Sparse Regularization of Inverse Problems

Gabriel Peyré

http://www.ceremade.dauphine.fr/~peyre/numerical-tour/

Overview

- Inverse Problems Regularization
- Sparse Prior in Orthogonal Bases
- Sparse Regularization in Orthogonal Bases
- Examples of applications:
  - *Sparse Diracs*: Seismic Deconvolution
  - *Sparse in 1D Wavelets*: Deblurring
  - *Sparse in 2D Wavelets*: Inpainting
- Redundant Representations
- Sparse Regularization in Redundant Dictionaries

Inverse Problems

Recovering f_0 from P noisy measurements $y = \Phi f_0 + w \in \mathbb{R}^P$.

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

$w[n] \sim \mathcal{N}(0, \sigma)$ white noise.

Inverse Problems

Recovering f_0 from P noisy measurements $y = \Phi f_0 + w \in \mathbb{R}^P$.

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

$w[n] \sim \mathcal{N}(0, \sigma)$ white noise.

Denoising: $\Phi = \mathrm{Id}_N$, $P = N$.

Inverse Problems

Recovering f_0 from P noisy measurements $y = \Phi f_0 + w \in \mathbb{R}^P$.

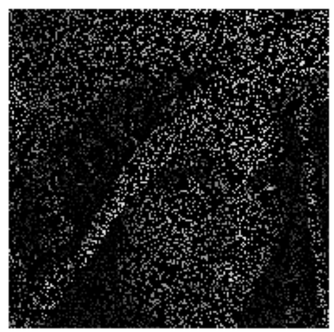
$\Phi : \mathbb{R}^N \mapsto \mathbb{R}^P$ with $P \ll N$ (missing information)

$w[n] \sim \mathcal{N}(0, \sigma)$ white noise.

Denoising: $\Phi = \mathrm{Id}_N$, $P = N$.

Inpainting: set $\Omega \subset \{0, \ldots, N-1\}$ of missing pixels, $P = N - |\Omega|$.

$\xrightarrow{\Phi}$



$$(\Phi f)(x) = \begin{cases} 0 & \text{if} \quad x \in \Omega, \\ f(x) & \text{if} \quad x \notin \Omega. \end{cases}$$

Inverse Problems

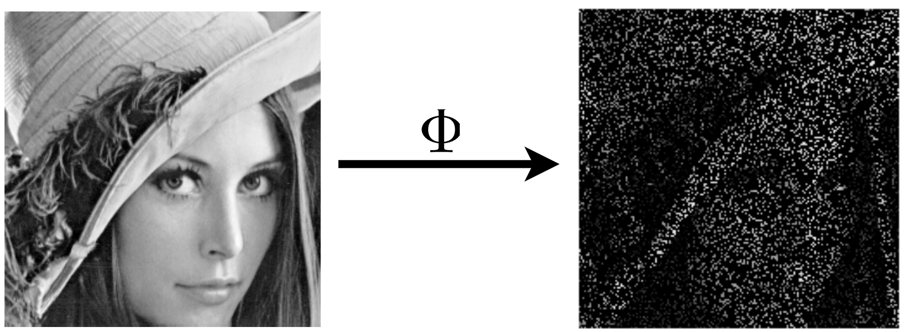
Recovering f_0 from P noisy measurements $y = \Phi f_0 + w \in \mathbb{R}^P$.

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

$w[n] \sim \mathcal{N}(0, \sigma)$ white noise.

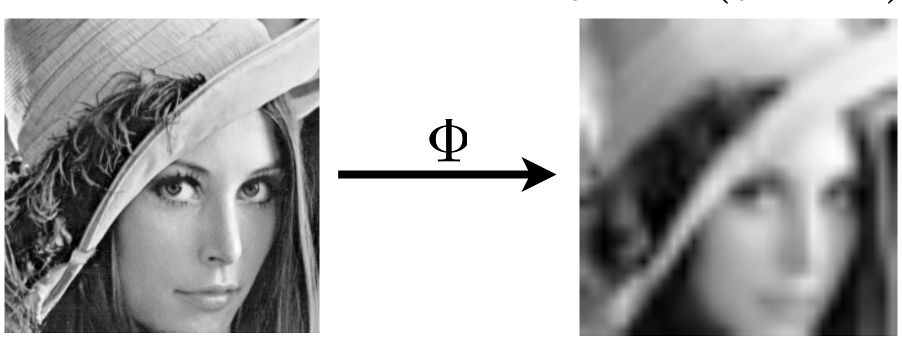
Denoising: $\Phi = \mathrm{Id}_N$, $P = N$.

Inpainting: set $\Omega \subset \{0, \ldots, N-1\}$ of missing pixels, $P = N - |\Omega|$.



$$(\Phi f)(x) = \begin{cases} 0 & \text{if} \quad x \in \Omega, \\ f(x) & \text{if} \quad x \notin \Omega. \end{cases}$$

Super-resolution: $\Phi f = (f \star h) \downarrow_k$, $P = N/k$.



Inverse Problems

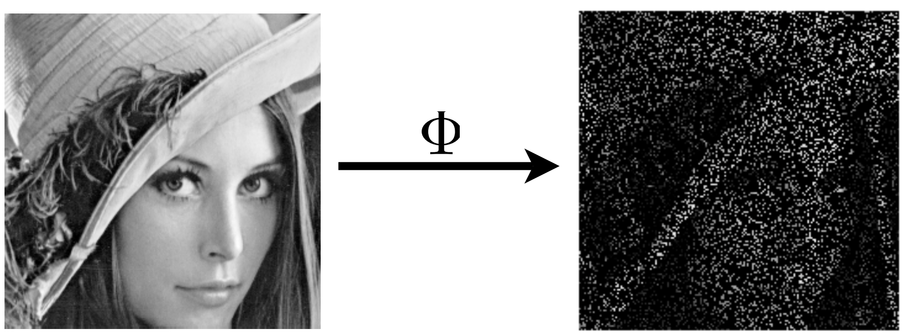
Recovering f_0 from P noisy measurements $y = \Phi f_0 + w \in \mathbb{R}^P$.

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

$w[n] \sim \mathcal{N}(0, \sigma)$ white noise.

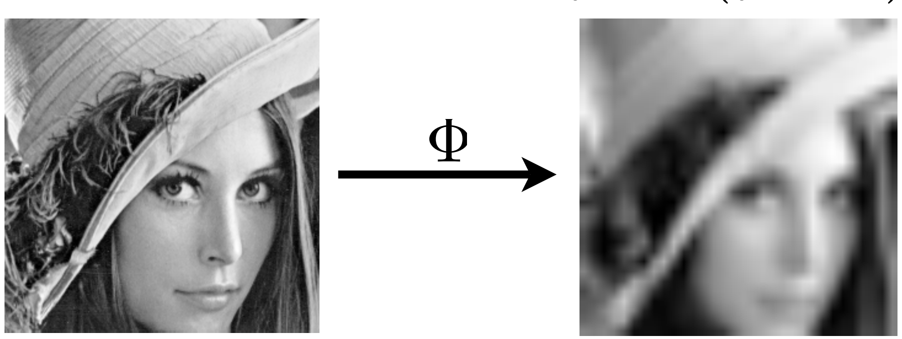
Denoising: $\Phi = \mathrm{Id}_N$, $P = N$.

Inpainting: set $\Omega \subset \{0, \ldots, N-1\}$ of missing pixels, $P = N - |\Omega|$.

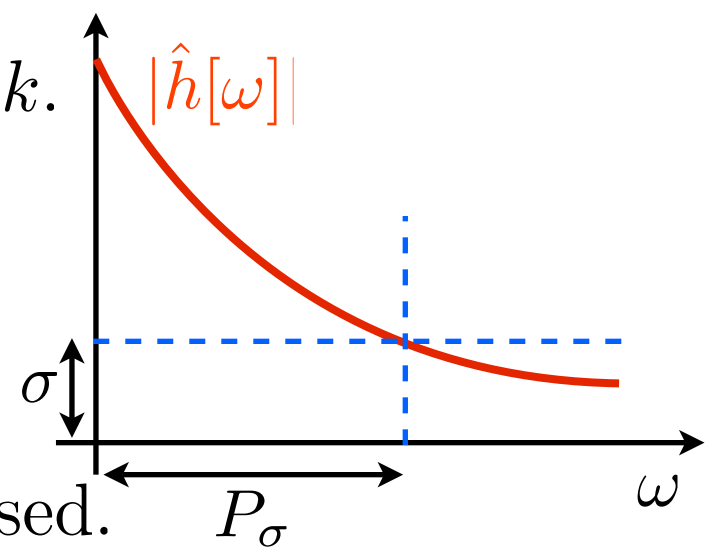


$$(\Phi f)(x) = \begin{cases} 0 & \text{if} \quad x \in \Omega, \\ f(x) & \text{if} \quad x \notin \Omega. \end{cases}$$

Super-resolution: $\Phi f = (f \star h) \downarrow_k$, $P = N/k$.



De-blurring: $\Phi f = f \star h$, $P = N$ but ill-posed.



Inverse Problem Regularization

Noisy measurements $y = \Phi f_0 + w \in \mathbb{R}^P$, $\quad w[n] \sim \mathcal{N}(0, \sigma)$.

Prior model: $J(f) \in \mathbb{R}$ such that $J(f_0)$ is small for $f_0 \in \Theta$.

Inverse Problem Regularization

Noisy measurements $y = \Phi f_0 + w \in \mathbb{R}^P$, $\quad w[n] \sim \mathcal{N}(0, \sigma)$.

Prior model: $J(f) \in \mathbb{R}$ such that $J(f_0)$ is small for $f_0 \in \Theta$.

Regularized inverse: $$f^\star = \underset{f \in \mathbb{R}^N}{\text{argmin}} \; \underbrace{\boxed{\frac{1}{2}\|y - \Phi f\|^2}}_{\text{Data fitting}} + \underbrace{\boxed{\lambda J(f)}}_{\text{Regularity}}$$

Bayesian interpretation: $p(f) \propto e^{-\tau J(f)}$, $\lambda = \tau\sigma^2$, MAP estimation.

Inverse Problem Regularization

Noisy measurements $y = \Phi f_0 + w \in \mathbb{R}^P, \quad w[n] \sim \mathcal{N}(0,\sigma)$.

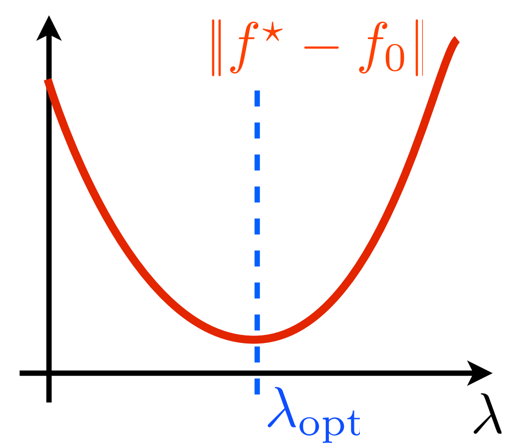
Prior model: $J(f) \in \mathbb{R}$ such that $J(f_0)$ is small for $f_0 \in \Theta$.

