

Parameter Selection for Sparse Regularization of Inverse Problems

Joint work with:

Gabriel Peyré

Samuel Vaiter Jalal Fadili
Charles {Dossal, Deledalle}



www.numerical-tours.com



Overview

- *Inverse Problems*
- Stein Unbiased Risk Estimators (SURE)
- SURE for Sparse Regularization

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)

Hypothesis: $W \sim \mathcal{N}(0, \sigma^2 \text{Id}_P)$, σ is known.

Inverse Problems

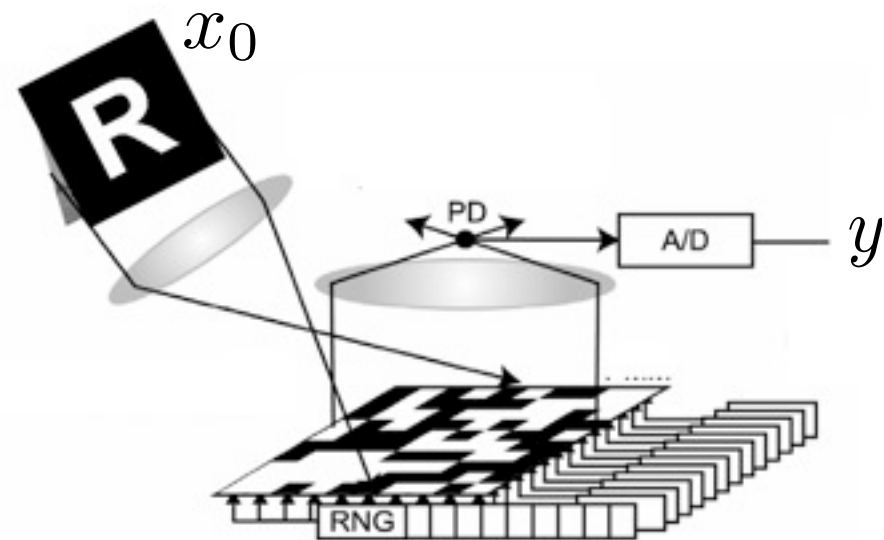
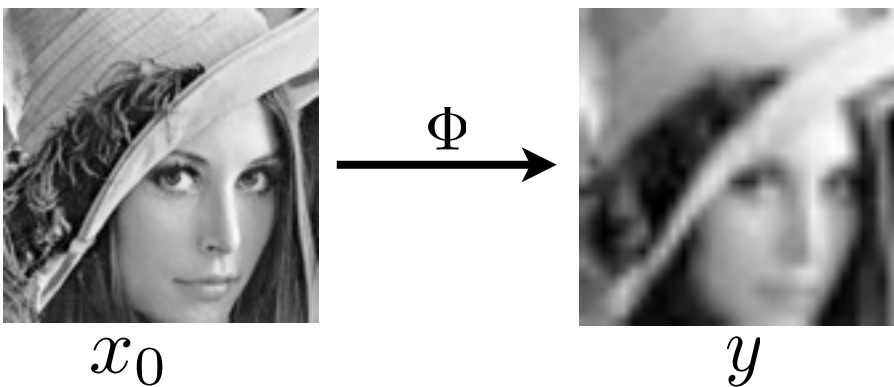
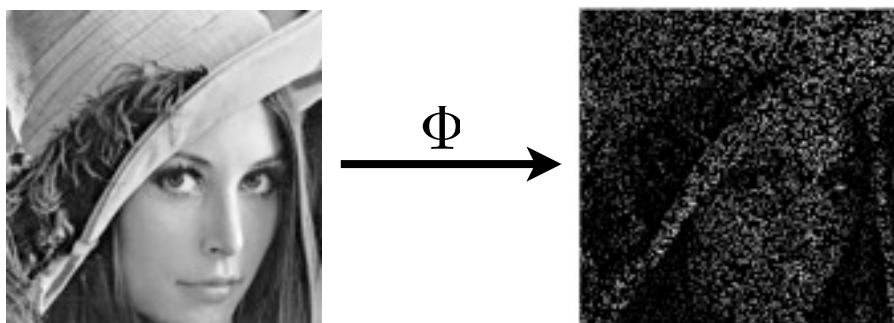
Recovering $x_0 \in \mathbb{R}^N$ from noisy observations

$$Y = \Phi x_0 + W \in \mathbb{R}^P$$

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

Hypothesis: $W \sim \mathcal{N}(0, \sigma^2 \text{Id}_P)$, σ is known.

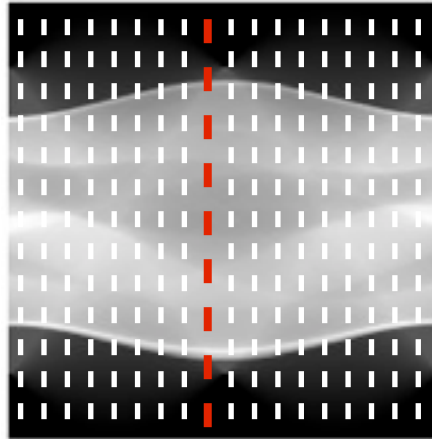
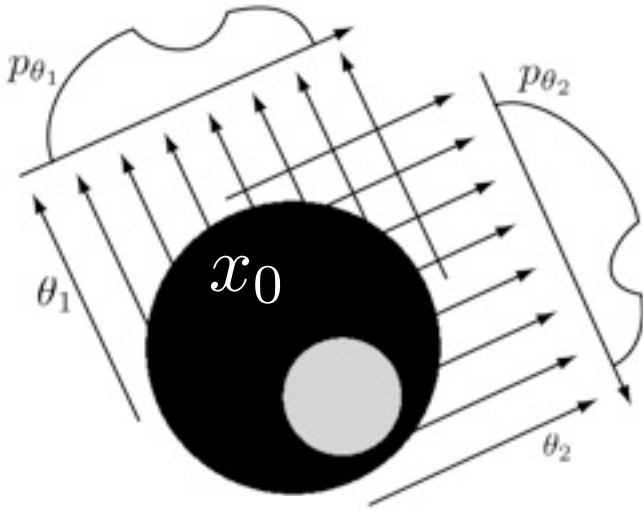
Examples: Inpainting, super-resolution, compressed-sensing



Inverse Problem in Medical Imaging

Tomography projection:

$$\Phi x_0 = (p_{\theta_k})_{1 \leq k \leq K}$$

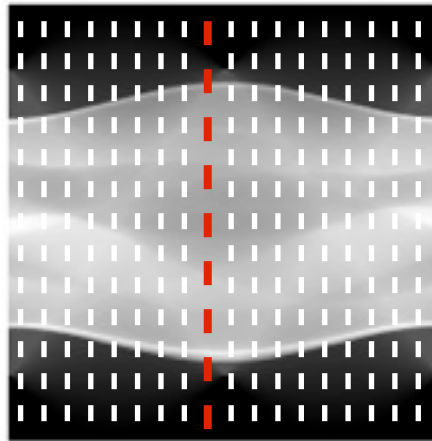
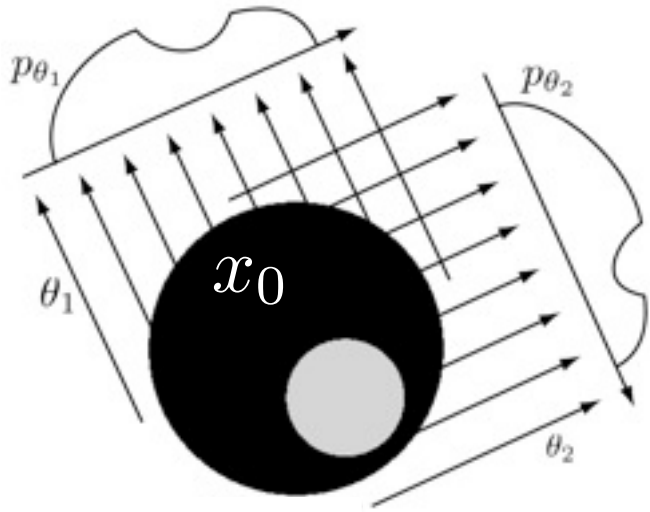


θ_k

Inverse Problem in Medical Imaging

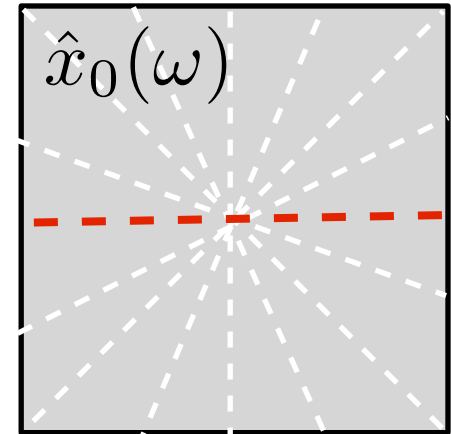
Tomography projection:

$$\Phi x_0 = (p_{\theta_k})_{1 \leq k \leq K}$$



θ_k

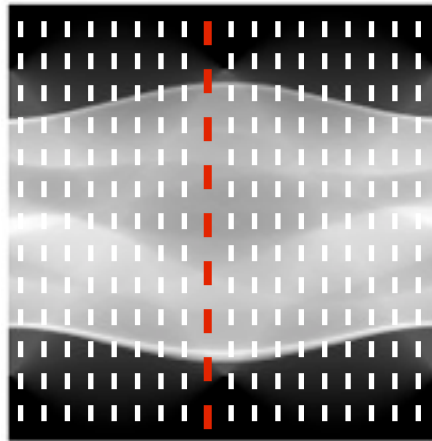
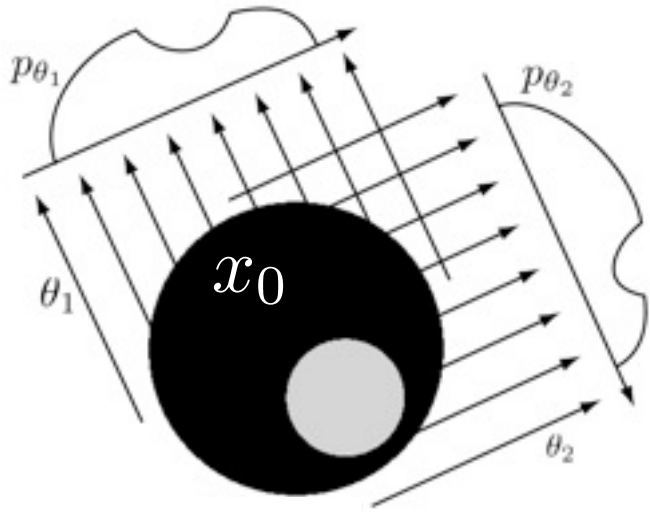
Fourier slice
 \iff
theorem



