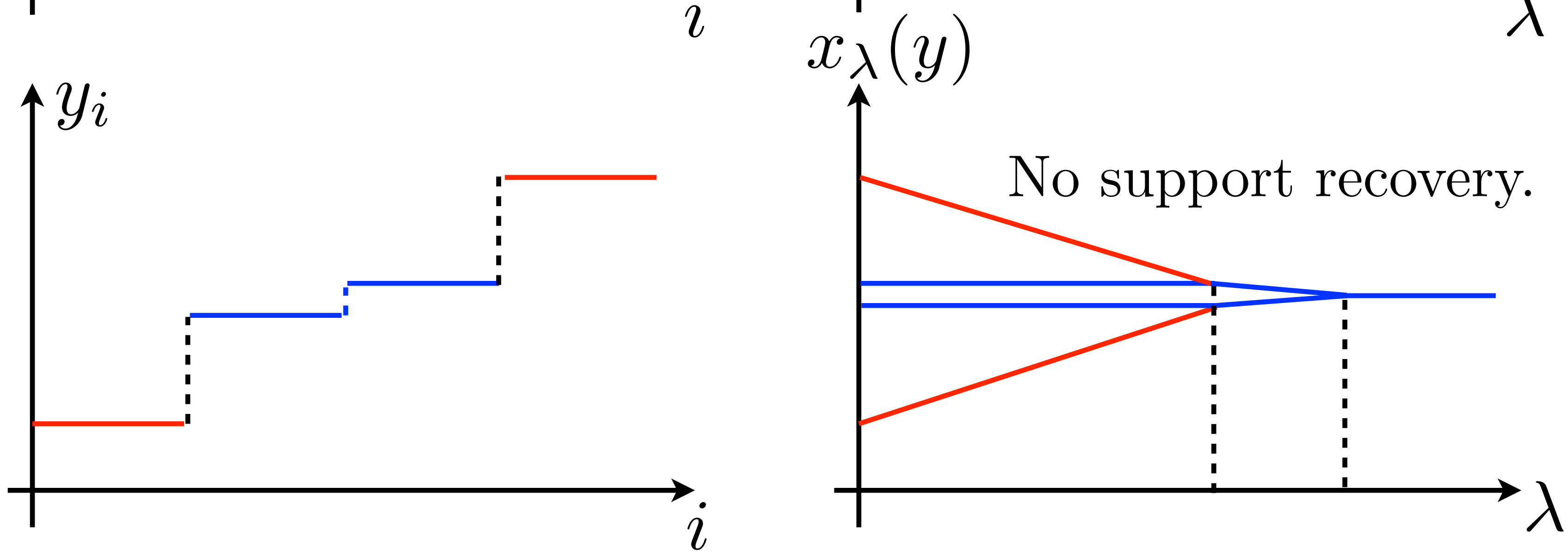


Staircasing Example

Case IC= 1:







Example: Total Variation in 2-D

Directional derivatives:

$$D_1^* x = (x_{i,j} - x_{i-1,j})_{i,j} \in \mathbb{R}^N$$

$$D_2^* x = (x_{i,j} - x_{i,j-1})_{i,j} \in \mathbb{R}^N$$

Gradient:

$$D^* x = (D_1^* x, D_2^* x) \in \mathbb{R}^{N \times 2}$$

Example: Total Variation in 2-D

Directional derivatives:

$$D_1^* x = (x_{i,j} - x_{i-1,j})_{i,j} \in \mathbb{R}^N$$

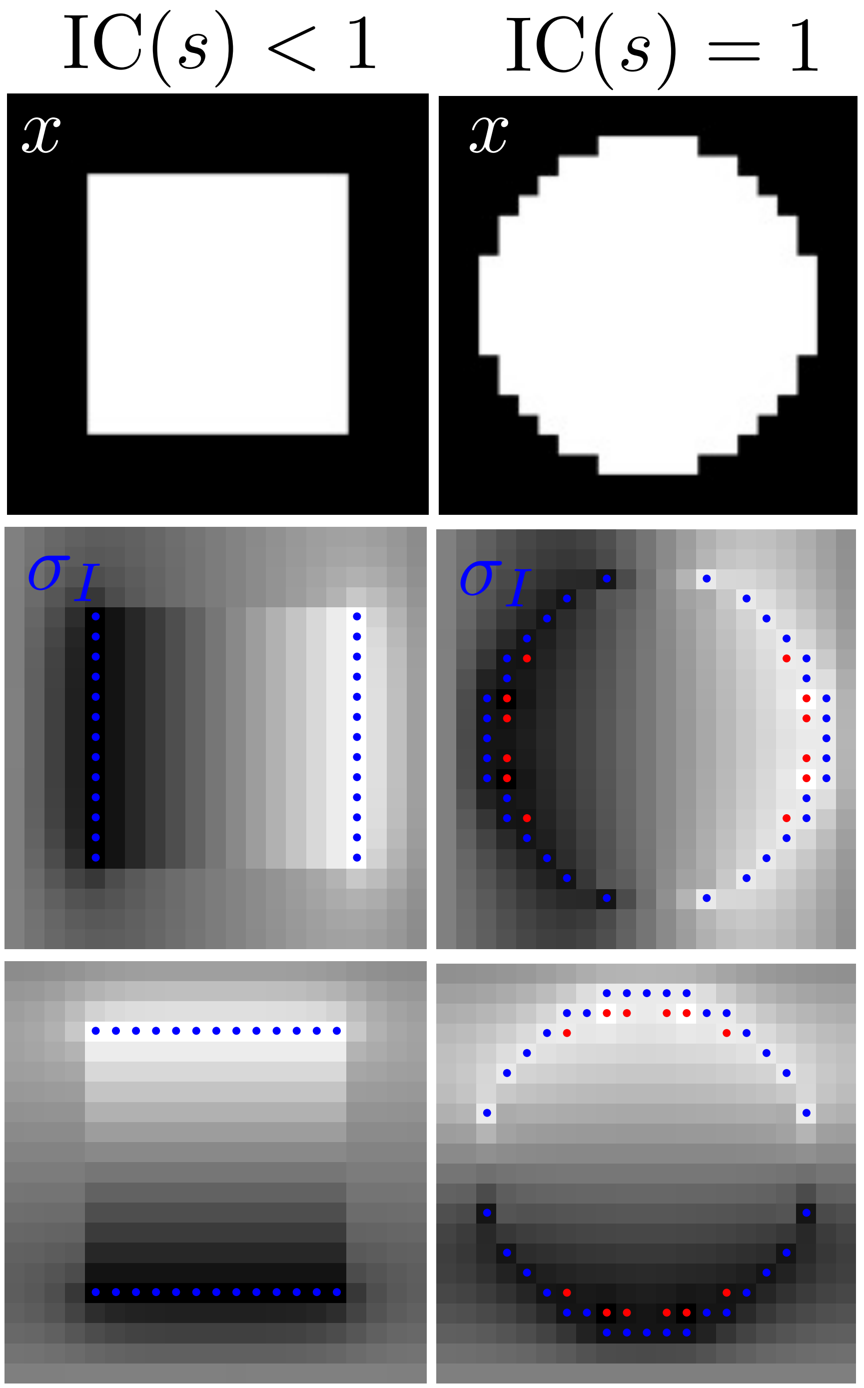
$$D_2^* x = (x_{i,j} - x_{i,j-1})_{i,j} \in \mathbb{R}^N$$

Gradient:

$$D^* x = (D_1^* x, D_2^* x) \in \mathbb{R}^{N \times 2}$$

Dual vector:

$$\begin{cases} \sigma_I = \text{sign}(D_I^* x) \\ \sigma_J = \Omega\, \text{sign}(D_I^* x) \end{cases}$$



Example: Fused Lasso

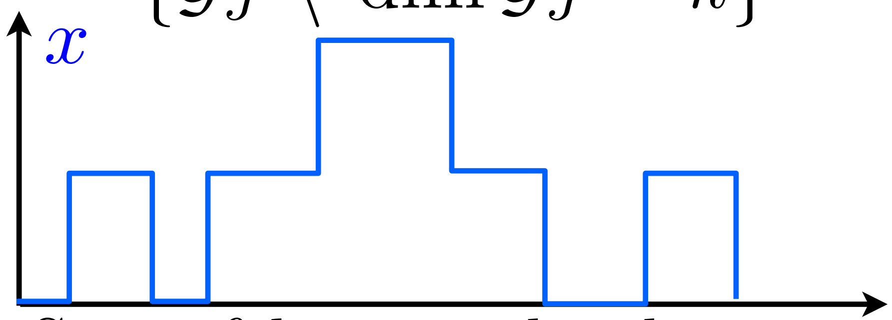
Total variation and ℓ^1 hybrid:

$$D^* x = (x_i - x_{i-1})_i \cup (\varepsilon x_i)_i$$

Compressed sensing:

Gaussian $\Phi \in \mathbb{R}^{Q \times N}$.