Regularized inverse:
$$f^\star = \underset{f \in \mathbb{R}^N}{\text{argmin}} \; \underbrace{\boxed{\frac{1}{2}\|y - \Phi f\|^2}}_{\text{Data fitting}} + \underbrace{\boxed{\lambda J(f)}}_{\text{Regularity}}$$

Bayesian interpretation: $p(f) \propto e^{-\tau J(f)}$, $\lambda = \tau\sigma^2$, MAP estimation.

Trade-off between denoising (λ increases with σ) and regularity of f_0.

Choice of λ: minimize $\|f^\star - f_0\|$ (oracle)



Inverse Problem Regularization

Noisy measurements $y = \Phi f_0 + w \in \mathbb{R}^P$, $\quad w[n] \sim \mathcal{N}(0, \sigma)$.

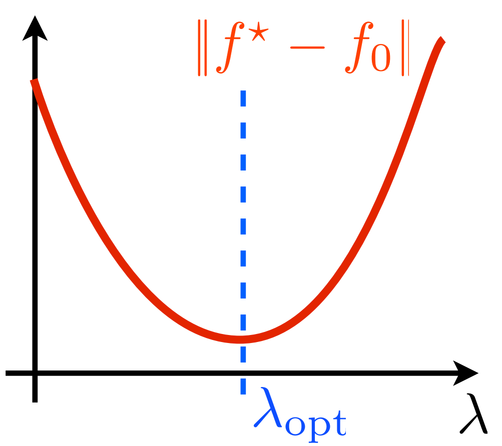
Prior model: $J(f) \in \mathbb{R}$ such that $J(f_0)$ is small for $f_0 \in \Theta$.

Regularized inverse: $$f^\star = \underset{f \in \mathbb{R}^N}{\text{argmin}} \; \underbrace{\boxed{\frac{1}{2}\|y - \Phi f\|^2}}_{\text{Data fitting}} + \underbrace{\boxed{\lambda J(f)}}_{\text{Regularity}}$$

Bayesian interpretation: $p(f) \propto e^{-\tau J(f)}$, $\lambda = \tau \sigma^2$, MAP estimation.

Trade-off between denoising (λ increases with σ) and regularity of f_0.

Choice of λ: minimize $\|f^\star - f_0\|$ (oracle)



No noise: $\sigma = 0$, $\lambda \to 0$, minimize

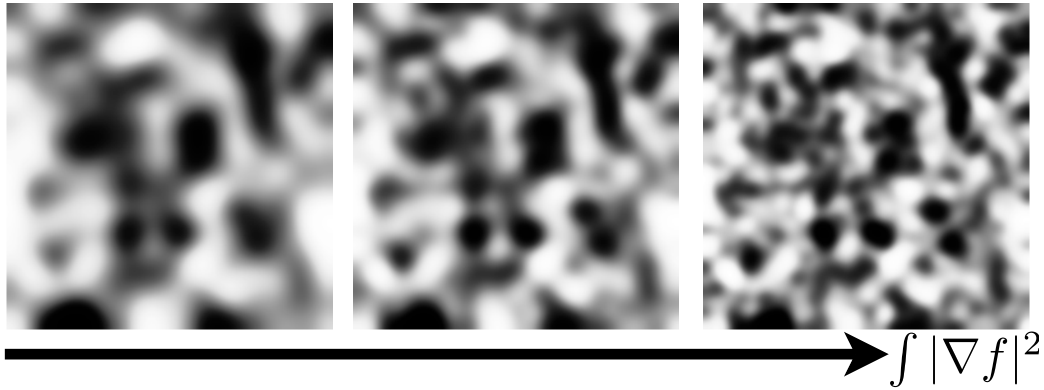
$$f^\star = \underset{f \in \mathbb{R}^N, \Phi f = y}{\text{argmin}} \; J(f)$$

Smooth and Cartoon Priors

Prior model: energy $J(f) \in \mathbb{R}$ low for images of the model $f \in \Theta$.

Sobolev pseudo-norm: $$J(f) = \frac{1}{2}\|f\|^2_{\text{Sob}} = \frac{1}{2}\int \|\nabla_x f\|^2 \mathrm{d}x$$

(norm: $\mu\|f\| + \|f\|_{\text{Sob}}$)



Smooth and Cartoon Priors

Prior model: energy $J(f) \in \mathbb{R}$ low for images of the model $f \in \Theta$.

Sobolev pseudo-norm: $$J(f) = \frac{1}{2}\|f\|^2_{\mathrm{Sob}} = \frac{1}{2}\int \|\nabla_x f\|^2 \mathrm{d}x$$

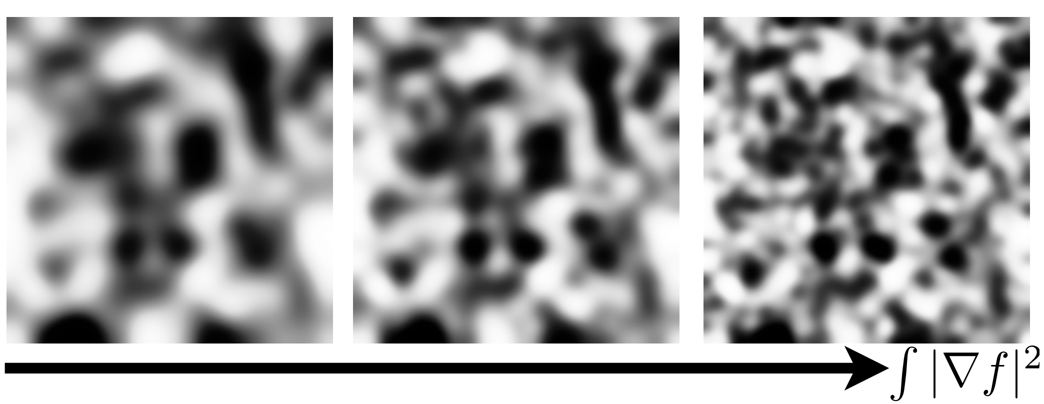
(norm: $\mu\|f\| + \|f\|_{\mathrm{Sob}}$)

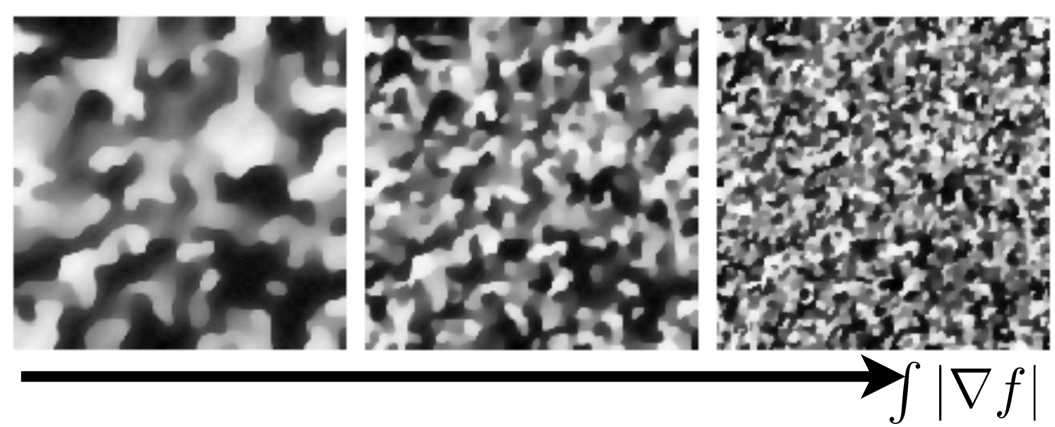
Total variation pseudo-norm: $$J(f) = \|f\|_{\mathrm{TV}} = \int \|\nabla_x f\| \mathrm{d}x$$

$\longrightarrow$ Extension to non-smooth functions $f \in \mathrm{BV}([0,1]^2)$

Co-area formula: $$\|f\|_{\mathrm{TV}} = \int_{\mathbb{R}} \mathrm{length}(\mathcal{C}_t)\mathrm{d}t$$

Level set $\mathcal{C}_t = \{x \setminus f(x) = t\}$





Sobolev and TV Regularization

Sobolev prior: $J(f) = \frac{1}{2}\|\nabla f\|^2$.

Sobolev regularization: $$f^\star = \underset{f \in \mathbb{R}^N}{\operatorname{argmin}} \; \|y - \Phi f\|^2 \; + \; \lambda \|\nabla f\|^2$$

$$f^\star = (\Phi^* \Phi - \lambda \Delta)^{-1} \Phi^* y$$

TV regularization: $$f^\star = \underset{f \in \mathbb{R}^N}{\operatorname{argmin}} \; \frac{1}{2}\|\Phi f - y\|^2 + \lambda J^\varepsilon(f)$$

"Smoothed" TV-norm: $$J^\varepsilon(f) = \sum_n \sqrt{\varepsilon^2 + \|\nabla f[n]\|^2}$$

Sobolev and TV Regularization

Sobolev prior: $J(f) = \frac{1}{2}\|\nabla f\|^2$.

Sobolev regularization: $$f^\star = \underset{f \in \mathbb{R}^N}{\text{argmin}} \; \|y - \Phi f\|^2 \; + \; \lambda \|\nabla f\|^2$$

$$f^\star = (\Phi^* \Phi - \lambda \Delta)^{-1} \Phi^* y$$

TV regularization: $$f^\star = \underset{f \in \mathbb{R}^N}{\text{argmin}} \; \frac{1}{2}\|\Phi f - y\|^2 + \lambda J^\varepsilon(f)$$

"Smoothed" TV-norm: $$J^\varepsilon(f) = \sum_n \sqrt{\varepsilon^2 + \|\nabla f[n]\|^2}$$

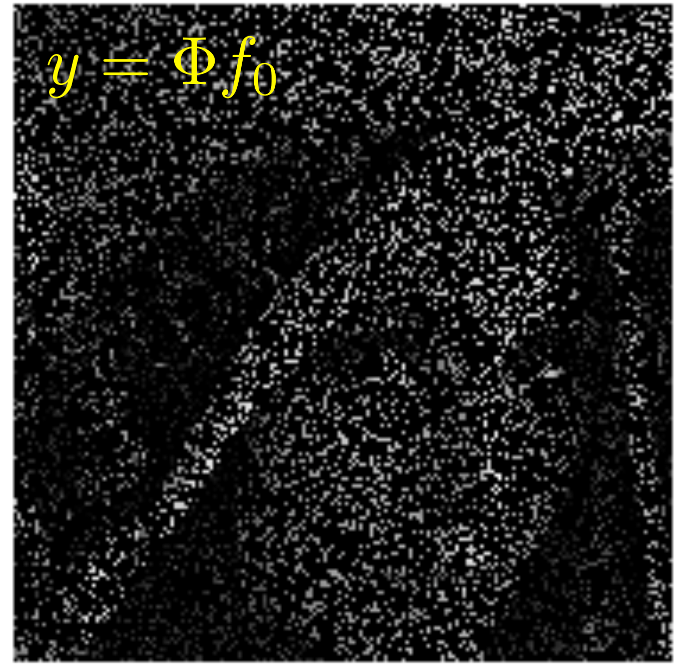
Numerical computation: gradient descent:

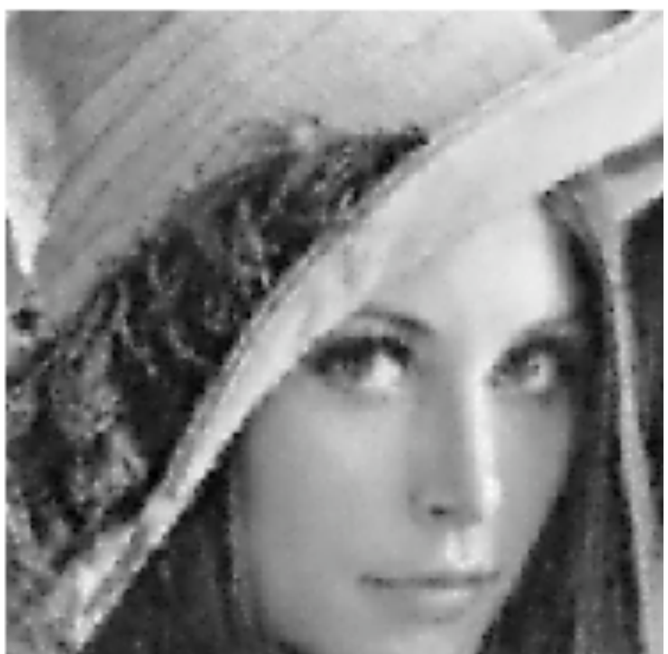
$$f^{(k+1)} = f^{(k)} - \tau \left[\Phi^*(\Phi f^{(k)} - y) + \text{grad}_{f^{(k)}} J \right]$$

Sobolev $\longrightarrow$ $$f^{(k+1)} = f^{(k)} - \tau \left[\Phi^*(\Phi f^{(k)} - y) - \lambda \Delta f^{(k)} \right]$$

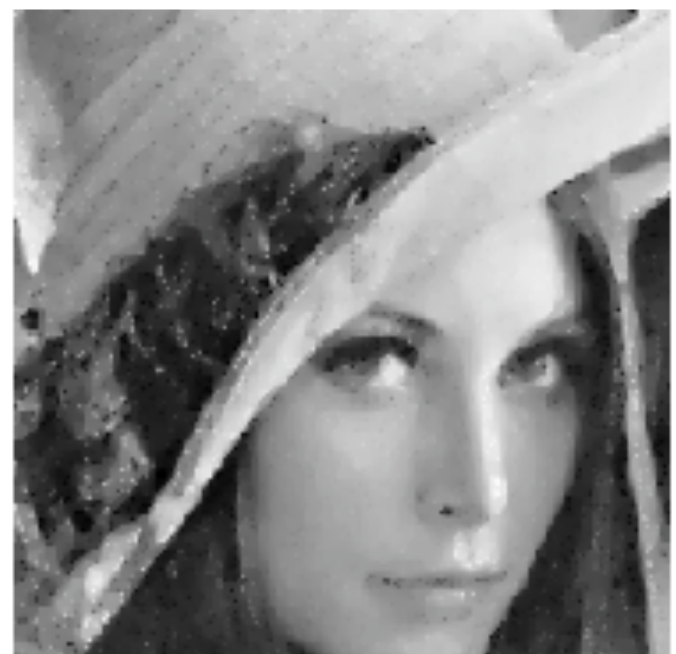
TV $\longrightarrow$ $$f^{(k+1)} = f^{(k)} - \tau \left[\Phi^*(\Phi f^{(k)} - y) - \lambda \text{div}\left(\frac{\nabla f^{(k)}}{\sqrt{\varepsilon^2 + \|\nabla f^{(k)}\|^2}} \right) \right]$$

Inpainting Results





Sobolev, SNR=19.8dB



TV, SNR=19dB

Overview

- Inverse Problems Regularization
- **Sparse Prior in Orthogonal Bases**
- Sparse Regularization in Orthogonal Bases
- Examples of applications:
  - *Sparse Diracs*: Seismic Deconvolution
  - *Sparse in 1D Wavelets*: Deblurring
  - *Sparse in 2D Wavelets*: Inpainting
- Redundant Representations
- Sparse Regularization in Redundant Dictionaries

Sparse Priors

Orthogonal basis $\mathcal{B} = \{\psi_m\}_m$ of $\mathbb{R}^N$.

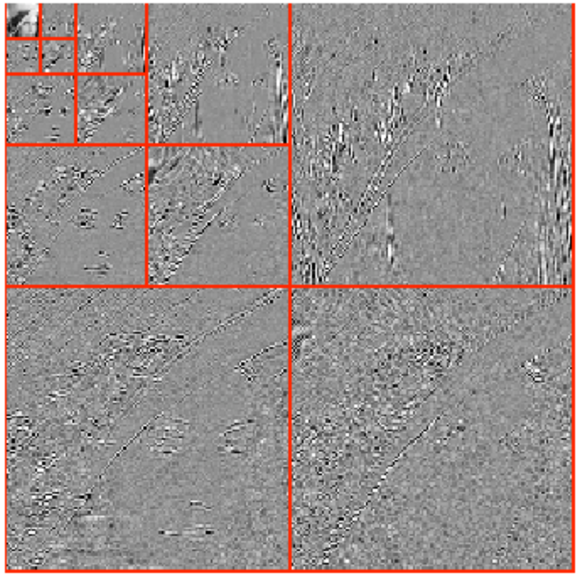
Example: Wavelet basis $\psi_m = \psi_{j,n}$, $m = (j, n)$.

Sparsity: most $\langle f,\, \psi_m \rangle$ are small.

Ideal sparsity: for most m, $\langle f,\, \psi_m \rangle = 0$.

Ideal prior: $J_0(f) = \#\{m \setminus \langle f,\, \psi_m \rangle \neq 0\}$

Image f



Coefficients $\{\langle f,\, \psi_m \rangle\}_m$

Sparse Priors

Orthogonal basis $\mathcal{B} = \{\psi_m\}_m$ of $\mathbb{R}^N$.

Example: Wavelet basis $\psi_m = \psi_{j,n}$, $m = (j, n)$.

Sparsity: most $\langle f,\, \psi_m \rangle$ are small.

Ideal sparsity: for most m, $\langle f,\, \psi_m \rangle = 0$.

Ideal prior: $J_0(f) = \#\left\{m \setminus \langle f,\, \psi_m \rangle \neq 0\right\}$

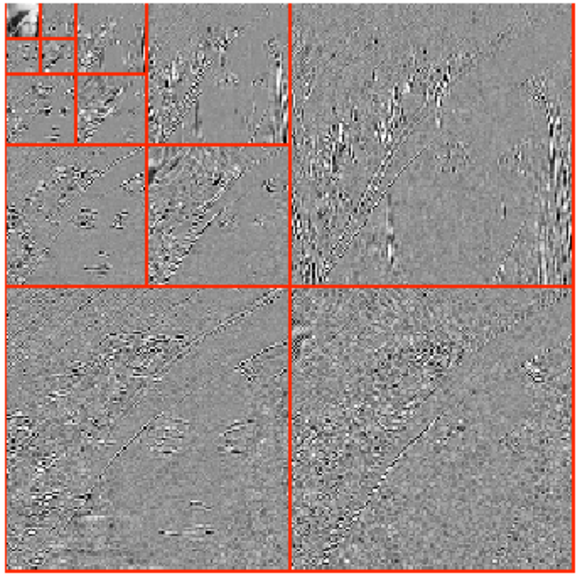
Best M-sparse approximation:

$$f_M = \sum_{|\langle f,\, \psi_m \rangle| > T} \langle f,\, \psi_m \rangle \psi_m.$$

$$M = J_0(f_M) = \#\left\{m \setminus |\langle f,\, \psi_m \rangle| > T\right\}$$

Approximate sparsity: $\|f - f_M\|$ is small.

Image f



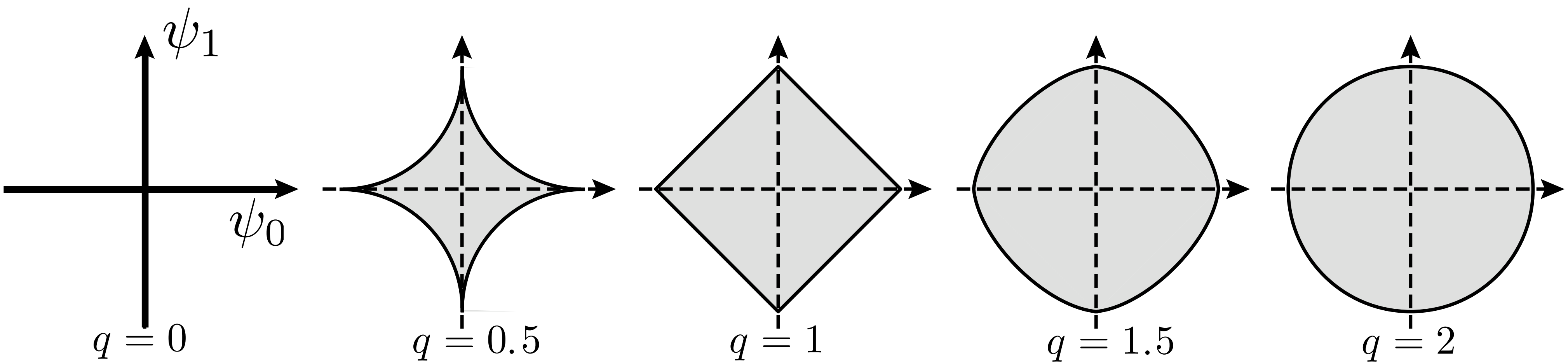
Coefficients $\{\langle f,\, \psi_m \rangle\}_m$

Convex Relaxation: L1 Prior

"Ideal" sparsity prior: $J_0(f) = \|(\langle f,\, \psi_m\rangle)_m\|_0 = \#\left\{m \setminus \langle f,\, \psi_m\rangle \neq 0\right\}$

ℓ^q priors: $J_q(f) = \sum_m |\langle f,\, \psi_m\rangle|^q$. convex (norms) for $q \geqslant 1$.