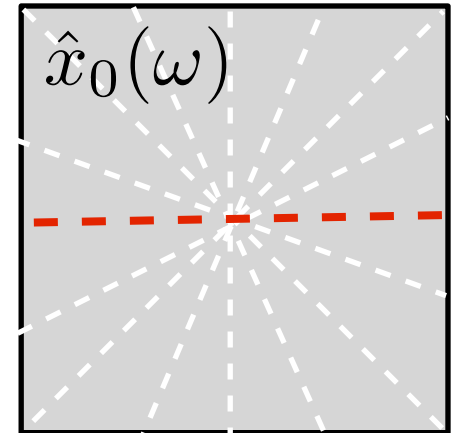
Inverse Problem in Medical Imaging

Tomography projection:

$$\Phi x_0 = (p_{\theta_k})_{1 \leq k \leq K}$$

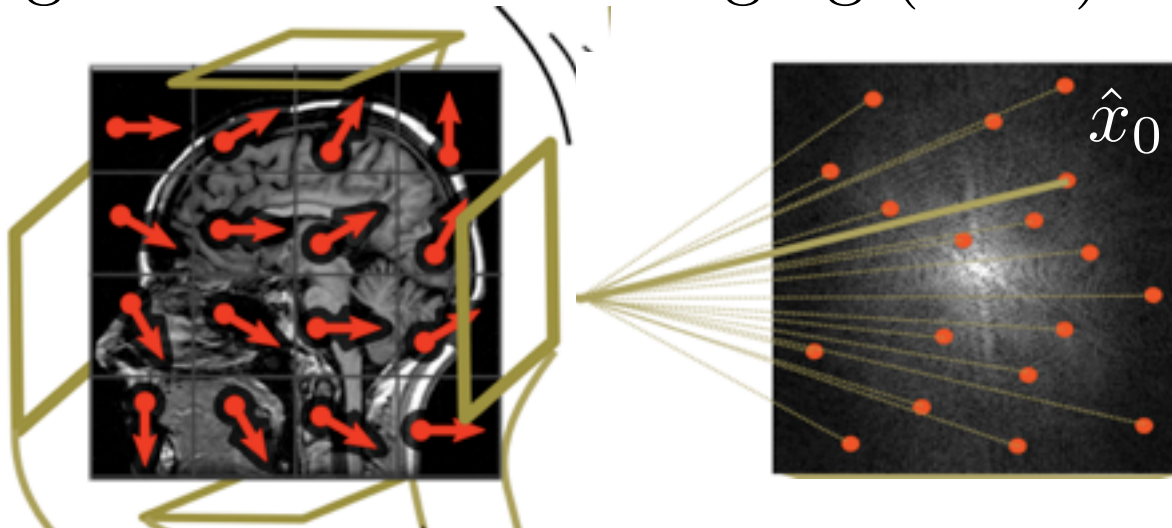


Fourier slice
 \iff
theorem



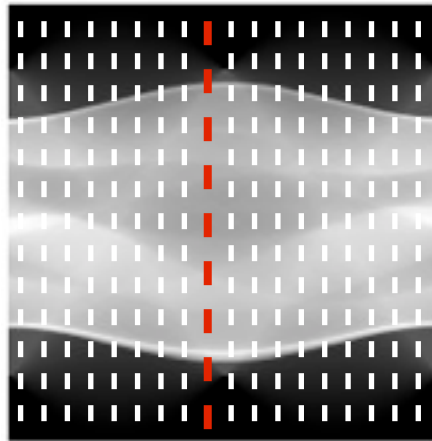
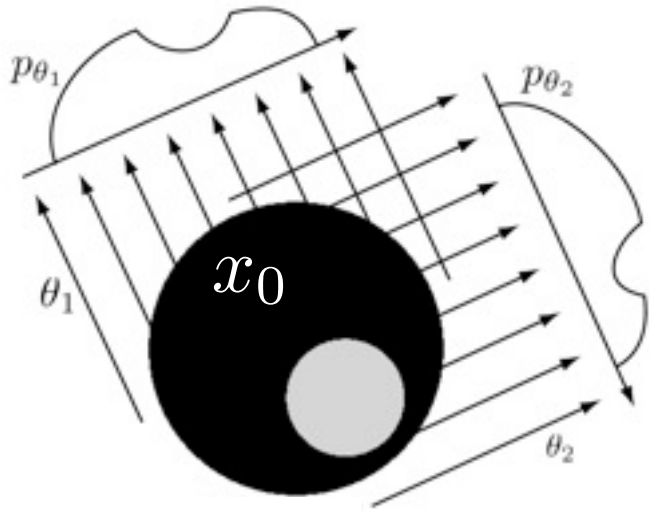
Magnetic resonance imaging (MRI):

$$\Phi x_0 = (\hat{x}_0(\omega))_{\omega \in \Omega}$$

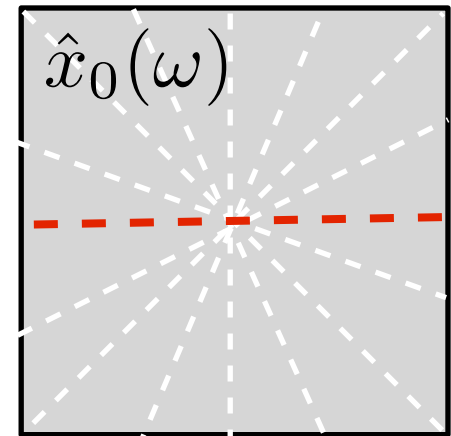


Inverse Problem in Medical Imaging

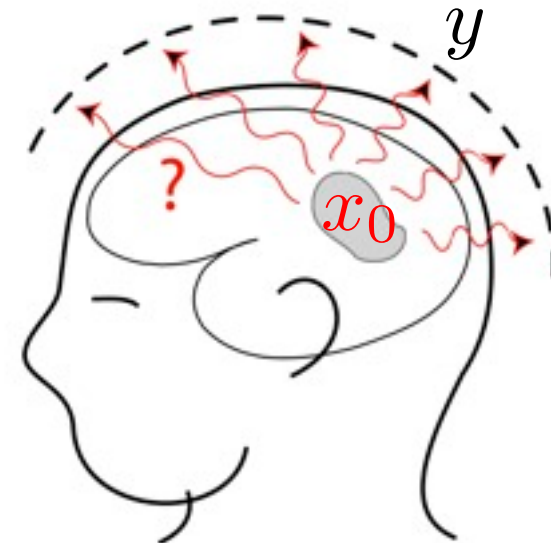
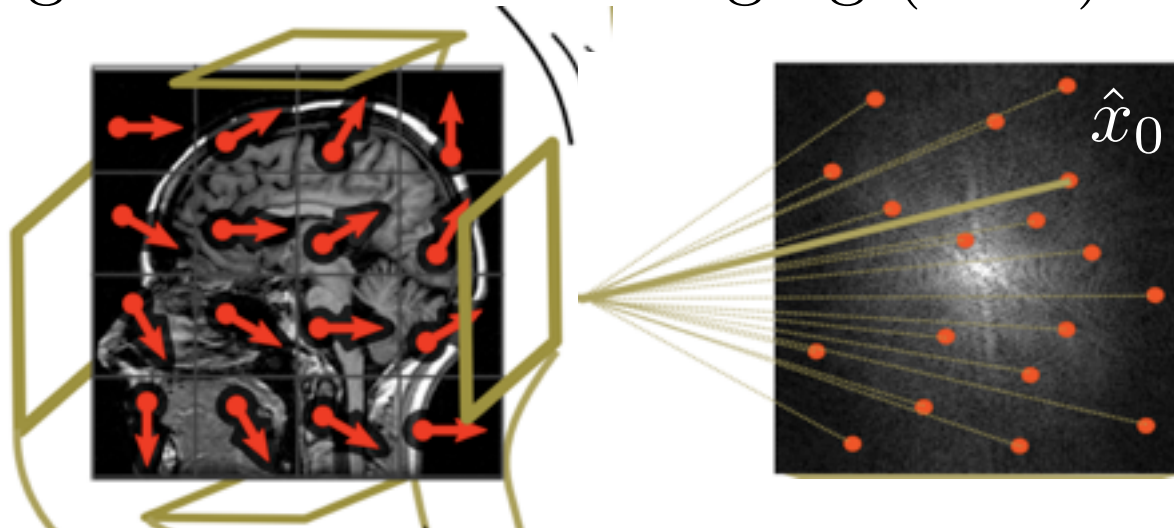
Tomography projection: $\Phi x_0 = (p_{\theta_k})_{1 \leq k \leq K}$



Fourier slice
 \iff
theorem



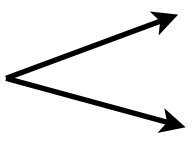
Magnetic resonance imaging (MRI): $\Phi x_0 = (\hat{x}_0(\omega))_{\omega \in \Omega}$



Other examples: MEG, EEG, ...

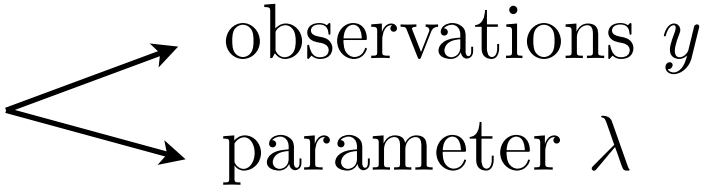
Estimators

Measures: $Y = \Phi x_0 + W$, $W \sim \mathcal{N}(0, \sigma^2 \text{Id}_P)$.

Estimator: $x_\lambda(y)$ depends only on  observations y
parameter λ

Estimators

Measures: $Y = \Phi x_0 + W$, $W \sim \mathcal{N}(0, \sigma^2 \text{Id}_P)$.

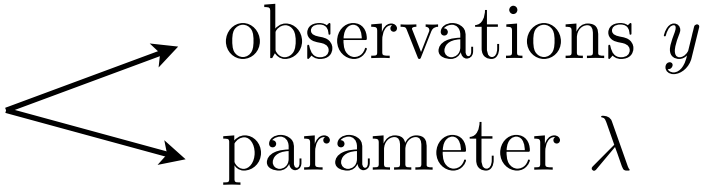
Estimator: $x_\lambda(y)$ depends only on  observations y
parameter λ

Example: variational methods

$$x_\lambda(y) \in \operatorname{argmin}_{x \in \mathbb{R}^N} \underbrace{\frac{1}{2} \|y - \Phi x\|^2}_{\text{Data fidelity}} + \lambda \underbrace{J(x)}_{\text{Regularity}}$$

Estimators

Measures: $Y = \Phi x_0 + W$, $W \sim \mathcal{N}(0, \sigma^2 \text{Id}_P)$.

Estimator: $x_\lambda(y)$ depends only on  observations y
parameter λ

Example: variational methods


$$x_\lambda(y) \in \operatorname{argmin}_{x \in \mathbb{R}^N} \underbrace{\frac{1}{2} \|y - \Phi x\|^2}_{\text{Data fidelity}} + \lambda \underbrace{J(x)}_{\text{Regularity}}$$

Goal: set up λ to minimize $\mathbb{E}_W(\|x_0 - x_\lambda(Y)\|^2)$.

Method: use risk estimators computed from y alone.

Estimators

Measures: $Y = \Phi x_0 + W$, $W \sim \mathcal{N}(0, \sigma^2 \text{Id}_P)$.

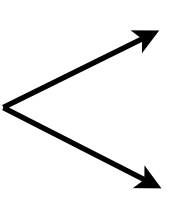
Estimator: $x_\lambda(y)$ depends only on  observations y
parameter λ

Example: variational methods

$$x_\lambda(y) \in \operatorname{argmin}_{x \in \mathbb{R}^N} \underbrace{\frac{1}{2} \|y - \Phi x\|^2}_{\text{Data fidelity}} + \lambda \underbrace{J(x)}_{\text{Regularity}}$$

Goal: set up λ to minimize $\mathbb{E}_W(\|x_0 - x_\lambda(Y)\|^2)$.

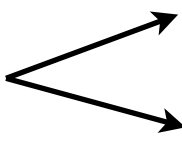
Method: use risk estimators computed from y alone.

 Theoretical approaches (closed form expressions).

Numerical approaches (recursive differentiation).

Estimators

Measures: $Y = \Phi x_0 + W$, $W \sim \mathcal{N}(0, \sigma^2 \text{Id}_P)$.

Estimator: $x_\lambda(y)$ depends only on  observations y
parameter λ

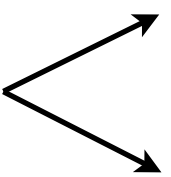
Example: variational methods

$$x_\lambda(y) \in \operatorname{argmin}_{x \in \mathbb{R}^N} \underbrace{\frac{1}{2} \|y - \Phi x\|^2}_{\text{Data fidelity}} + \lambda \underbrace{J(x)}_{\text{Regularity}}$$

Goal: set up λ to minimize $\mathbb{E}_W(\|x_0 - x_\lambda(Y)\|^2)$.

Method: use risk estimators computed from y alone.

Today's focus

 Theoretical approaches (closed form expressions).

Numerical approaches (recursive differentiation).