$\{\mathcal{G}_J \setminus \dim \mathcal{G}_J = k\}$



Sum of k interval indicators.

Example: Fused Lasso

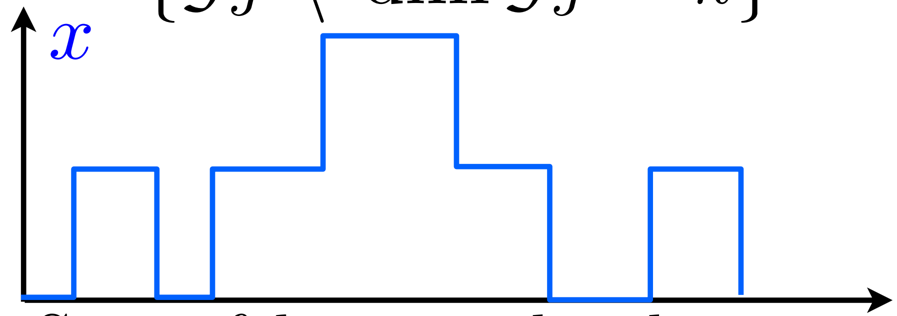
Total variation and ℓ^1 hybrid:

$$D^* x = (x_i - x_{i-1})_i \cup (\varepsilon x_i)_i$$

Compressed sensing:

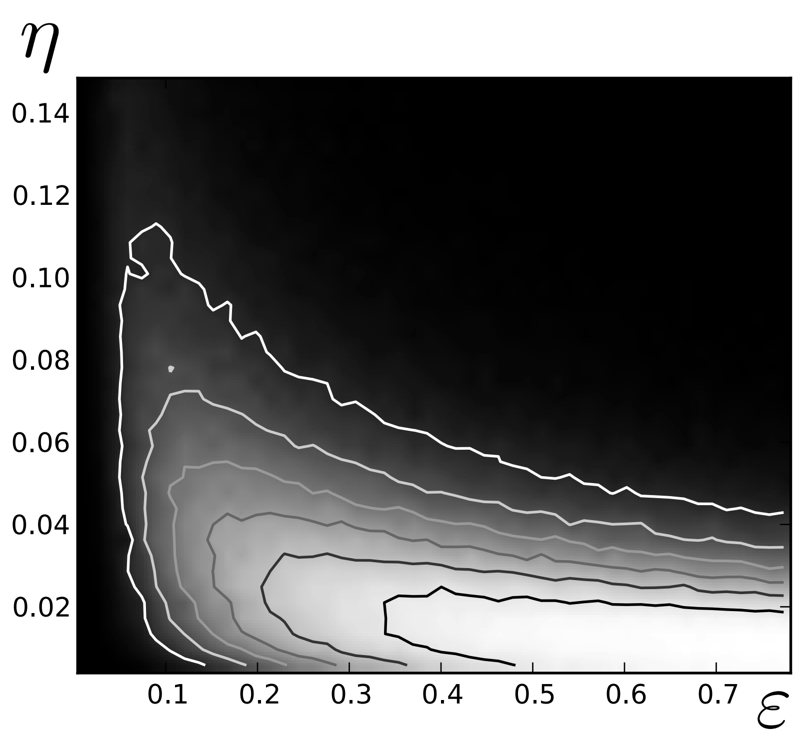
Gaussian $\Phi \in \mathbb{R}^{Q \times N}$.