Bayesian interpretation: $\langle f,\, \psi_m\rangle$ i.i.d. $\sim$ generalized Gaussian $p(x) \propto e^{-\tau|x|^q}$

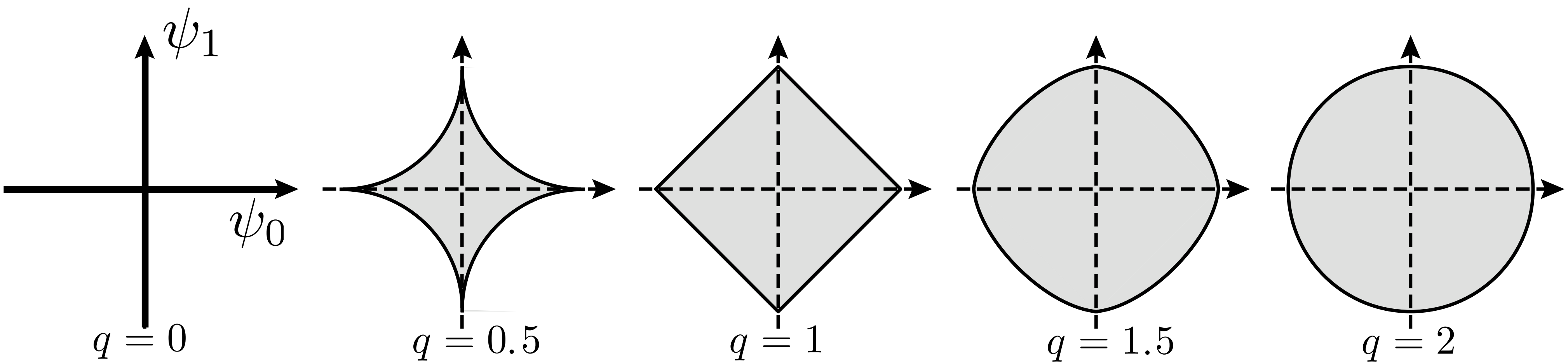


Convex Relaxation: L1 Prior

"Ideal" sparsity prior: $J_0(f) = \|(\langle f,\, \psi_m\rangle)_m\|_0 = \#\,\{m \setminus \langle f,\, \psi_m\rangle \neq 0\}$

ℓ^q priors: $J_q(f) = \sum_m |\langle f,\, \psi_m\rangle|^q.$ convex (norms) for $q \geqslant 1.$

Bayesian interpretation: $\langle f,\, \psi_m\rangle$ i.i.d. $\sim$ generalized Gaussian $p(x) \propto e^{-\tau|x|^q}$



ℓ^1 norm: ℓ^q norm the "closest" to the ℓ^0 ideal sparsity.

Sparse ℓ^1 prior: $\{\psi_m\}_m$ orthogonal basis of $\mathbb{R}^N$

$$J_1(f) = \|(\langle f,\, \psi_m\rangle)\|_1 = \sum_m |\langle f,\, \psi_m\rangle|$$

Sparse Regularization Denoising

Noisy image: $f = f_0 + w$, $\quad w[n] \sim \mathcal{N}(0,\sigma)$ white noise.

Regularization denoising: $\quad f^\star_\lambda = \underset{g \in \mathbb{R}^N}{\text{argmin}} \frac{1}{2}\|f - g\|^2 + \lambda J(g)$

ℓ^0 regularization: $f^\star_{\lambda,0} = \underset{g \in \mathbb{R}^N}{\text{argmin}} \frac{1}{2}\|f - g\|^2 + \lambda \# \left\{ m \setminus \langle g,\, \psi_m \rangle \neq 0 \right\}$

ℓ^1 regularization: $f^\star_{\lambda,1} = \underset{g \in \mathbb{R}^N}{\text{argmin}} \frac{1}{2}\|f - g\|^2 + \lambda \sum_m |\langle g,\, \psi_m \rangle|$

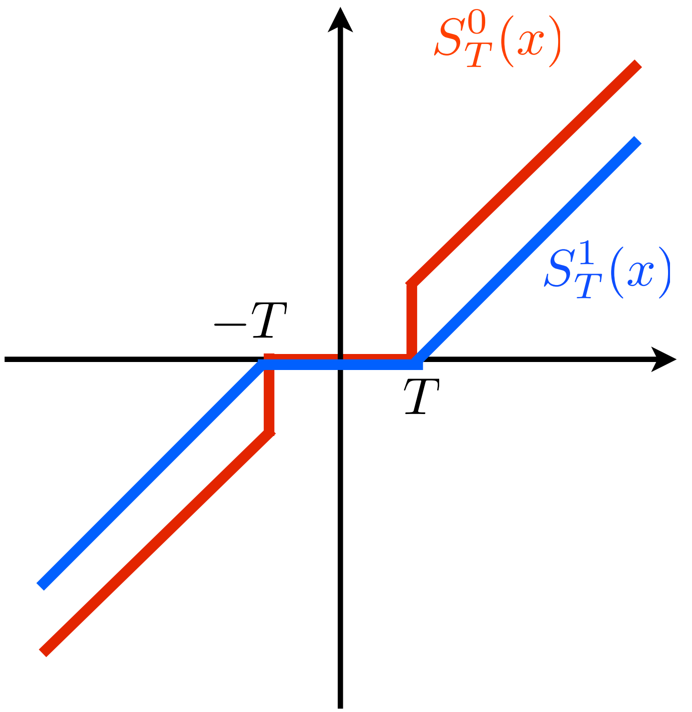
Sparse Regularization and Thresholding

$$f^\star_{\lambda,q} = \underset{g \in \mathbb{R}^N}{\operatorname{argmin}} \frac{1}{2}\|f-g\|^2 + \lambda \sum_m |\langle g,\, \psi_m\rangle|^q \qquad x^0 = \delta(x)$$

Orthogonal basis $\{\psi_m\}$.

Theorem: $$f_{\lambda,q} = \sum_m S_T^q(\langle f,\, \psi_m\rangle)\psi_m$$

where $\begin{cases} T = \sqrt{2\lambda} \quad \text{for} \quad q = 0, \\ T = \lambda \quad \text{for} \quad q = 1. \end{cases}$



In practice: $T \approx 3\sigma$.

Proof: Coefficients $a[m] = \langle f,\, \psi_m\rangle$,

$$a^\star_{\lambda,q}[m] = \langle f^\star_{\lambda,q},\, \psi_m\rangle$$

$$a^\star_{\lambda,q} = \underset{b \in \mathbb{R}^N}{\operatorname{argmin}} \sum_m \frac{1}{2}|a[m]-b[m]|^2 + \lambda |b[m]|^q$$

Overview

- Inverse Problems Regularization
- Sparse Prior in Orthogonal Bases
- **Sparse Regularization in Orthogonal Bases**
- Examples of applications:
  - *Sparse Diracs*: Seismic Deconvolution
  - *Sparse in 1D Wavelets*: Deblurring
  - *Sparse in 2D Wavelets*: Inpainting
- Redundant Representations
- Sparse Regularization in Redundant Dictionaries

Noiseless Sparse Regularization

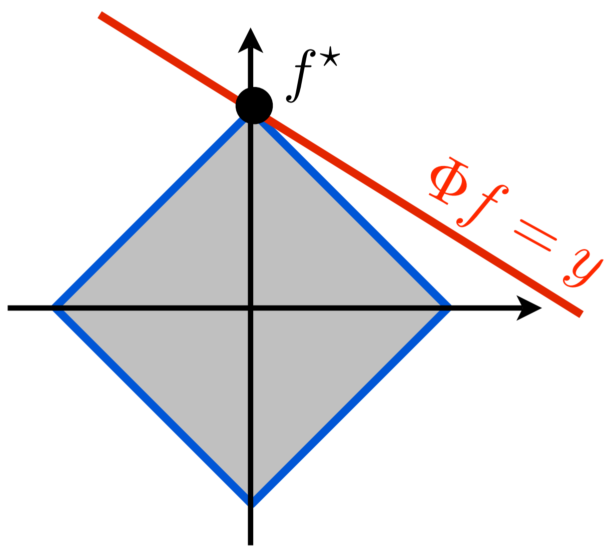
Orthogonal basis $\{\psi_m\}_m$ of $\mathbb{R}^N$. Sparse prior: $J(f) = \sum_m |\langle f,\, \psi_m \rangle|$

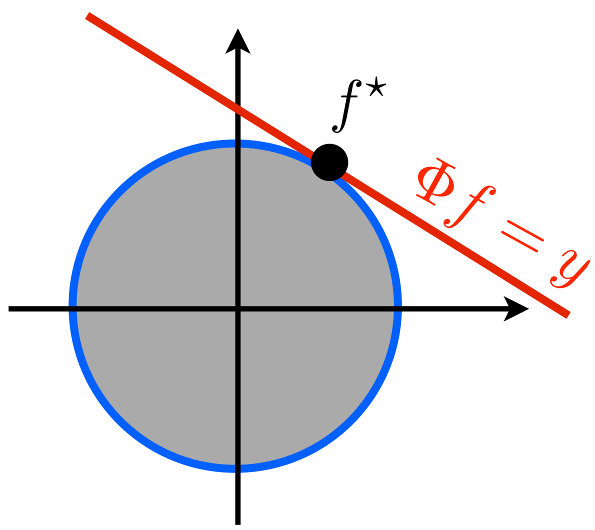
Noiseless measurements $y = \Phi f_0$: $\quad f^\star = \underset{\Phi f = y}{\text{argmin}} \sum_m |\langle f,\, \psi_m \rangle|$

Convex linear program.

→ Interior points, cf. [Chen, Donoho, Saunders] "basis pursuit".

→ Douglas-Rachford splitting, see [Combettes, Pesquet].





$f^\star = \underset{\Phi f = y}{\text{argmin}} \sum_m |\langle f,\, \psi_m \rangle|$ $\qquad$ $f^\star = \underset{\Phi f = y}{\text{argmin}} \sum_m |\langle f,\, \psi_m \rangle|^2 = \|f\|^2$

Noisy Sparse Regularization

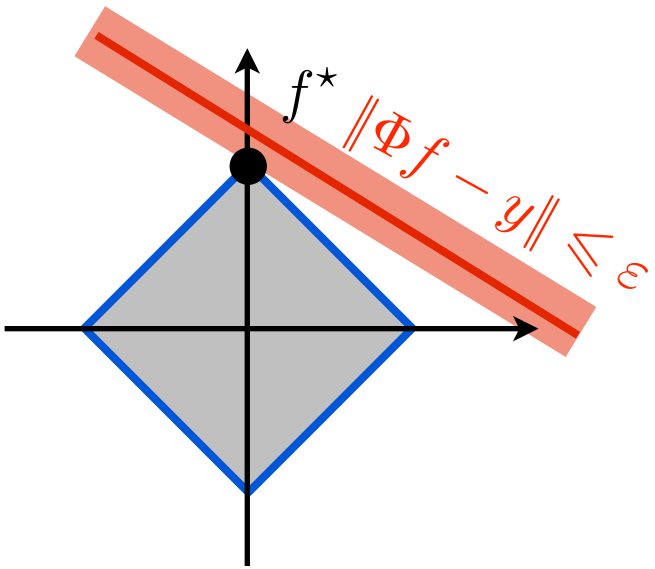
Noisy measurements: $y = \Phi f_0 + w$:
$$f^\star = \underset{\|\Phi f - y\| \leqslant \varepsilon}{\text{argmin}} \sum_m |\langle f,\, \psi_m \rangle|.$$

Convex program, can be solved with Lagrangian relaxation.

$$f^\star = \underset{f \in \mathbb{R}^N}{\text{argmin}} \underbrace{\boxed{\frac{1}{2}\|\Phi f - y\|^2}}_{\text{Data fitting}} + \underbrace{\boxed{\lambda \sum_m |\langle f,\, \psi_m \rangle|.}}_{\text{Sparsity}}$$

→ Iterative thresholding, see [Daubechies et al], [Pesquet et al], etc.