Overview

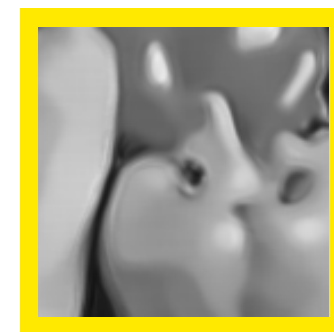
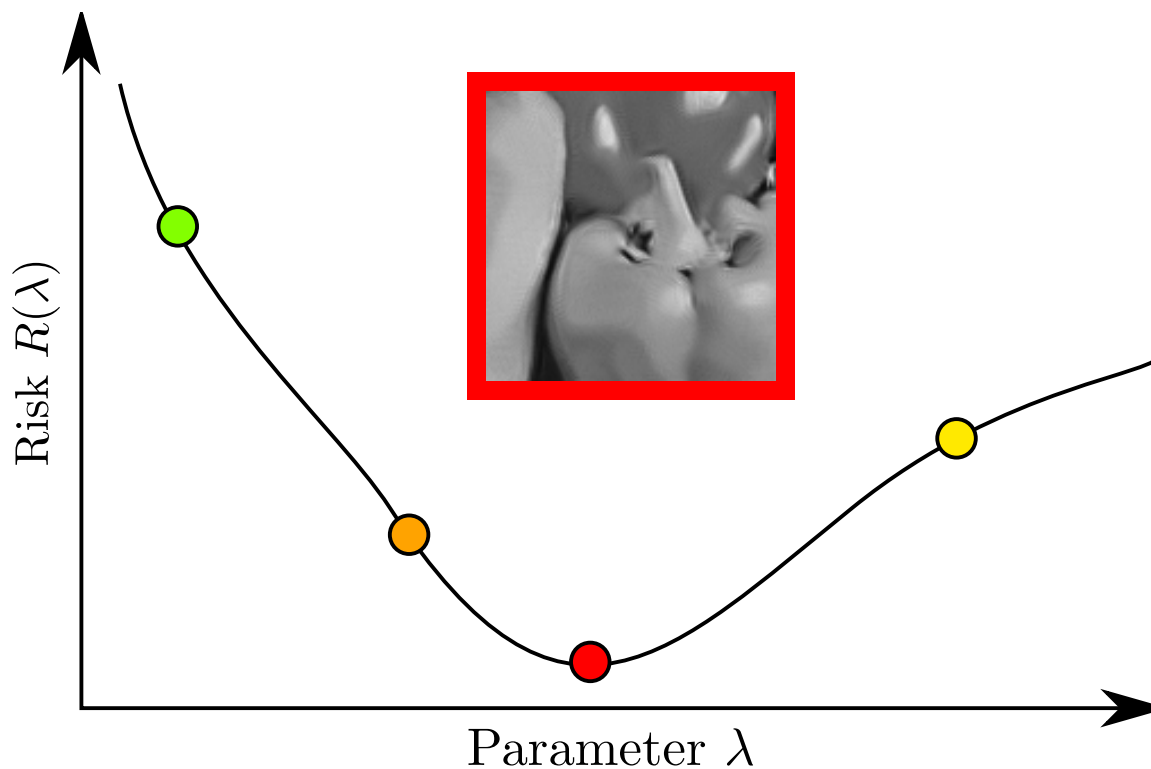
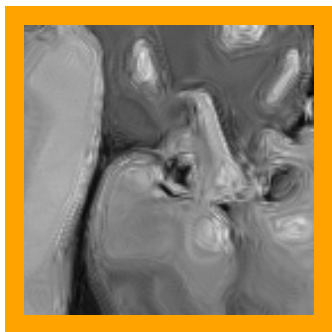
- Inverse Problems
- *Stein Unbiased Risk Estimators (SURE)*
- SURE for Sparse Regularization

Risk Minimization

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

$$\lambda^* = \underset{\lambda}{\operatorname{argmin}} R(\lambda)$$

Plugin-estimator: $x_{\lambda^*}(Y)$

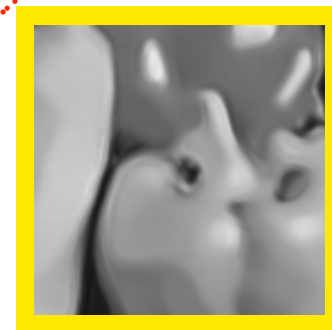
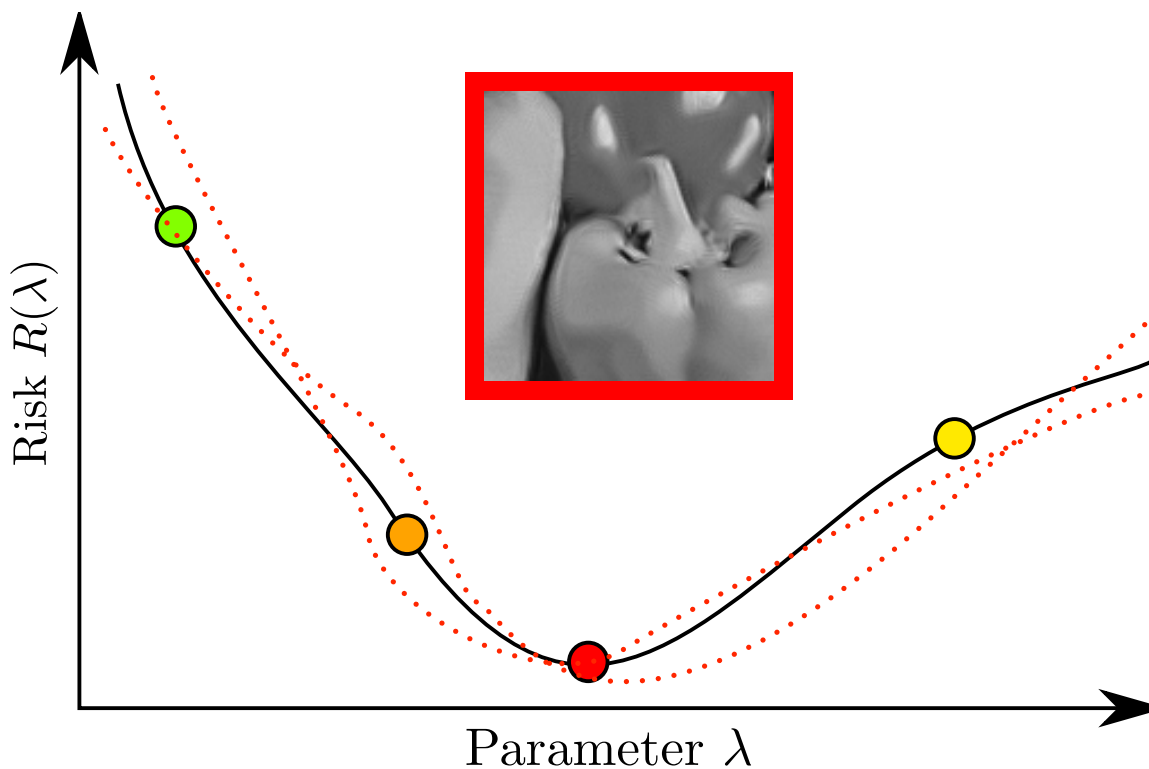
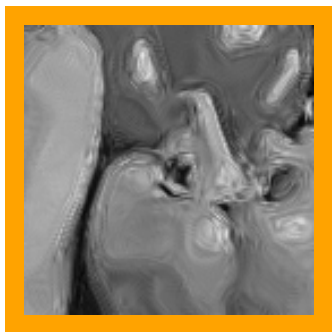


Risk Minimization

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

$$\lambda^* = \operatorname{argmin}_\lambda R(\lambda)$$

Plugin-estimator: $x_{\lambda^*}(Y)$



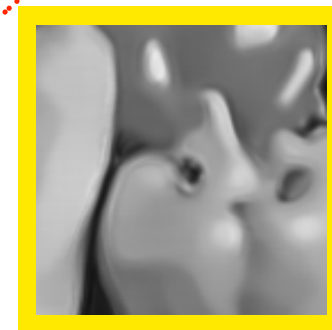
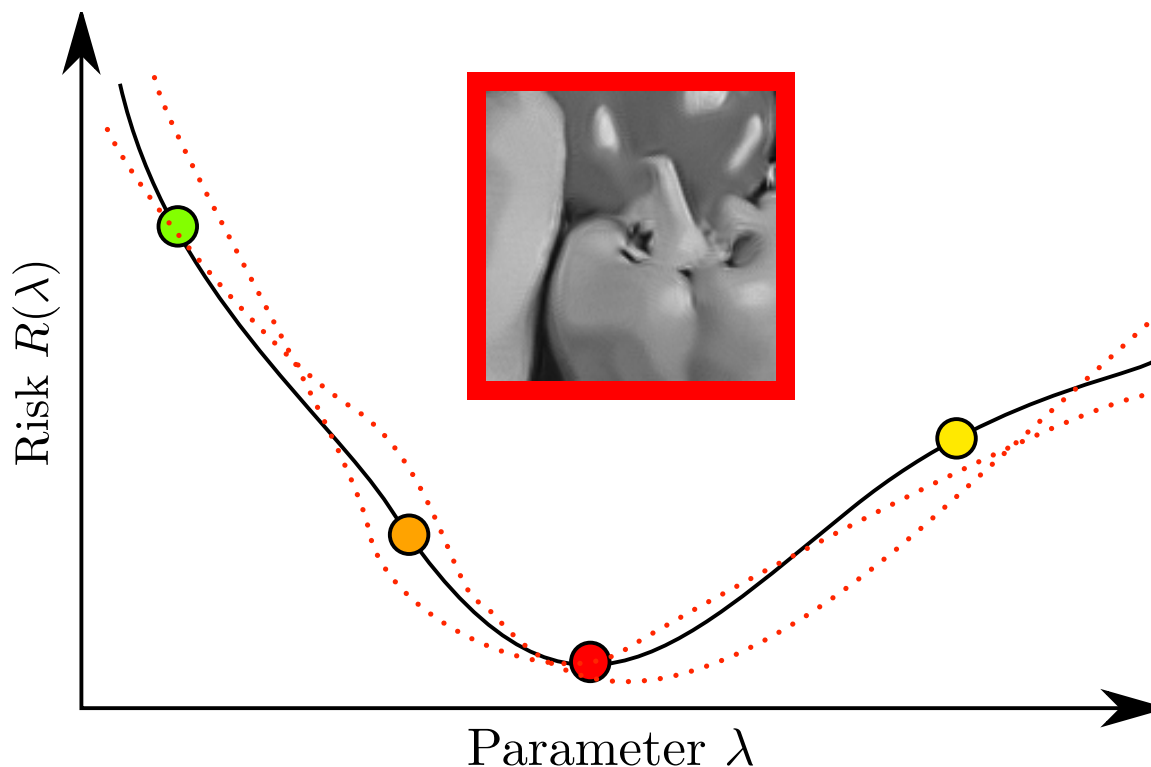
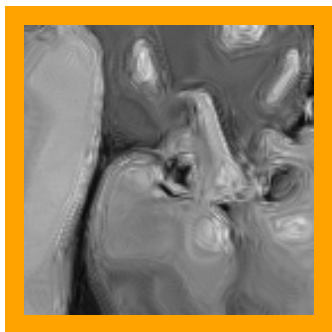
But: \mathbb{E}_W is not accessible \rightarrow use one observation.