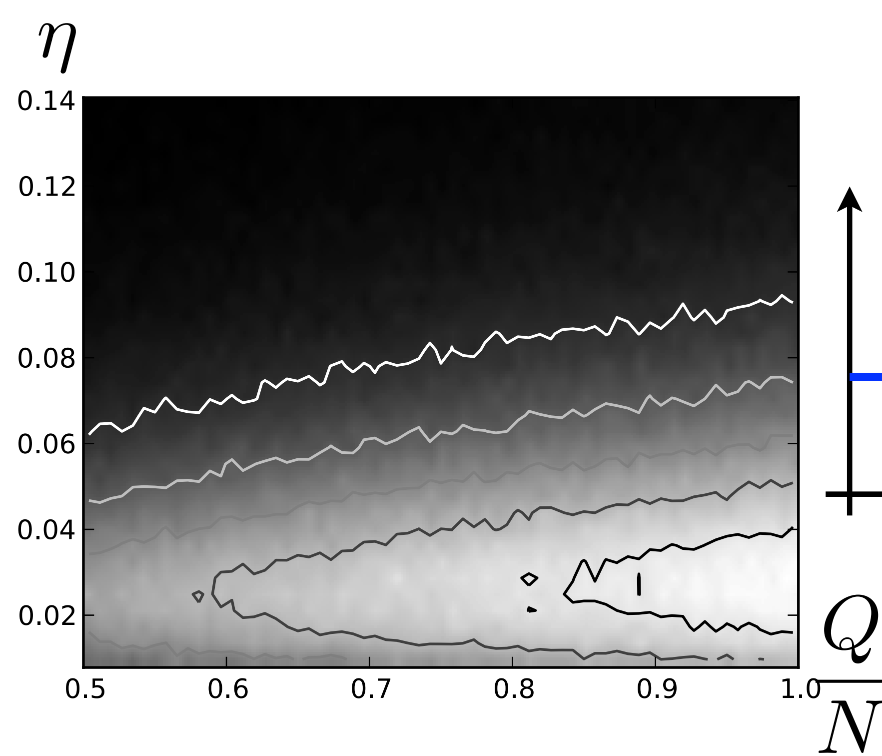
$\{\mathcal{G}_J \setminus \dim \mathcal{G}_J = k\}$

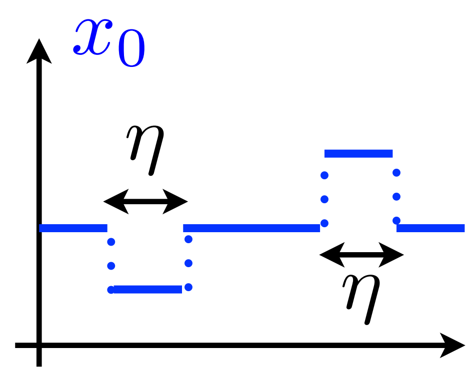


Sum of k interval indicators.

Probabilité $P(\eta, \varepsilon, \frac{Q}{N})$ of the even IC<1.



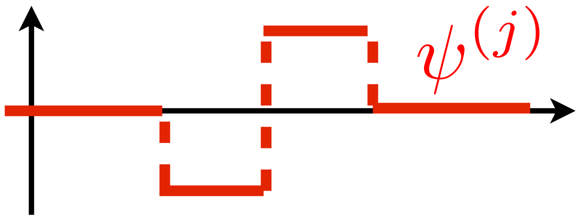




Example: Invariant Haar Analysis

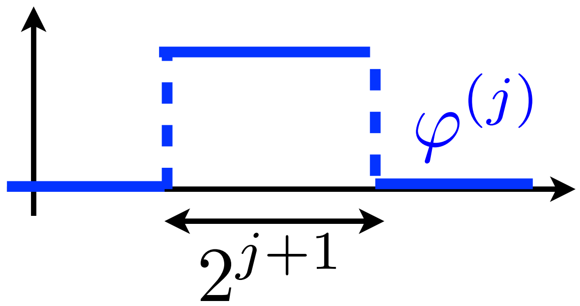
Haar wavelets:

$$\psi_i^{(j)} = \frac{1}{2^{\tau(j+1)}} \begin{cases} +1 & \text{if} \quad 0 \leqslant i < 2^j \\ -1 & \text{if} \quad -2^j \leqslant i < 0 \\ 0 & \text{otherwise.} \end{cases}$$



Haar TI analysis:

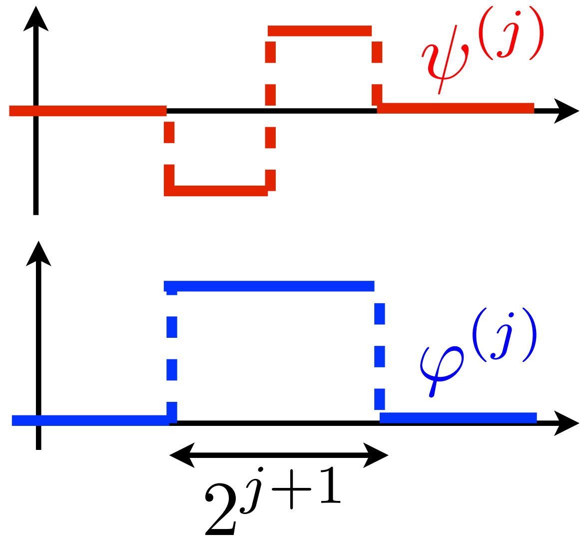
$$\|D^* x\|_1 = \sum_j \|x \star \psi^{(j)}\|_1 = \sum_j \|x \star \varphi^{(j)}\|_{\mathrm{TV}}$$



Example: Invariant Haar Analysis

Haar wavelets:

$$\psi_i^{(j)} = \frac{1}{2^{\tau(j+1)}} \begin{cases} +1 & \text{if} \quad 0 \leqslant i < 2^j \\ -1 & \text{if} \quad -2^j \leqslant i < 0 \\ 0 & \text{otherwise.} \end{cases}$$