→ Nesterov multi-steps schemes.



Surrogate Functionals

Sparse regularization: $\quad f^\star = \underset{f}{\text{argmin}}\ E(f).$

Energy: $\quad E(f) = \frac{1}{2}\|\Phi f - y\|^2 + \lambda \sum_m |\langle f,\, \psi_m\rangle|.$

Surrogate functional:

$$E(f, f^{(k)}) = E(f) - \frac{1}{2}\|\Phi f - \Phi f^{(k)}\|^2 + \frac{1}{2\tau}\|f - f^{(k)}\|^2.$$

Surrogate Functionals

Sparse regularization: $\quad f^\star = \underset{f}{\text{argmin}}\ E(f).$

Energy: $\quad E(f) = \frac{1}{2}\|\Phi f - y\|^2 + \lambda \sum_m |\langle f,\, \psi_m \rangle|.$

Surrogate functional:

$$E(f, f^{(k)}) = E(f) - \frac{1}{2}\|\Phi f - \Phi f^{(k)}\|^2 + \frac{1}{2\tau}\|f - f^{(k)}\|^2.$$

$$= C + \frac{1}{2\tau}\|f - f^{(k)} + \tau\Phi^*(\Phi f^{(k)} - y)\|^2 + \lambda \sum_m |\langle f,\, \psi_m \rangle|$$

Theorem: $f \mapsto E(f, f^{(k)})$ is strictly convex and

$$\underset{f \in \mathbb{R}^N}{\text{argmin}}\ E(f, f^{(k)}) = \sum_m S^1_{\lambda\tau}(\langle \tilde{f}^{(k)},\, \psi_m \rangle)\psi_m$$

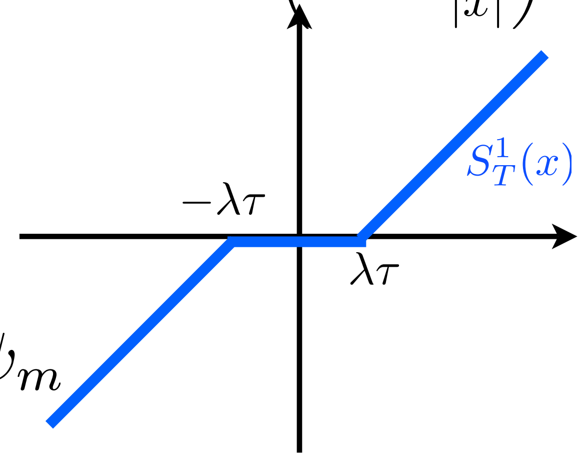
where $\quad \tilde{f}^{(k)} = f^{(k)} - \tau\Phi^*(\Phi f^{(k)} - y).$

Iterative Thresholding

$$S^1_{\lambda\tau}(x) = \max\left(0, 1 - \frac{\lambda\tau}{|x|}\right) x$$

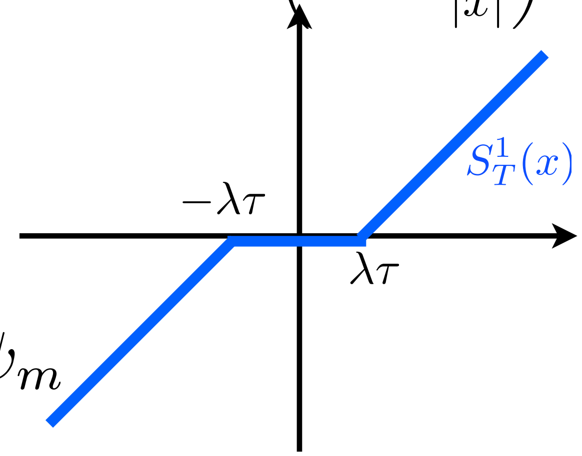
Algorithm: $f^{(k+1)} = \underset{f \in \mathbb{R}^N}{\operatorname{argmin}}\, E(f, f^{(k)})$

- *Inversion:* $\tilde{f}^{(k)} = f^{(k)} - \tau \Phi^* (\Phi f^{(k)} - y).$
- *Denoising:* $f^{(k+1)} = \sum_m S^1_{\lambda\tau}(\langle \tilde{f}^{(k)}, \psi_m \rangle) \psi_m$



Iterative Thresholding

$$S^1_{\lambda\tau}(x) = \max\left(0, 1 - \frac{\lambda\tau}{|x|}\right) x$$



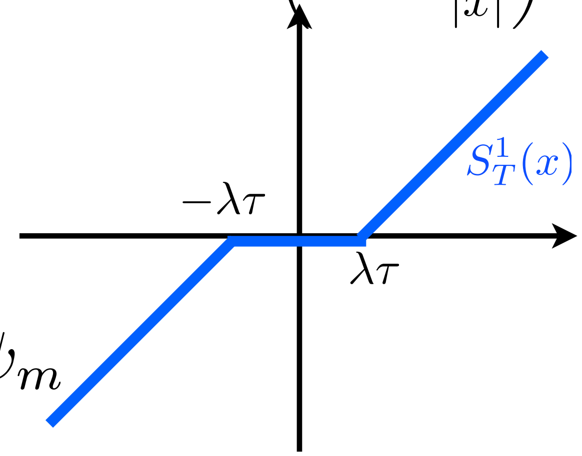
Algorithm: $$f^{(k+1)} = \underset{f \in \mathbb{R}^N}{\text{argmin}}\ E(f, f^{(k)})$$

- *Inversion:* $\tilde{f}^{(k)} = f^{(k)} - \tau \Phi^* (\Phi f^{(k)} - y).$
- *Denoising:* $f^{(k+1)} = \sum_m S^1_{\lambda\tau}(\langle \tilde{f}^{(k)}, \psi_m \rangle) \psi_m$

Theorem: if $\tau < 2/\|\Phi^*\Phi\|_S$, then $f^{(k)} \rightarrow f^\star$.

Iterative Thresholding

$$S^1_{\lambda\tau}(x) = \max\left(0, 1 - \frac{\lambda\tau}{|x|}\right) x$$



Algorithm: $$f^{(k+1)} = \underset{f \in \mathbb{R}^N}{\text{argmin}}\, E(f, f^{(k)})$$

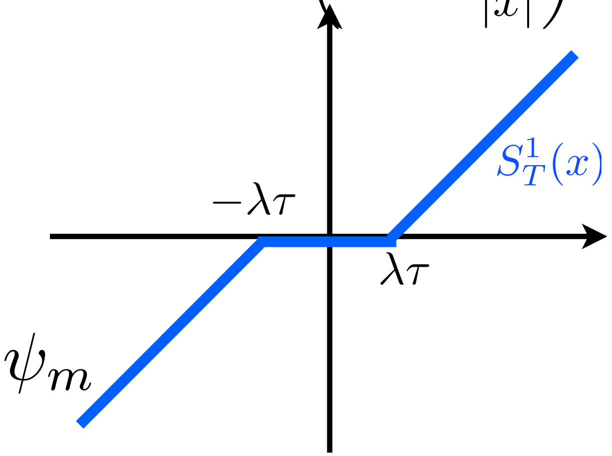
- *Inversion:* $\tilde{f}^{(k)} = f^{(k)} - \tau \Phi^* (\Phi f^{(k)} - y)$.
- *Denoising:* $f^{(k+1)} = \sum_m S^1_{\lambda\tau}(\langle \tilde{f}^{(k)}, \psi_m \rangle) \psi_m$

Theorem: if $\tau < 2/\|\Phi^*\Phi\|_S$, then $f^{(k)} \to f^\star$.

Noiseless case $\sigma = 0$: use a decaying value of $\lambda = \lambda_{\max}/k$.

Iterative Thresholding

$$S^1_{\lambda\tau}(x) = \max\left(0, 1 - \frac{\lambda\tau}{|x|}\right) x$$



Algorithm: $$f^{(k+1)} = \underset{f \in \mathbb{R}^N}{\text{argmin}}\, E(f, f^{(k)})$$

- *Inversion:* $\tilde{f}^{(k)} = f^{(k)} - \tau \Phi^*(\Phi f^{(k)} - y)$.
- *Denoising:* $f^{(k+1)} = \sum_m S^1_{\lambda\tau}(\langle \tilde{f}^{(k)}, \psi_m \rangle) \psi_m$

$$\boxed{\textit{Theorem:} \text{ if } \tau < 2/\|\Phi^*\Phi\|_S, \text{ then } f^{(k)} \to f^\star.}$$

Noiseless case $\sigma = 0$: use a decaying value of $\lambda = \lambda_{\max}/k$.

Constraint minimization: $$f^\star = \underset{\|\Phi f - y\| \leqslant \varepsilon}{\text{argmin}} \sum_m |\langle f, \psi_m \rangle|.$$

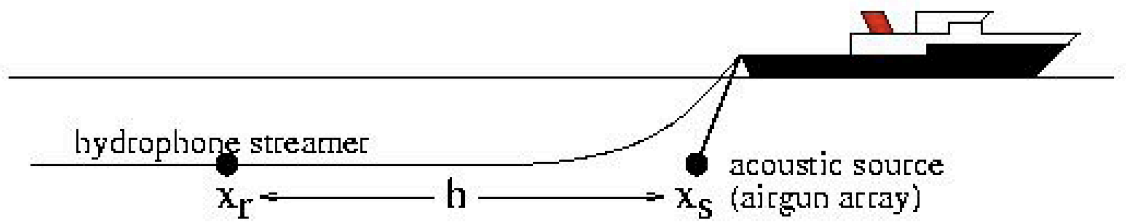
Mapping $\lambda \leftrightarrow \varepsilon$ depends on Φ and y.