Risk Minimization

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

$$\lambda^* = \underset{\lambda}{\operatorname{argmin}} R(\lambda)$$

Plugin-estimator: $x_{\lambda^*}(Y)$



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:

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

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

Theorem: [Stein, 1981] $\mathbb{E}_W(\text{SURE}_\lambda(Y)) = \mathbb{E}_W(\|\Phi x_0 - \mu_\lambda(Y)\|^2)$

Prediction Risk Estimation

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

Sensitivity analysis: if μ_λ is weakly differentiable

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

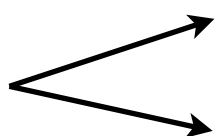
Stein Unbiased Risk Estimator:

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

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

Theorem: [Stein, 1981] $\mathbb{E}_W(\text{SURE}_\lambda(Y)) = \mathbb{E}_W(\|\Phi x_0 - \mu_\lambda(Y)\|^2)$

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 $\ker(\Phi)^\perp$

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

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

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

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

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

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

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

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

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

Overview

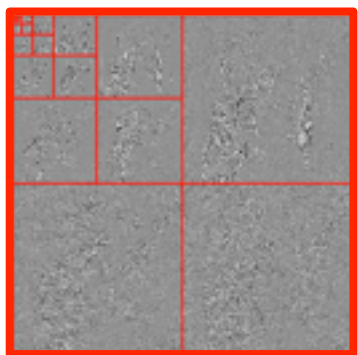
- Inverse Problems
- Stein Unbiased Risk Estimators (SURE)
- **SURE for Sparse Regularization**

Sparse Regularizations

Synthesis regularization

$$J(\boldsymbol{x}) = \|\boldsymbol{x}\|_1 = \sum_i |\boldsymbol{x}_i|$$

$$\min_{\boldsymbol{x} \in \mathbb{R}^N} \frac{1}{2} \|y - (\Phi\Psi)\boldsymbol{x}\|_2^2 + \lambda \|\boldsymbol{x}\|_1$$



Coefficients \boldsymbol{x}

Ψ



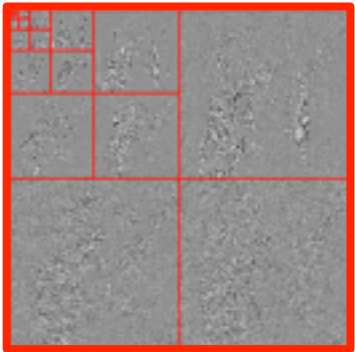
Image $\Psi\boldsymbol{x}$

Sparse Regularizations

Synthesis regularization

$$J(\boldsymbol{x}) = \|\boldsymbol{x}\|_1 = \sum_i |\boldsymbol{x}_i|$$

$$\min_{\boldsymbol{x} \in \mathbb{R}^N} \frac{1}{2} \|y - (\Phi\Psi)\boldsymbol{x}\|_2^2 + \lambda \|\boldsymbol{x}\|_1$$



Ψ



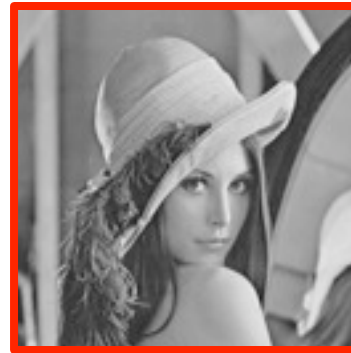
Coefficients \boldsymbol{x}

Image $\Psi\boldsymbol{x}$

Analysis regularization

$$J(\boldsymbol{x}) = \|D^*\boldsymbol{x}\|_1 = \sum_i |\langle \boldsymbol{x}, d_i \rangle|$$

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



D^*



Image \boldsymbol{x}

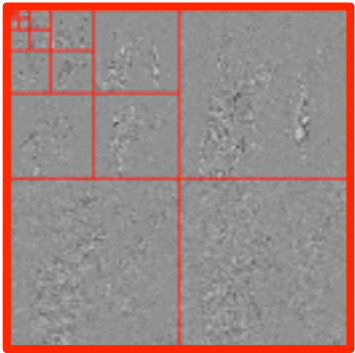
Gradient $D^*\boldsymbol{x}$

Sparse Regularizations

Synthesis regularization

$$J(\boldsymbol{x}) = \|\boldsymbol{x}\|_1 = \sum_i |\boldsymbol{x}_i|$$

$$\min_{\boldsymbol{x} \in \mathbb{R}^N} \frac{1}{2} \|y - (\Phi\Psi)\boldsymbol{x}\|_2^2 + \lambda \|\boldsymbol{x}\|_1$$



Ψ



Coefficients \boldsymbol{x}

Image $\Psi\boldsymbol{x}$

Analysis regularization

$$J(\boldsymbol{x}) = \|D^*\boldsymbol{x}\|_1 = \sum_i |\langle \boldsymbol{x}, d_i \rangle|$$

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



D^*



Image \boldsymbol{x}

Gradient $D^*\boldsymbol{x}$

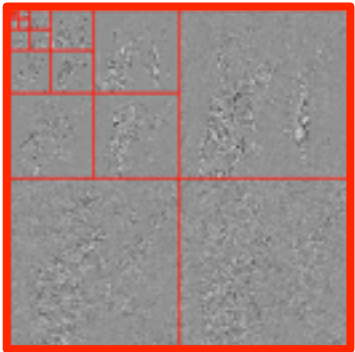
→ Unless $D = \Psi$ is orthogonal, produces different results.

Sparse Regularizations

Synthesis regularization

$$J(\boldsymbol{x}) = \|\boldsymbol{x}\|_1 = \sum_i |\boldsymbol{x}_i|$$

$$\min_{\boldsymbol{x} \in \mathbb{R}^N} \frac{1}{2} \|y - (\Phi\Psi)\boldsymbol{x}\|_2^2 + \lambda \|\boldsymbol{x}\|_1$$



Ψ



Coefficients \boldsymbol{x}

Image $\Psi\boldsymbol{x}$

Analysis regularization

$$J(\boldsymbol{x}) = \|D^*\boldsymbol{x}\|_1 = \sum_i |\langle \boldsymbol{x}, d_i \rangle|$$

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



D^*



Image \boldsymbol{x}

Gradient $D^*\boldsymbol{x}$

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

→ Analysis more general when using $\Phi\Psi$ as operator.

Sparse Analysis Local Variations

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

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

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

Sparse Analysis Local Variations

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

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

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

Theorem: μ_λ is Lipschitz (hence weakly differentiable).

Sparse Analysis Local Variations

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

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

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

Theorem: μ_λ is Lipschitz (hence weakly differentiable).

Questions:

—→ Find a closed form expression for $\partial\mu_\lambda(y)$.

—→ Prove that this formula is valid almost everywhere.

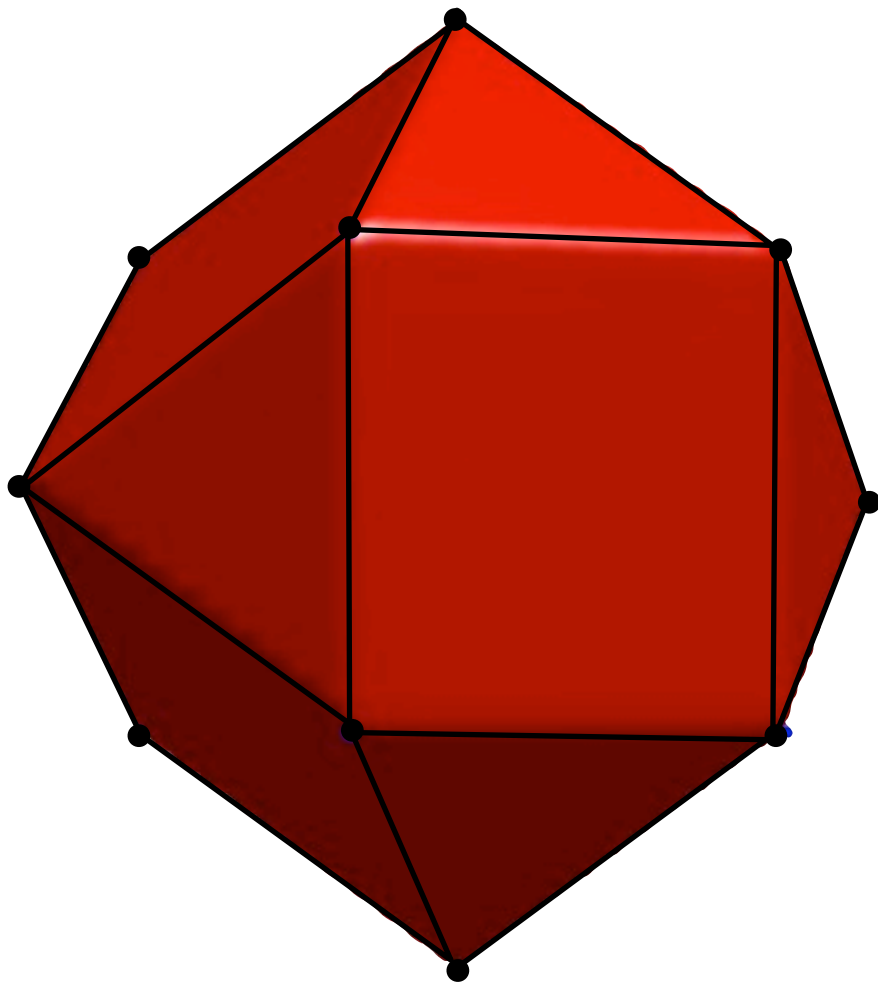
Polytope of TV in 1-D

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

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

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

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



Copyright © Mael

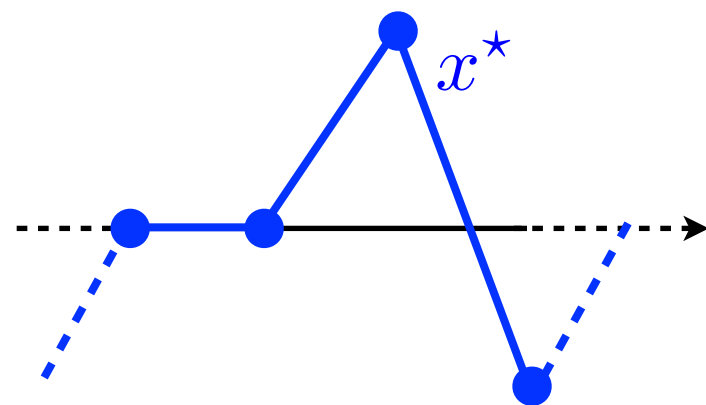
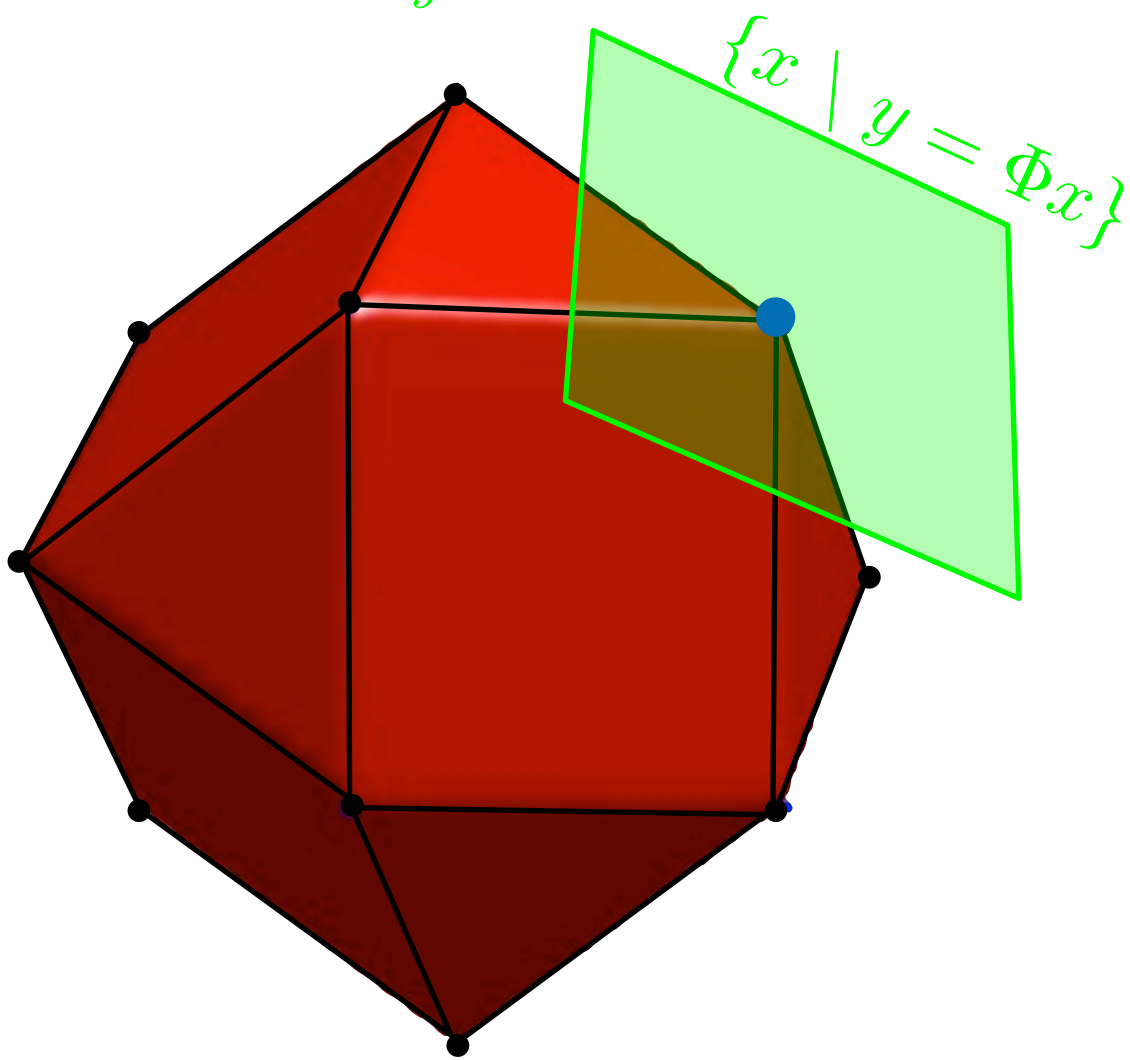
Polytope of TV in 1-D