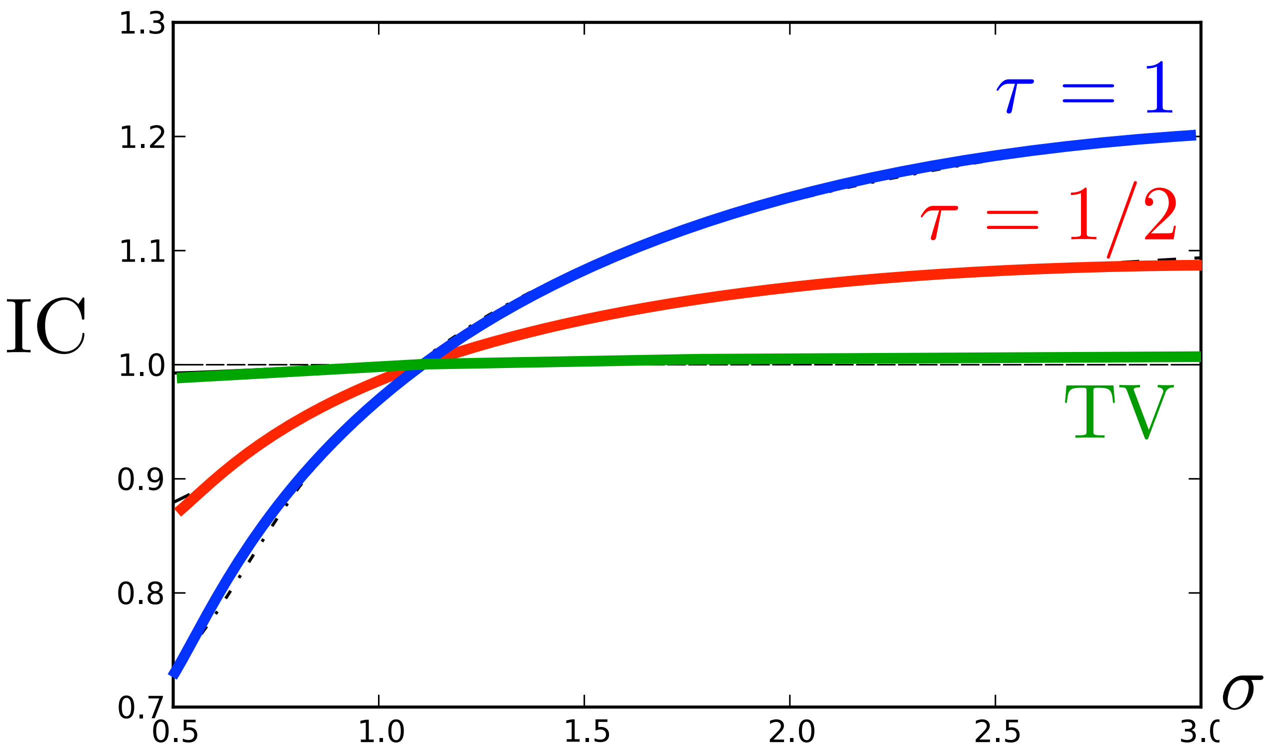


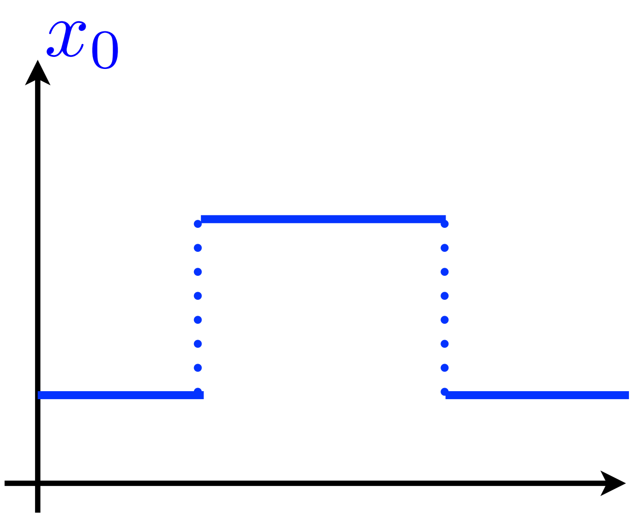
Haar TI analysis:

$$\|D^* x\|_1 = \sum_j \|x \star \psi^{(j)}\|_1 = \sum_j \|x \star \varphi^{(j)}\|_{\mathrm{TV}}$$

De-blurring:

$$\Phi x = \varphi \star x, \quad \varphi(t) \propto e^{-\frac{t^2}{2\sigma^2}}$$



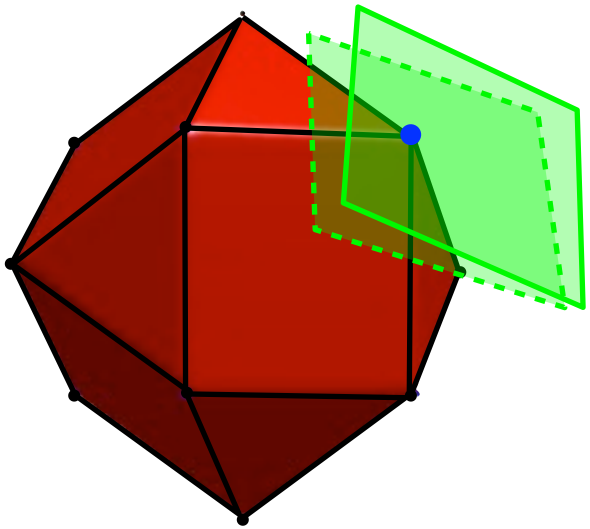


Conclusion

SURE risk estimation:

Computing risk estimates

$\iff$ Sensitivity analysis.



Conclusion

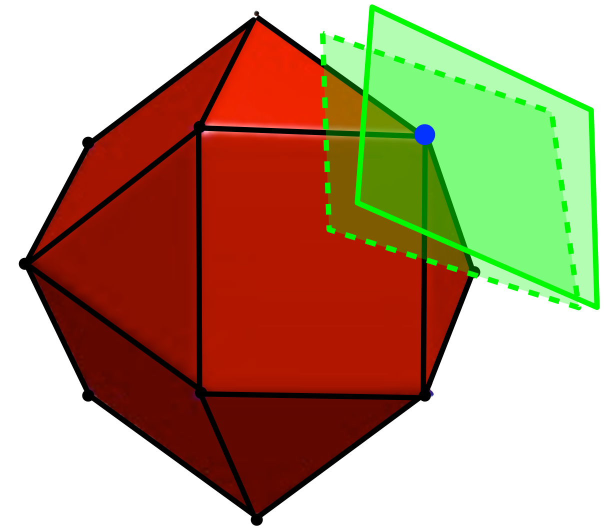
SURE risk estimation:

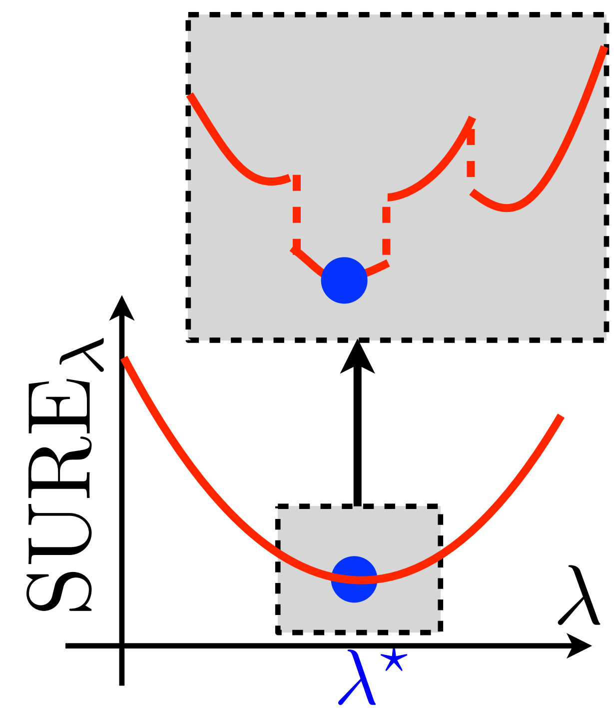
Computing risk estimates

$\iff$ Sensitivity analysis.

Open problem:

Fast algorithms to optimize λ.



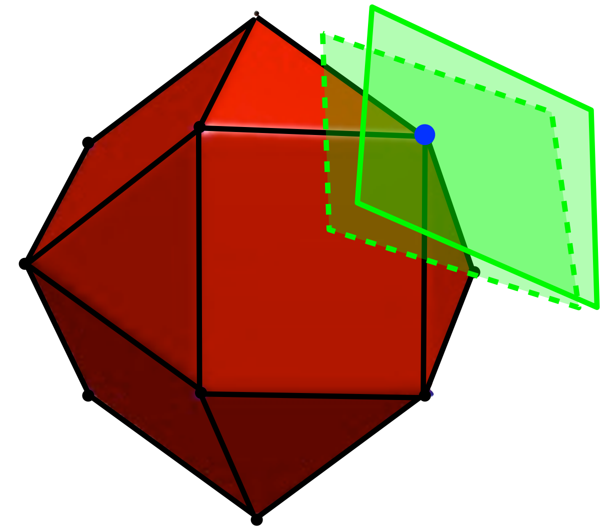


Conclusion

SURE risk estimation:

Computing risk estimates

$\iff$ Sensitivity analysis.

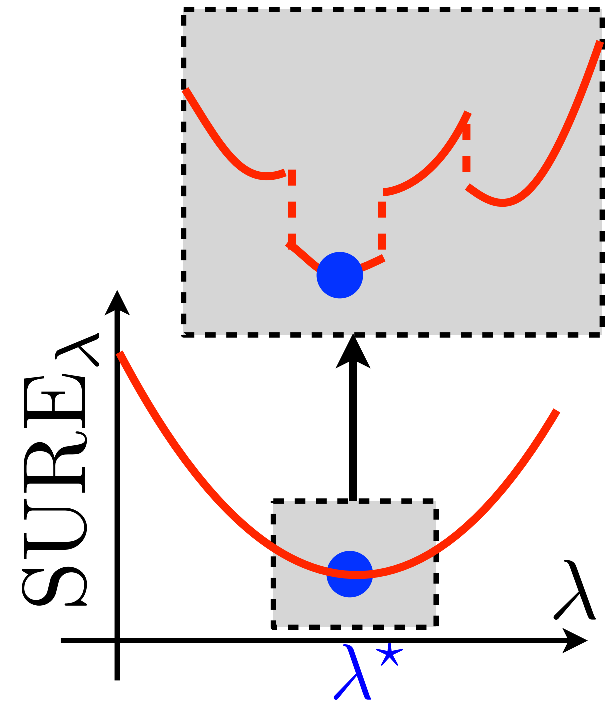


Open problem:

Fast algorithms to optimize λ.

Analysis vs. synthesis regularization:

Analysis support is less stable.



Conclusion

SURE risk estimation:

Computing risk estimates

$\iff$ Sensitivity analysis.

Open problem:

Fast algorithms to optimize λ.

Analysis vs. synthesis regularization:

Analysis support is less stable.

Open problem:

Robustness without support stability.