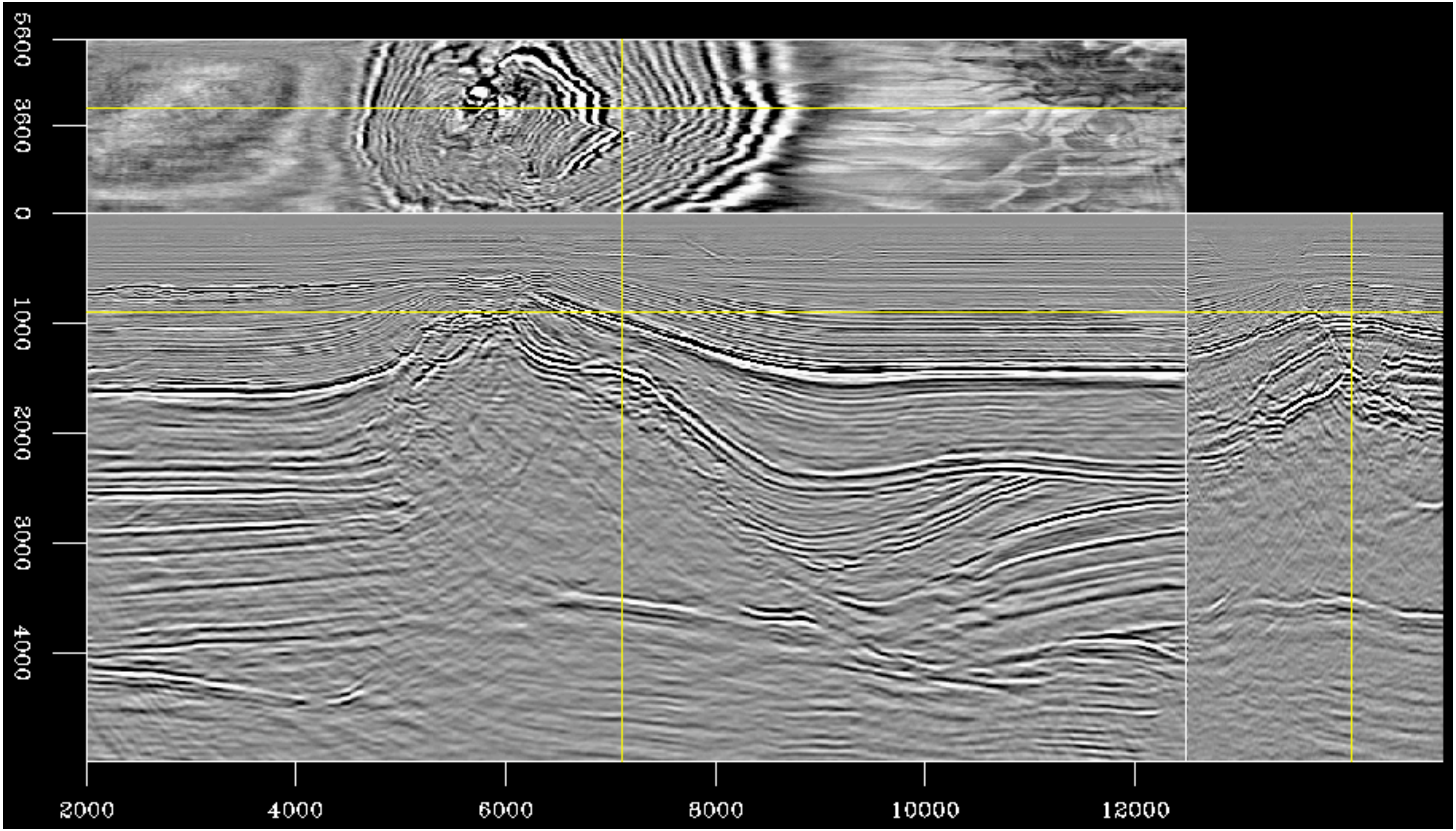
Iteratively update $\lambda = \lambda^{(k)}$: $$\lambda^{(k+1)} = \lambda^{(k)} \frac{\varepsilon}{\|\Phi f^{(k)} - y\|}$$

Overview

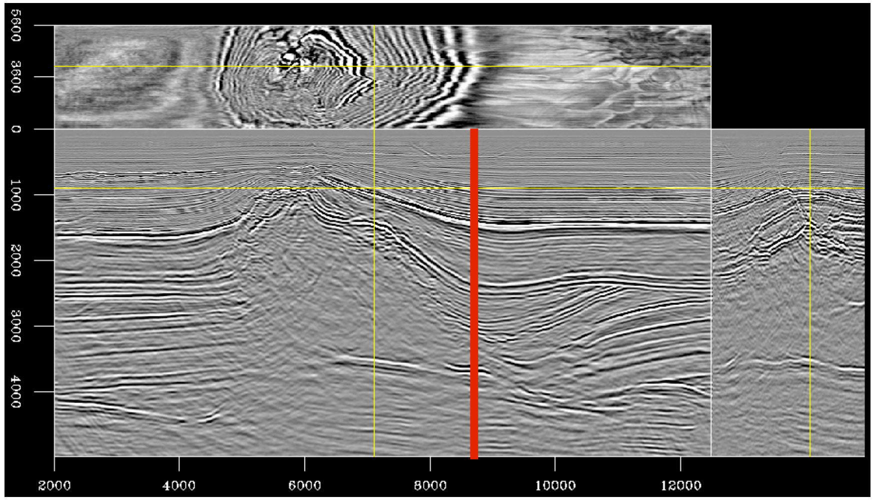
- Inverse Problems Regularization
- Sparse Prior in Orthogonal Bases
- Sparse Regularization in Orthogonal Bases
- Examples of applications:
  - *Sparse Diracs:* Seismic Deconvolution
  - *Sparse in 1D Wavelets*: Deblurring
  - *Sparse in 2D Wavelets*: Inpainting
- Redundant Representations
- Sparse Regularization in Redundant Dictionaries

Seismic Imaging





1D Idealization

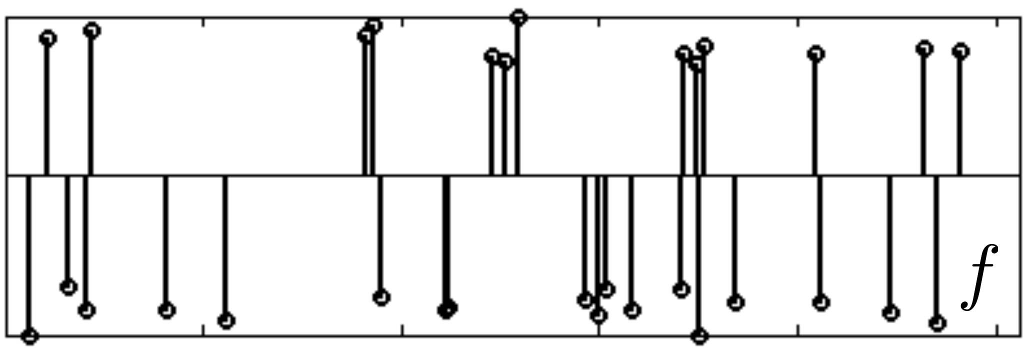


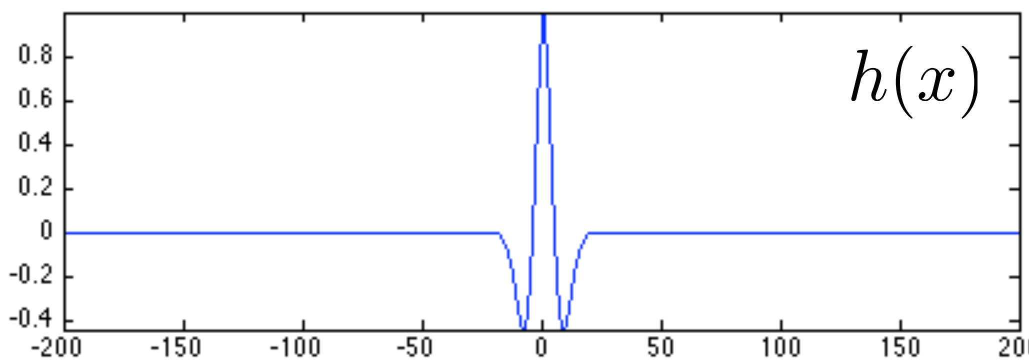
Initial condition: "wavelet"