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

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

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

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



Copyright © Mael

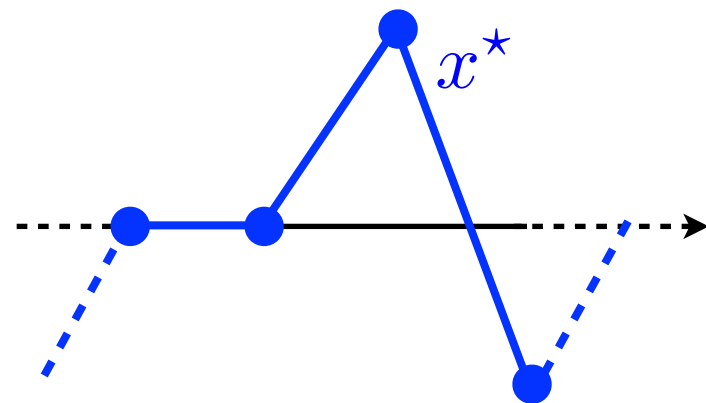
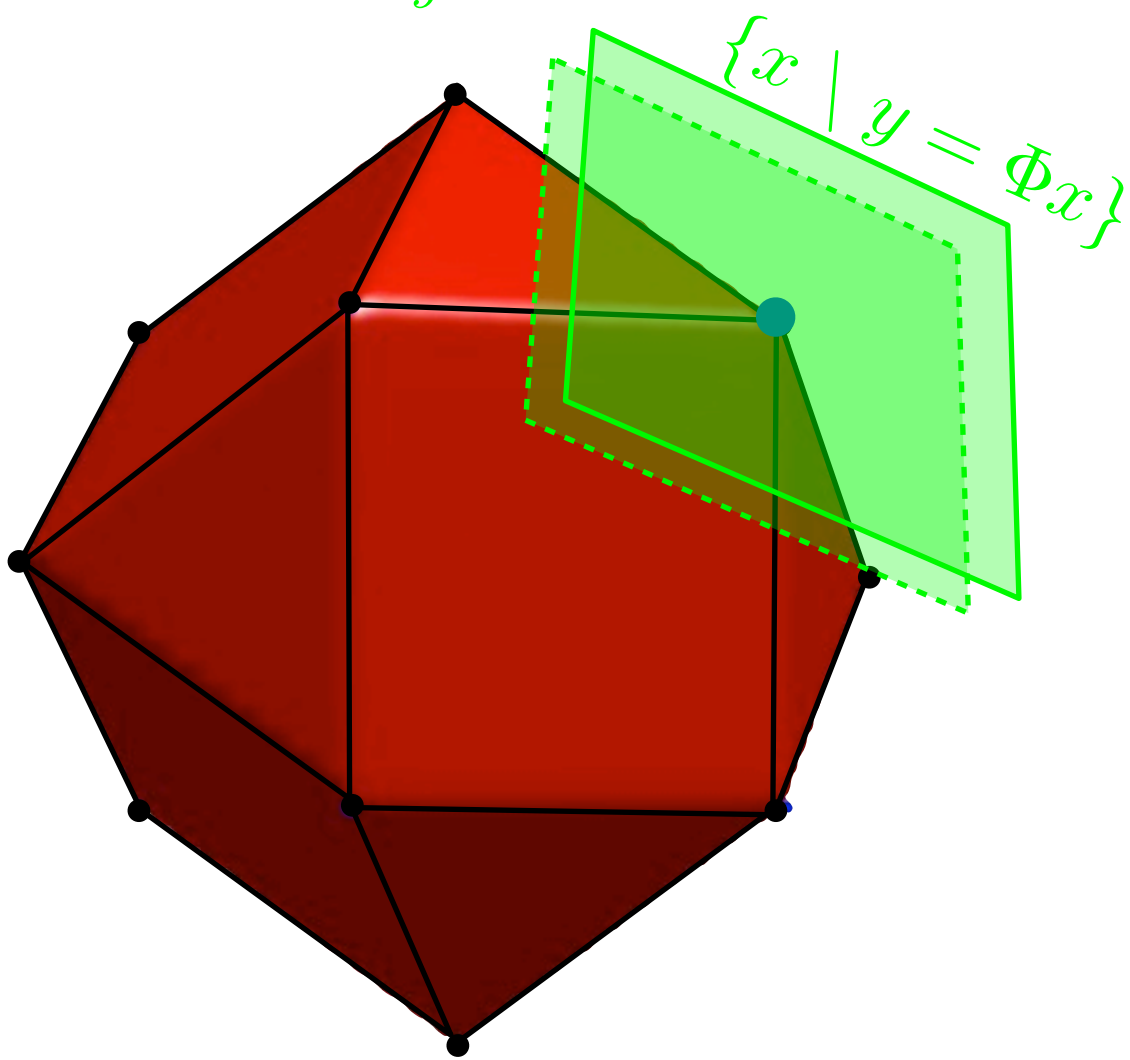
Polytope of TV in 1-D

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

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

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

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



Copyright © Mael

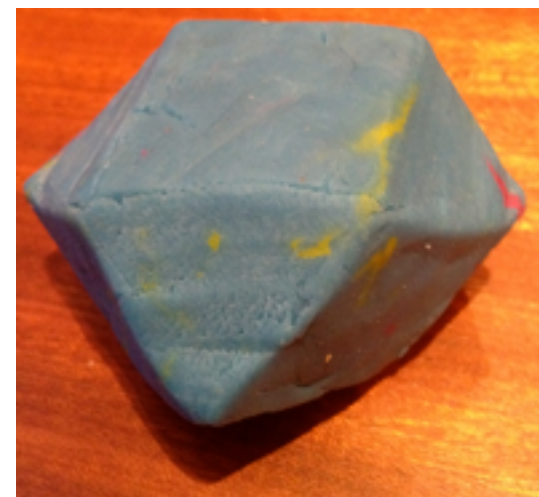
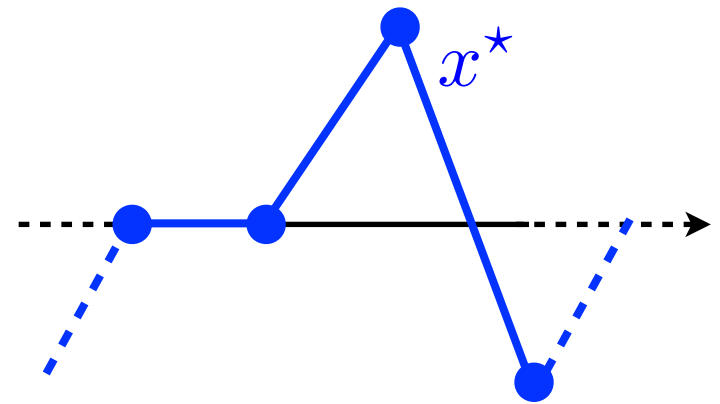
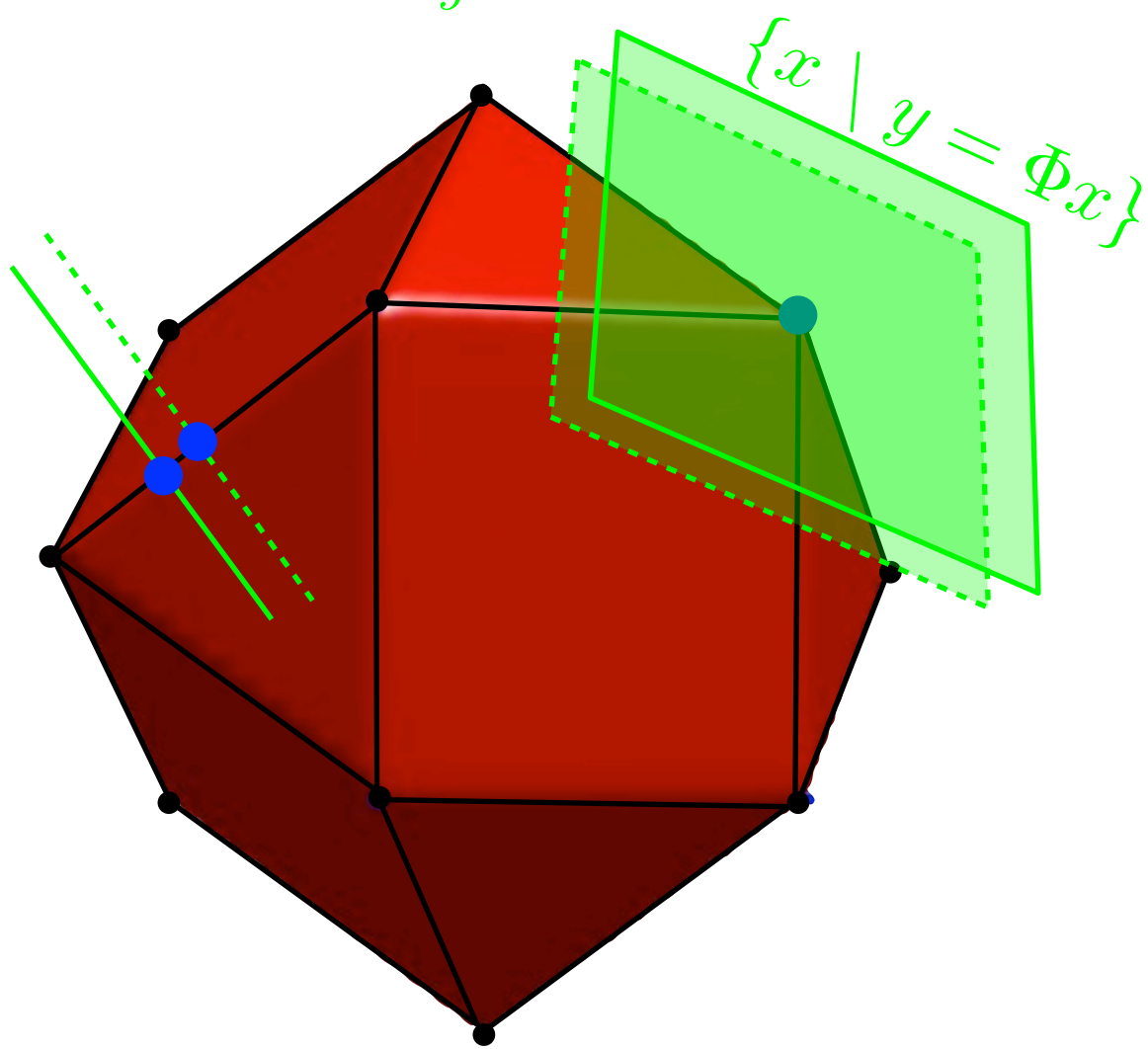
Polytope of TV in 1-D

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

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

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

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



Copyright © Mael

Union of Subspaces Model

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

Support of the solution:

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

$$J = I^c$$

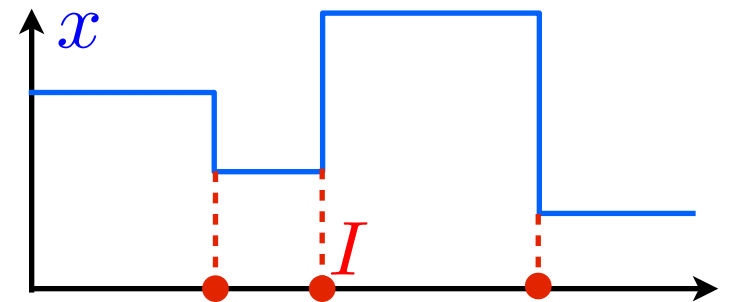
Union of Subspaces Model

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

Support of the solution:

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

$$J = I^c$$



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

I = positions of discontinuities.

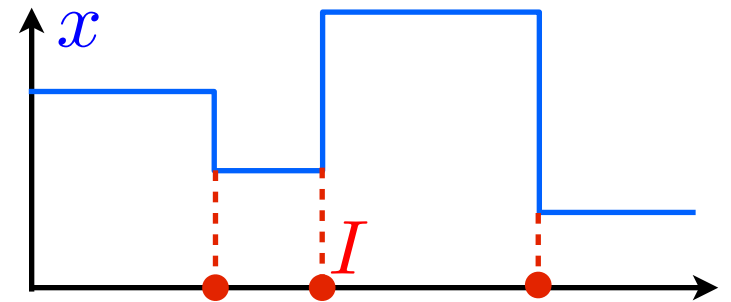
Union of Subspaces Model

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

Support of the solution:

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

$$J = I^c$$



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

I = positions of discontinuities.

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

For TV: \mathcal{G}_J = signals with prescribed discontinuities.

Local Sign Stability

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

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

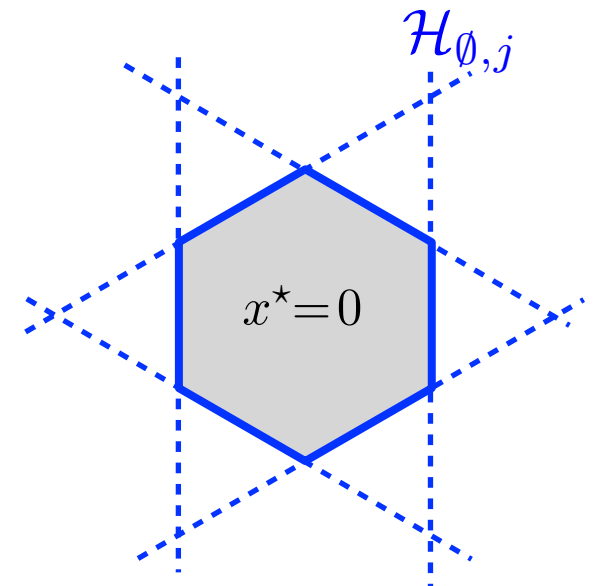
To be understood: there exists a solution with same sign.

Local Sign Stability

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

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

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



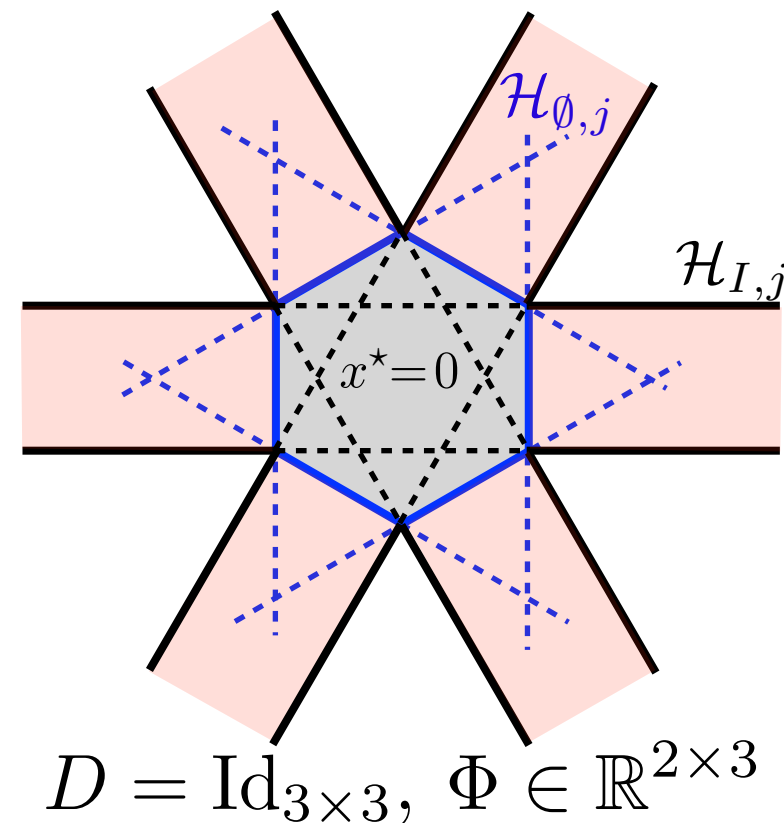
$$D = \operatorname{Id}_{3 \times 3}, \quad \Phi \in \mathbb{R}^{2 \times 3}$$

Local Sign Stability

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

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

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



Local Sign Stability

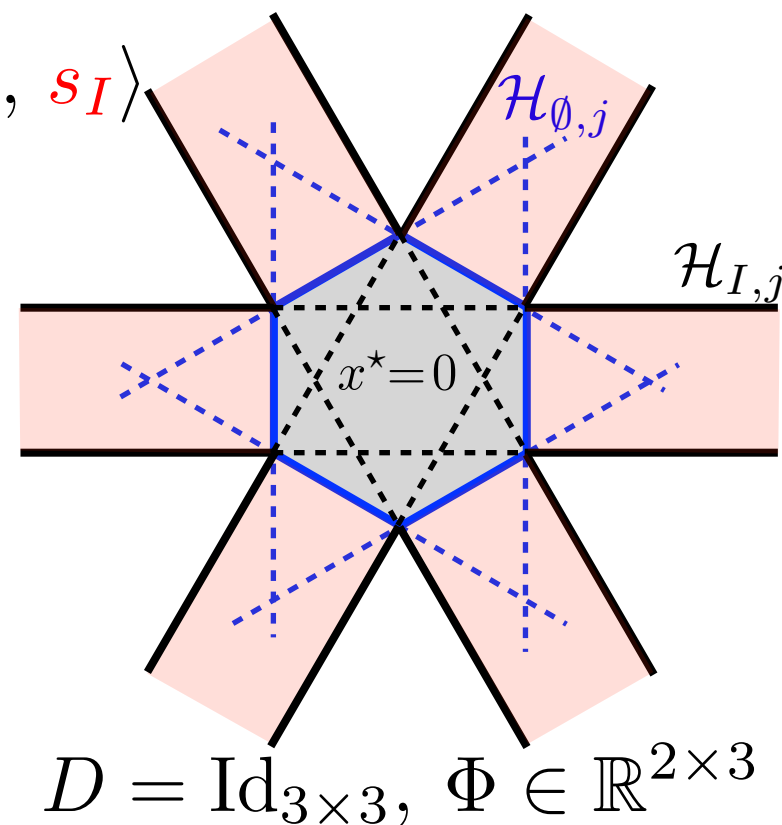
$$x_\lambda(y) \in \operatorname{argmin}_x \frac{1}{2} \|\Phi x - y\|^2 + \lambda \|D^* x\|_1 \quad (\mathcal{P}_\lambda(y))$$

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

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

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

$$\begin{aligned} \hat{x}_\lambda(\bar{y}) &= \operatorname{argmin}_{x \in \mathcal{G}_J} \frac{1}{2} \|\Phi x - \bar{y}\|^2 + \lambda \langle D_I^* x, s_I \rangle \\ &= A^{[J]} (\Phi^* \bar{y} - \lambda D_I s_I) \end{aligned}$$



Local Sign Stability

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

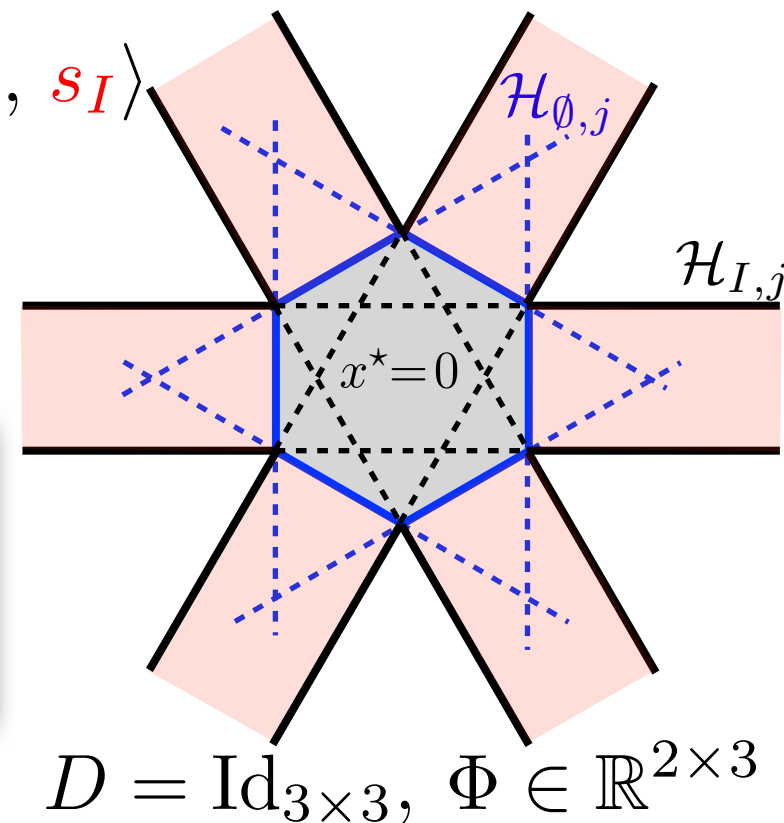
Lemma: $\operatorname{sign}(D^* x_\lambda(y))$ is constant around $y \notin \mathcal{H}$.

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

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

$$\begin{aligned} \hat{x}_\lambda(\bar{y}) &= \operatorname{argmin}_{x \in \mathcal{G}_J} \frac{1}{2} \|\Phi x - \bar{y}\|^2 + \lambda \langle D_I^* x, s_I \rangle \\ &= A^{[J]} (\Phi^* \bar{y} - \lambda D_I s_I) \end{aligned}$$

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



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.

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

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

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

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.

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

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

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

are unbiased estimators of df and gdf.

Trick: $\text{tr}(A) = \mathbb{E}_z(\langle Az, z \rangle)$, $z \sim \mathcal{N}(0, \text{Id}_P)$. [Ramani et al'08]

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.

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

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

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

are unbiased estimators of df and gdf.

Trick: $\text{tr}(A) = \mathbb{E}_z(\langle Az, z \rangle)$, $z \sim \mathcal{N}(0, \text{Id}_P)$. [Ramani et al'08]

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

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

Group Lasso Variations

Partition in blocks: $\{1, \dots, N\} = \bigcup_{b \in \mathcal{B}} b$

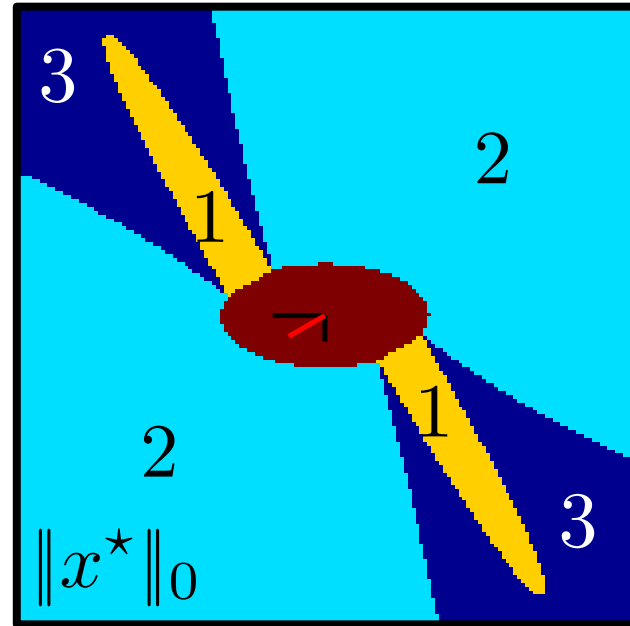
$\ell^1 - \ell^2$ regularization: $J(x) = \sum_{b \in \mathcal{B}} \|x_b\|, \quad x_b = (x_i)_{i \in b}$

Group Lasso Variations

Partition in blocks: $\{1, \dots, N\} = \bigcup_{b \in \mathcal{B}} b$

$\ell^1 - \ell^2$ regularization: $J(x) = \sum_{b \in \mathcal{B}} \|x_b\|$, $x_b = (x_i)_{i \in b}$

Transition set \mathcal{H} : semi-algebraic with zero measure.



$$J(x) = (x_1^2 + x_2^2)^{\frac{1}{2}} + |x_3|$$

Group Lasso Variations

Partition in blocks: $\{1, \dots, N\} = \bigcup_{b \in \mathcal{B}} b$

$\ell^1 - \ell^2$ regularization: $J(x) = \sum_{b \in \mathcal{B}} \|x_b\|$, $x_b = (x_i)_{i \in b}$

Transition set \mathcal{H} : semi-algebraic with zero measure.