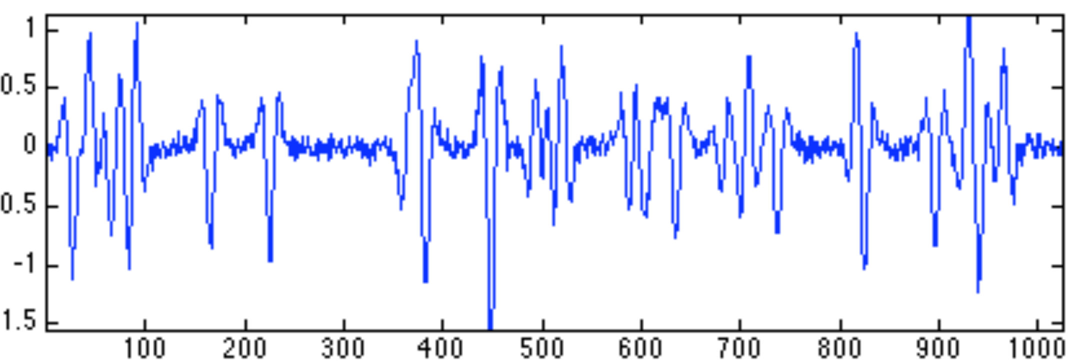
$= \text{band pass filter } h$

1D propagation $\approx$ convolution

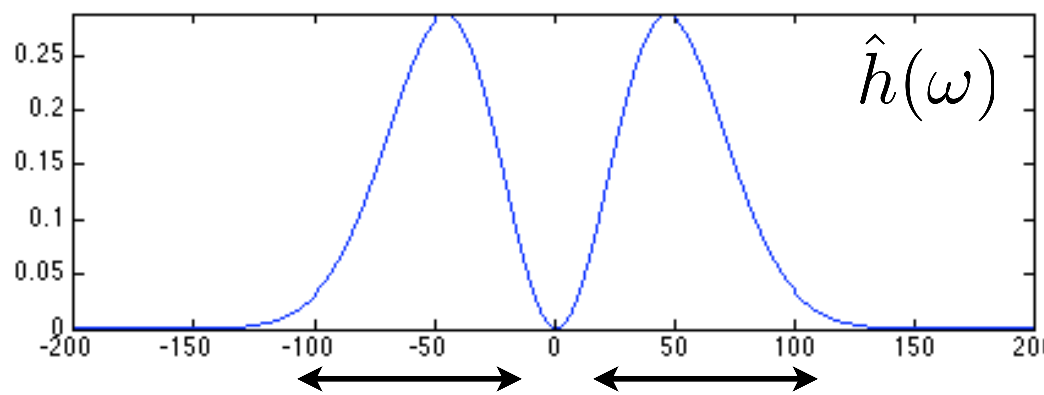
$$\Phi f = h \star f$$







$$y = f_0 \star h + w$$



P

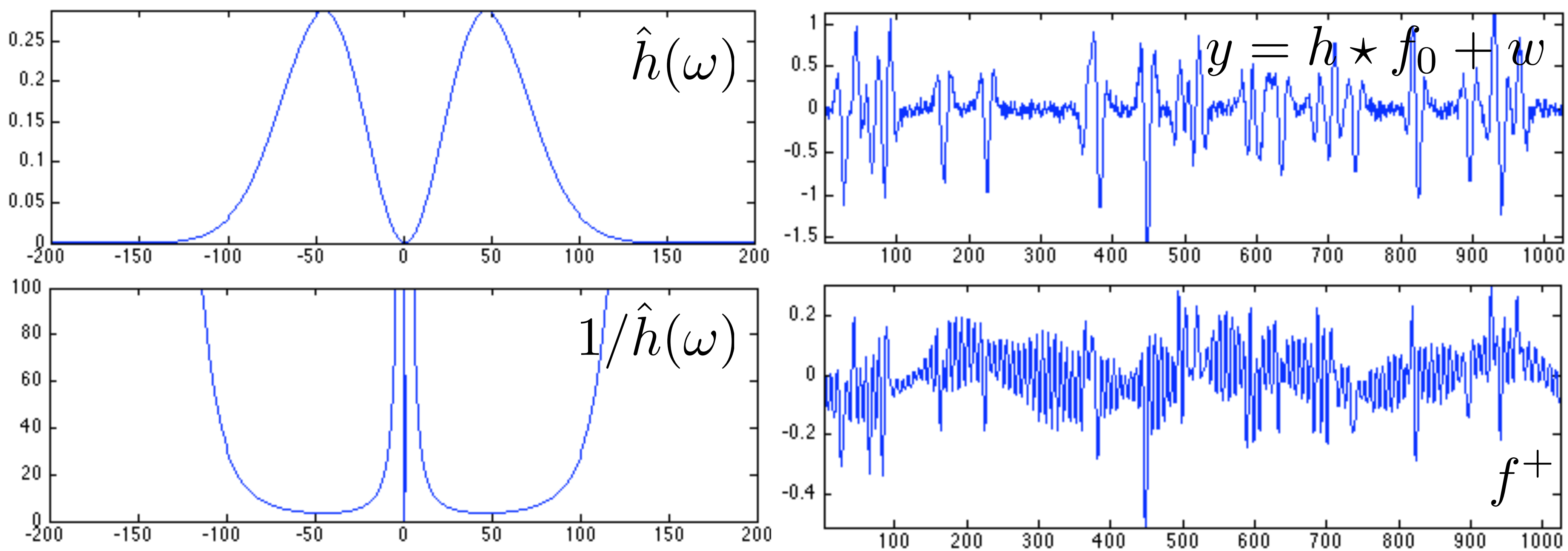
Pseudo Inverse

Pseudo-inverse: $$f^+ = \underset{\Phi f = y}{\text{argmin}} \; \|f\| = \Phi^*(\Phi\Phi^*)^{-1} y$$

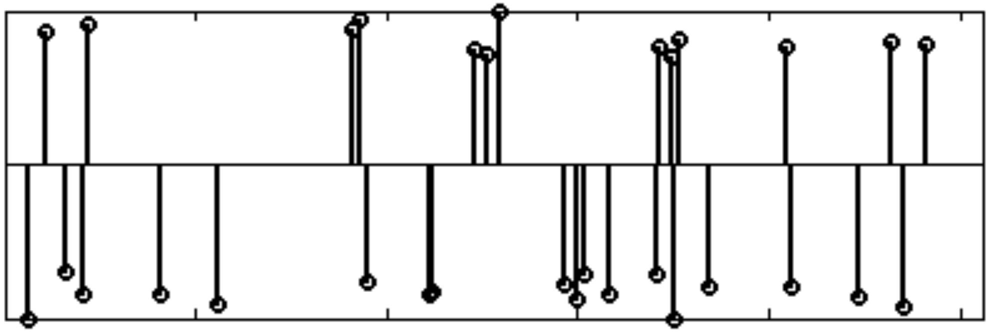
$$\hat{f}^+(\omega) = \frac{\hat{y}(\omega)}{h(\omega)} \quad \Longrightarrow \quad f^+ = h^+ \star y = f_0 + h^+ \star w$$

$$\text{where} \quad \hat{h}^+(\omega) = \hat{h}(\omega)^{-1}$$

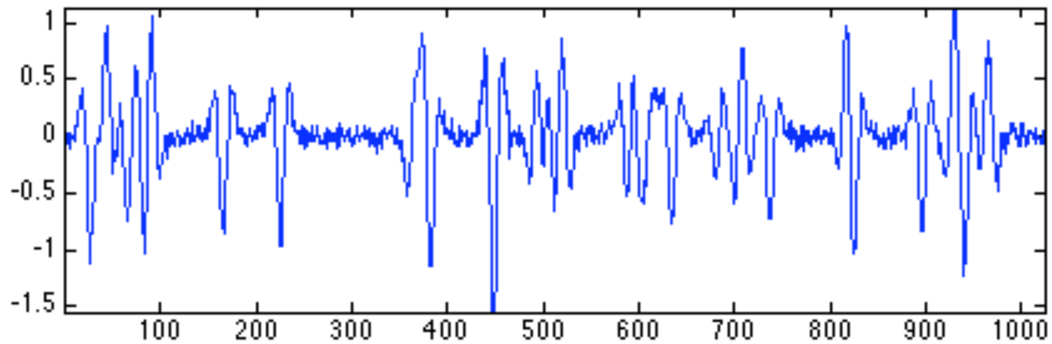
Stabilization: $\hat{h}^+(\omega) = 0 \quad \text{if} \quad |\hat{h}(\omega)| \leqslant \eta$



Sparse Spikes Deconvolution



f with small $\|f\|_0$

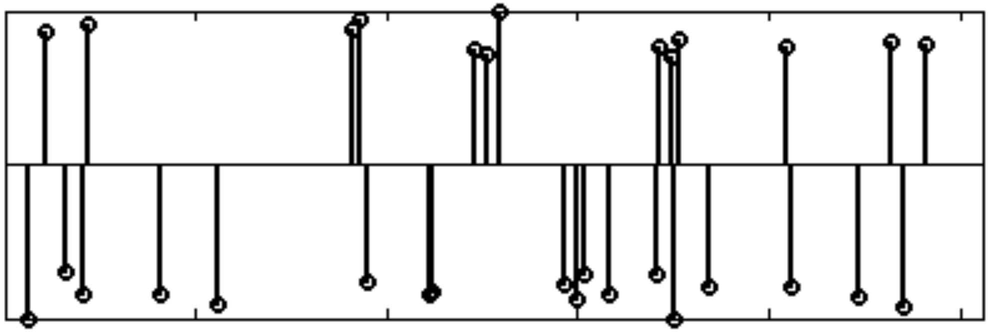


$y = f \star h + w$

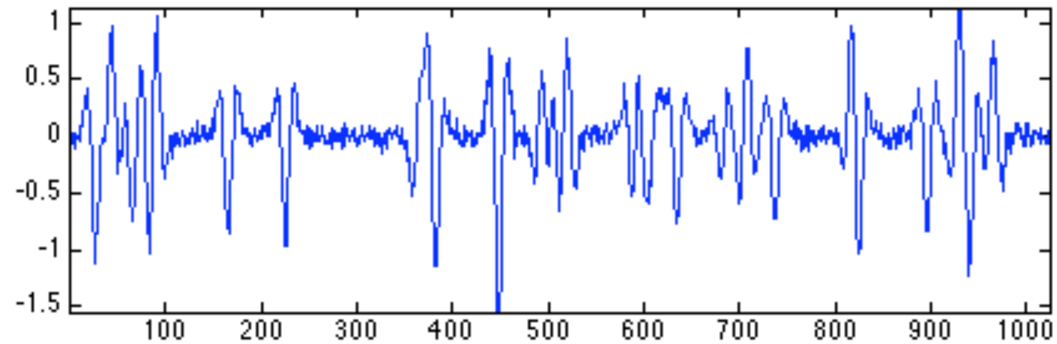
Sparsity basis: Diracs $\qquad \psi_m[x] = \delta[x-m]$

$$f^\star = \underset{f \in \mathbb{R}^N}{\text{argmin}} \; \frac{1}{2}\|f \star h - y\|^2 + \lambda \sum_m |f[m]|.$$

Sparse Spikes Deconvolution



f with small $\|f\|_0$



$y = f \star h + w$

Sparsity basis: Diracs $\qquad \psi_m[x] = \delta[x - m]$

$$f^\star = \underset{f \in \mathbb{R}^N}{\text{argmin}} \; \frac{1}{2}\|f \star h - y\|^2 + \lambda \sum_m |f[m]|.$$

Algorithm: $\qquad \tau < 2/\|\Phi^*\Phi\| = 2/\max_\omega |\hat{h}(\omega)|^2$

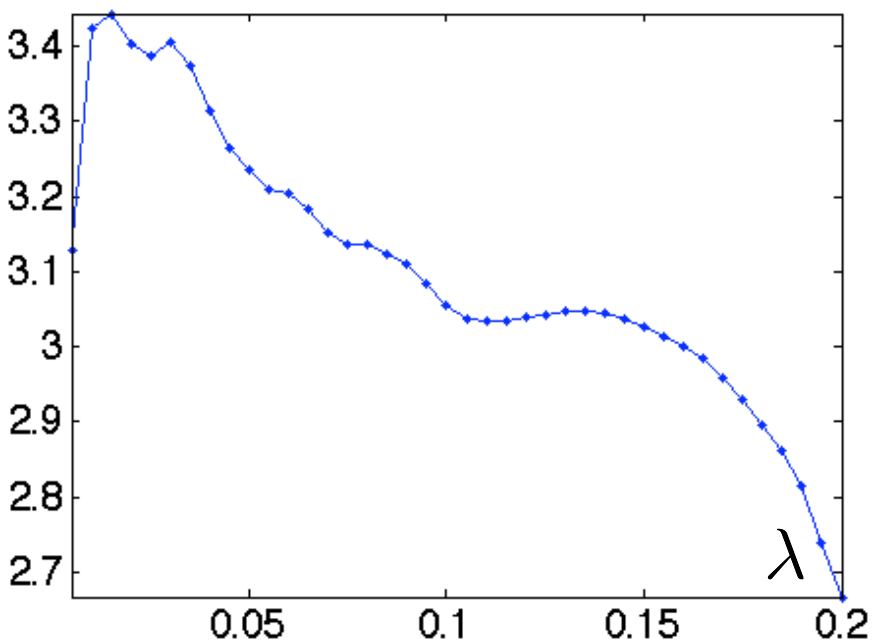
- *Inversion:* $\tilde{f}^{(k)} = f^{(k)} - \tau h \star (h \star f^{(k)} - y).$
- *Denoising:* $f^{(k+1)}[m] = S^1_{\lambda\tau}(\tilde{f}^{(k)}[m])$