Theorem:

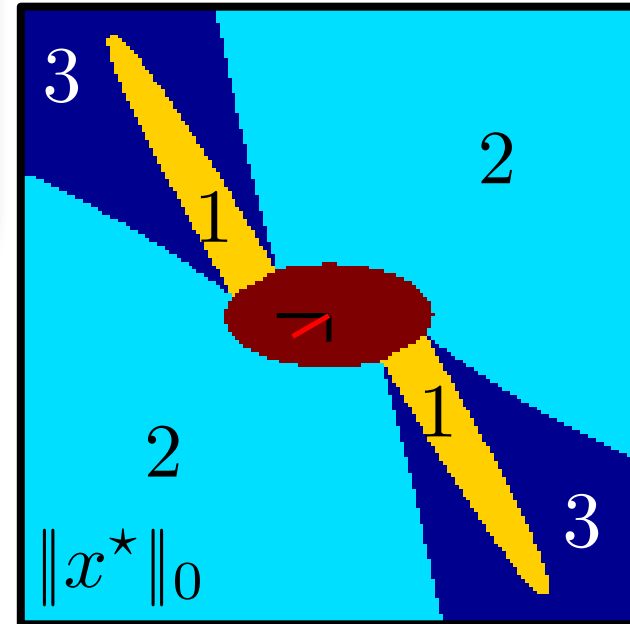
$$\forall y \notin \mathcal{H}, \partial \mu_I(y) = \Phi_I \Psi_I(x_\lambda(y))^{-1} \Phi_I^*$$

where:

$$\Psi(x_\lambda(y)) = \Phi_I^* \Phi_I + \lambda \delta_x \circ P_x$$

$$\delta_x = \text{diag}(\text{Id}_b / \|x_b\|)_{b \in I}$$

$$P_x = \text{Id}_I - \text{diag}(x_b x_b^*)_{b \in I}$$

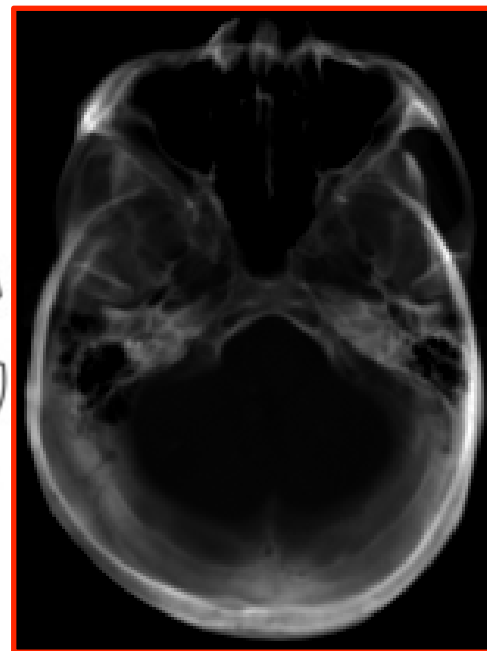
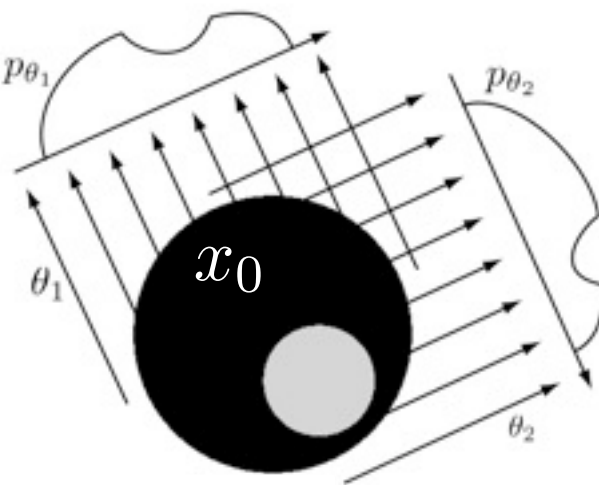
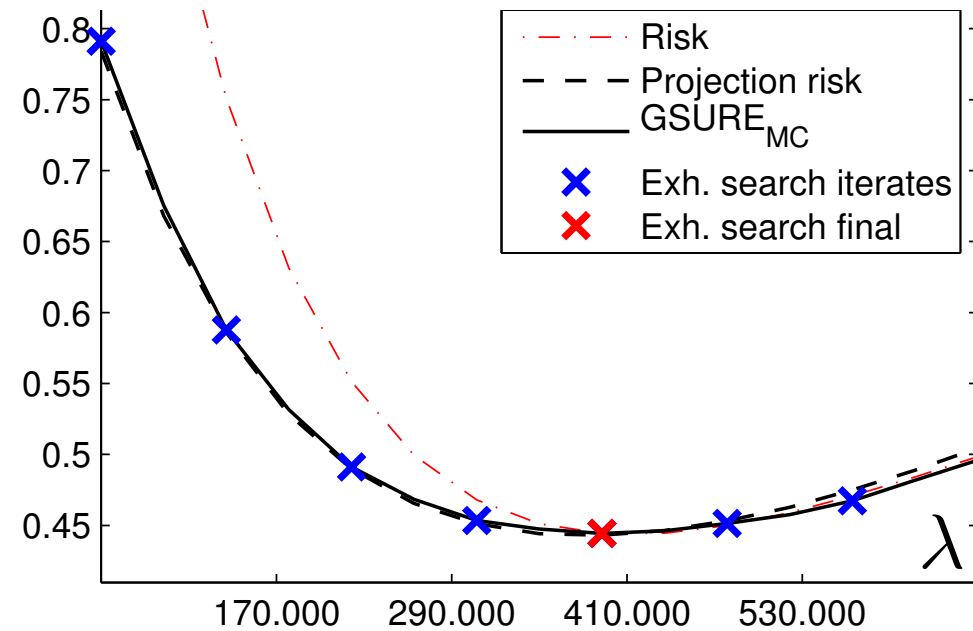


$$J(x) = (x_1^2 + x_2^2)^{\frac{1}{2}} + |x_3|$$

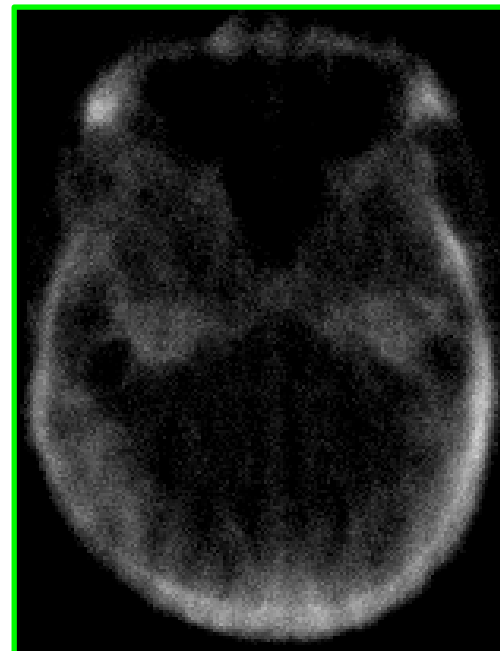
CT Scan (Tomography)

Φ : sub-sampled
radon transform.
#projections=16.

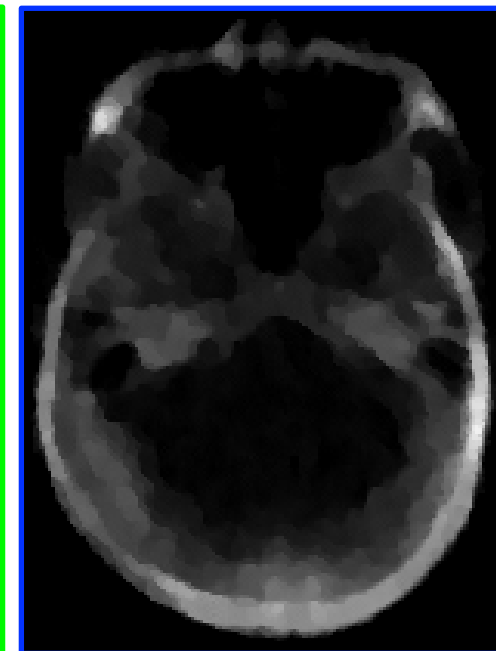
Finite differences gradient:
 $D = [\partial_1, \partial_2]$



x_0

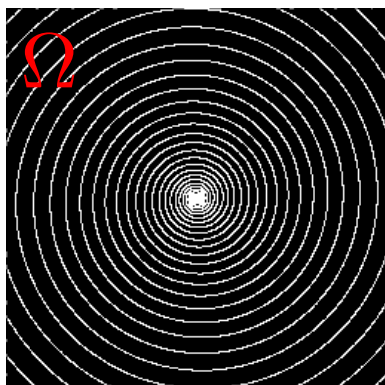
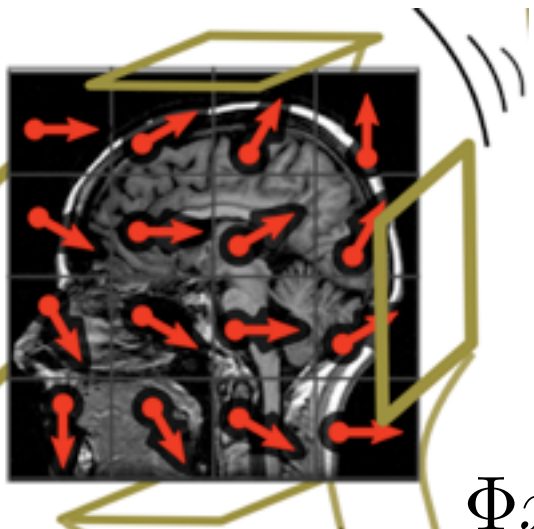


$\Phi^+ y$

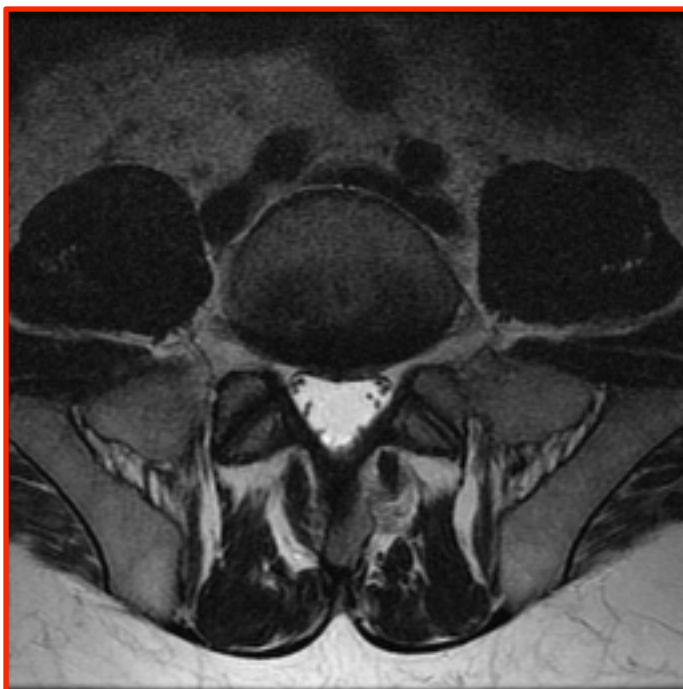
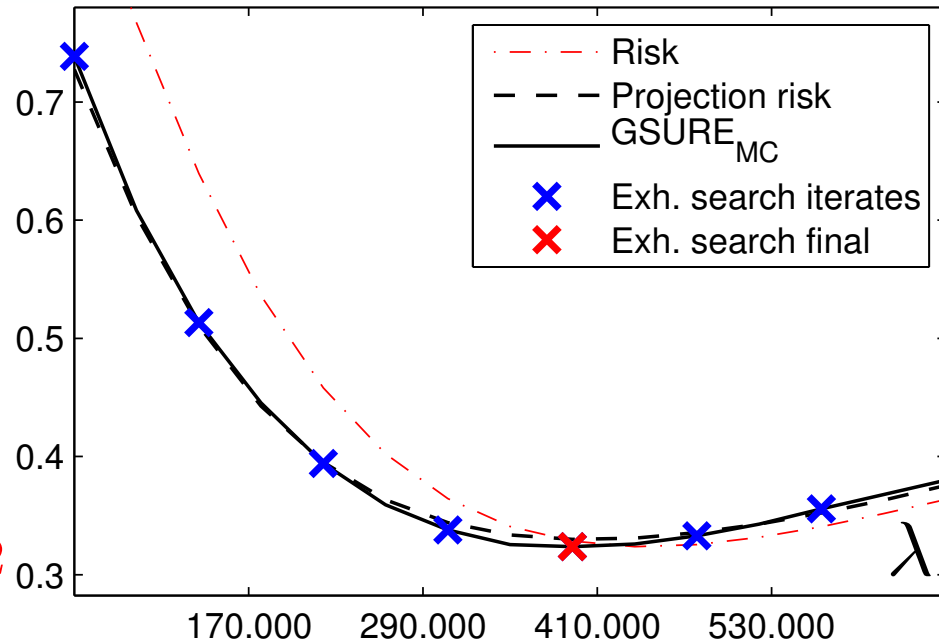


$x_{\lambda^*}(y)$

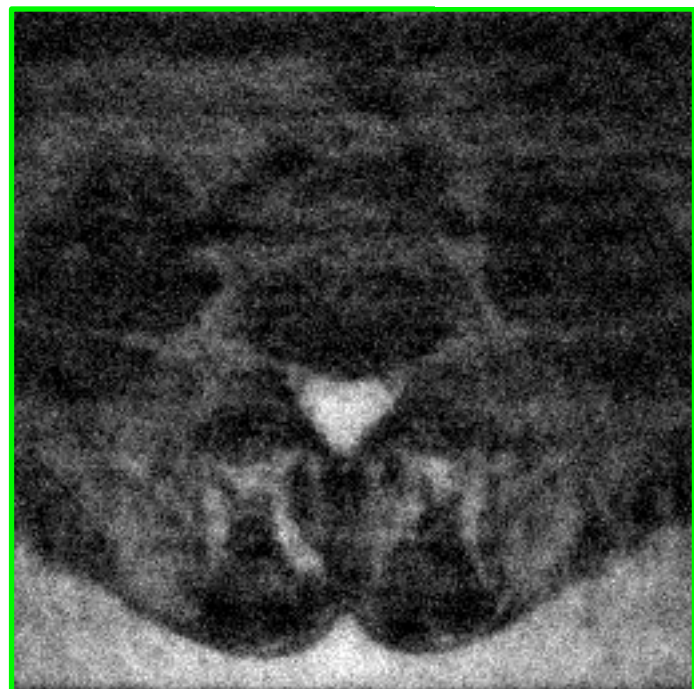
MRI (Fourier Sampling)



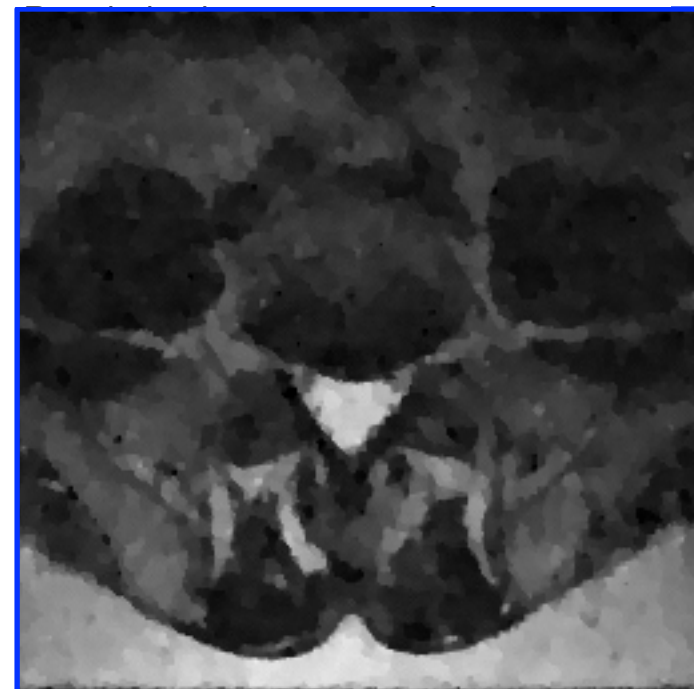
$$\Phi x_0 = (\hat{x}_0(\omega))_{\omega \in \Omega}$$



x_0



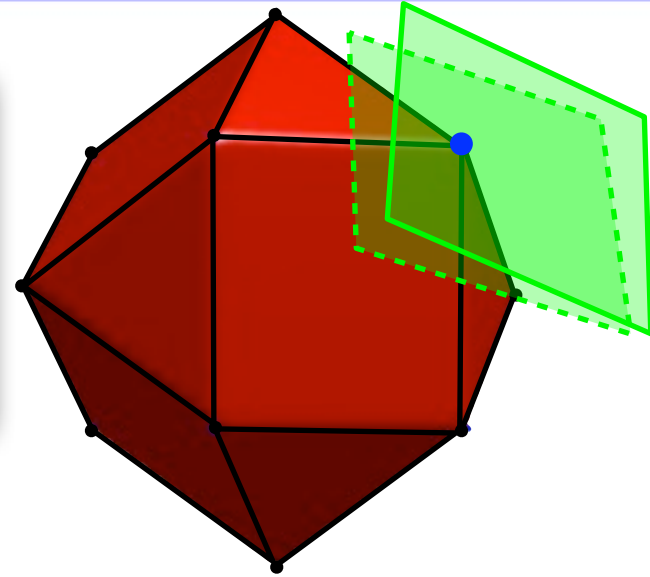
$\Phi^+ y$



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

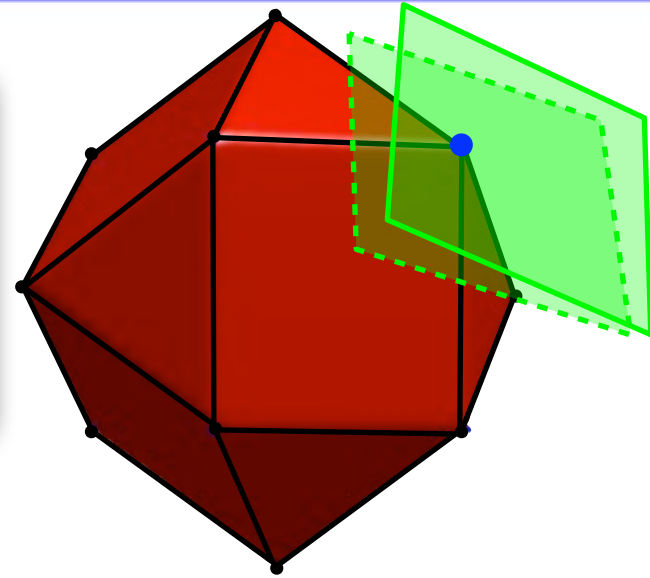
Conclusion

Computing risk estimates
 \iff
Sensitivity of the estimator



Conclusion

Computing risk estimates
 \iff
Sensitivity of the estimator

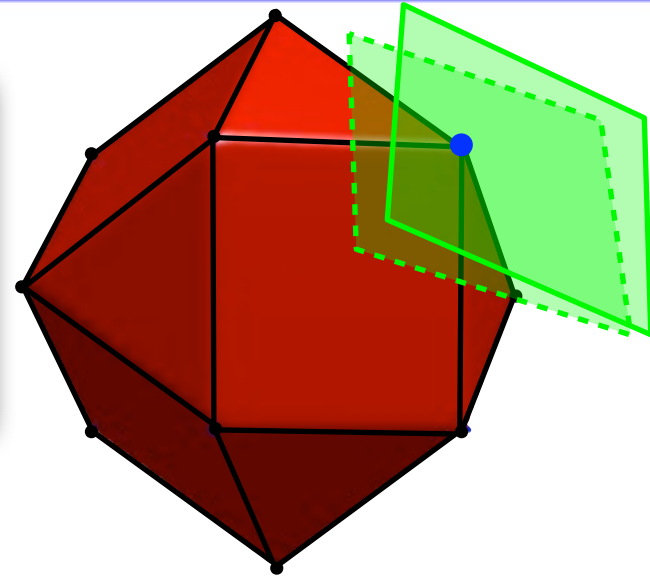


Closed form expression:

\rightarrow quadratic, ℓ^1 regularization, ...

Conclusion

Computing risk estimates
 \iff
Sensitivity of the estimator



Closed form expression:

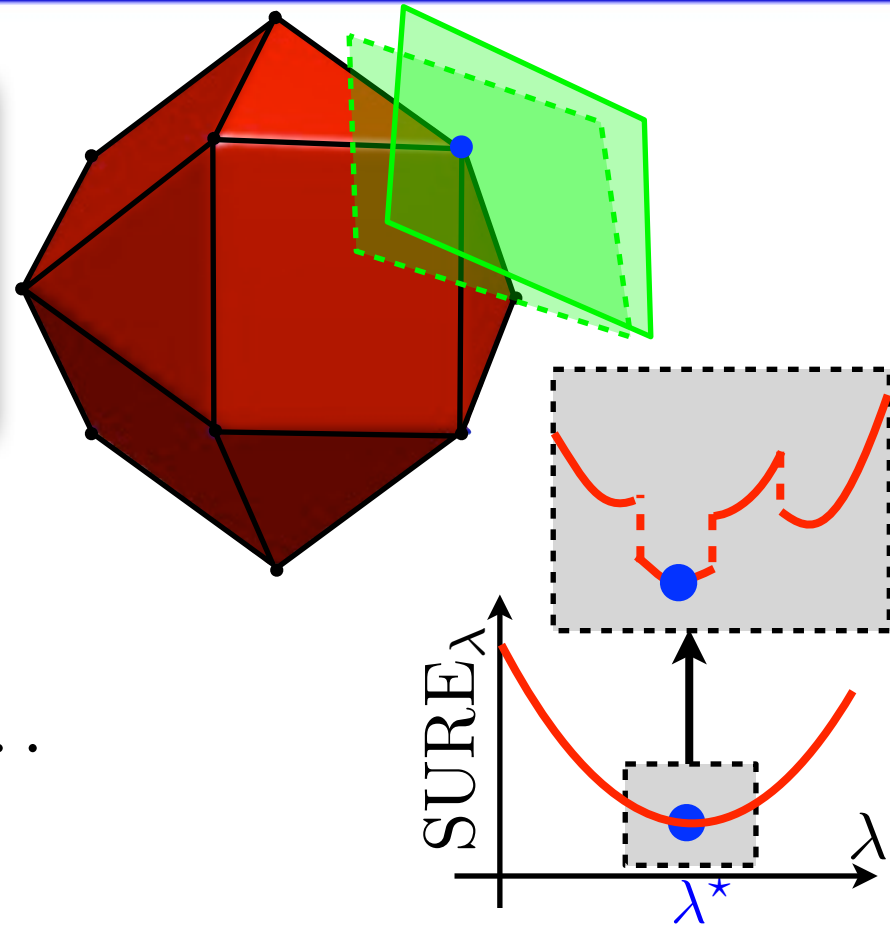
\rightarrow quadratic, ℓ^1 regularization, ...

Open problems:

- Estimate the risk in $\ker(\Phi)$.
- Fast algorithms to optimize high dimensional λ .

Conclusion

Computing risk estimates
 \iff
Sensitivity of the estimator



Closed form expression:

\rightarrow quadratic, ℓ^1 regularization, ...

Open problems:

- Estimate the risk in $\ker(\Phi)$.
- Fast algorithms to optimize high dimensional λ .