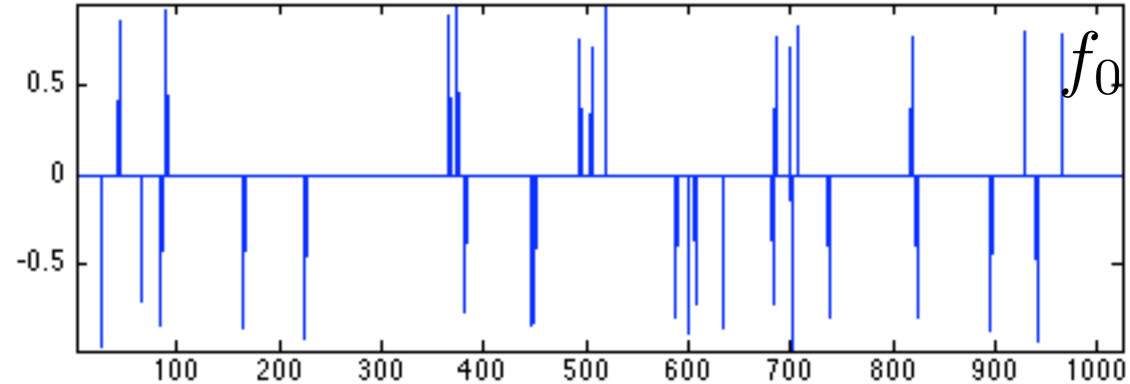
Numerical Example

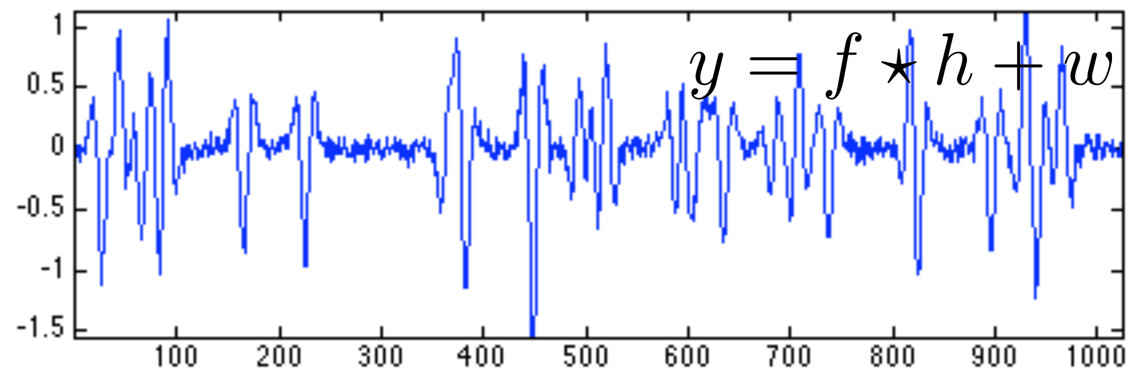
Choosing optimal λ:

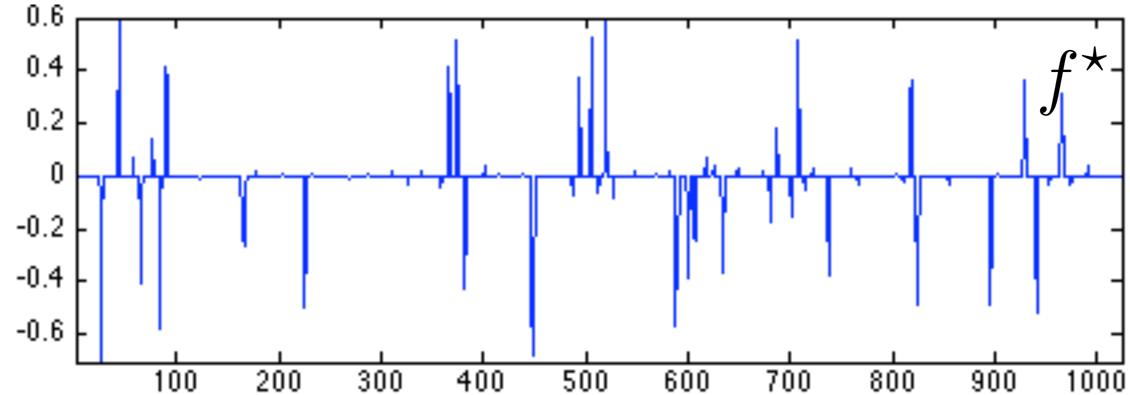
oracle, minimize $\|f_0 - f^\star\|$

$\mathrm{SNR}(f_0, f^\star)$







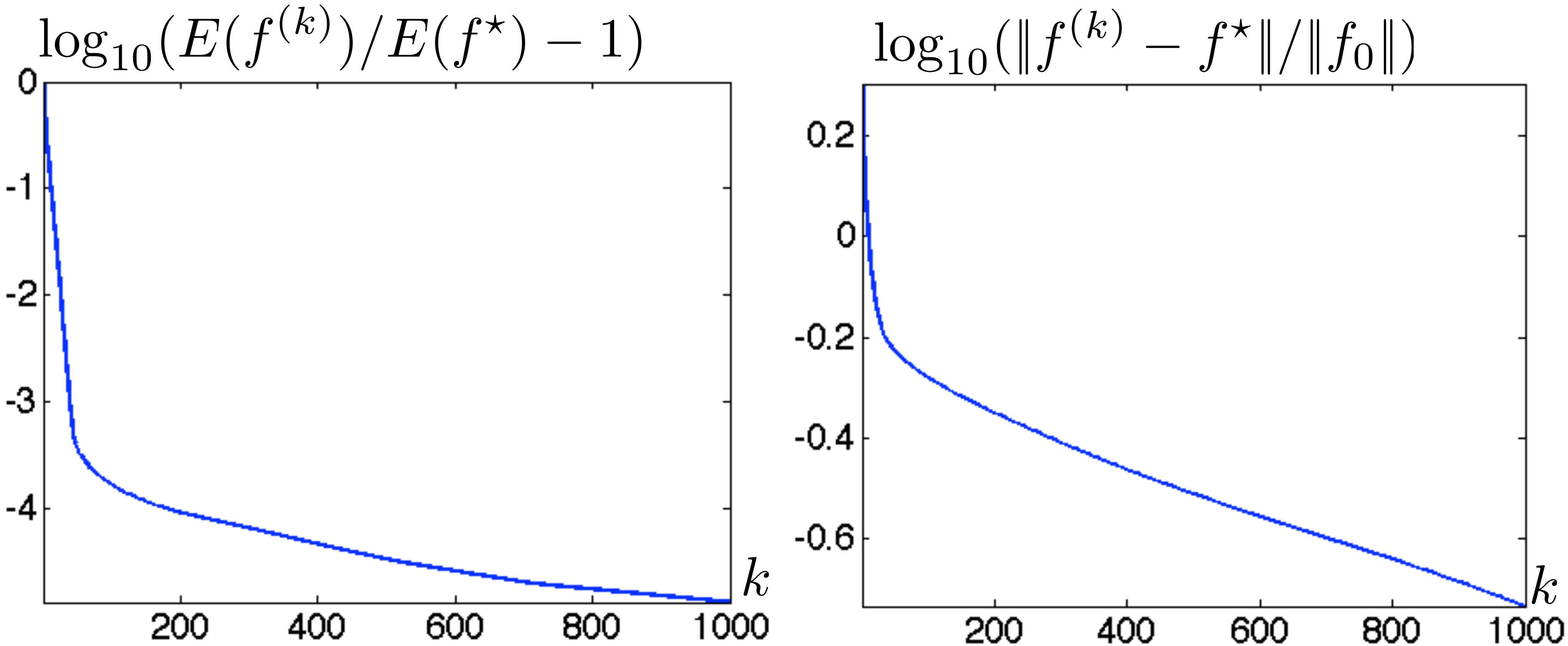


Convergence Study

Sparse deconvolution: $\quad f^\star = \underset{f \in \mathbb{R}^N}{\text{argmin}}\ E(f).$

$$\textit{Energy:} \quad E(f) = \frac{1}{2}\|h \star f - y\|^2 + \lambda \sum_m |f[m]|.$$

Not strictly convex $\implies$ no convergence speed.



Overview

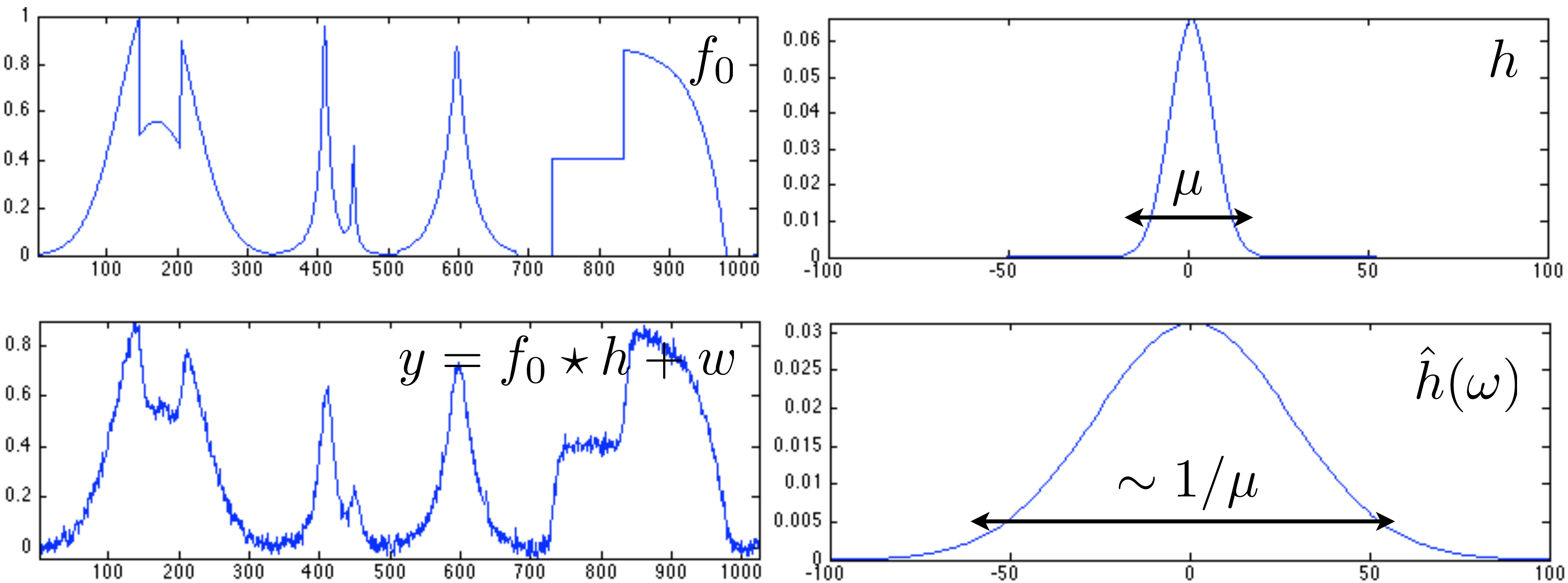
- Inverse Problems Regularization
- Sparse Prior in Orthogonal Bases
- Sparse Regularization in Orthogonal Bases
- Examples of applications:
  - *Sparse Diracs*: Seismic Deconvolution
  - ***Sparse in 1D Wavelets*: Deblurring**
  - *Sparse in 2D Wavelets*: Inpainting
- Redundant Representations
- Sparse Regularization in Redundant Dictionaries

De-blurring Problem

De-blurring: $\quad y = f_0 \star h + w.$

Optical blur: h Gaussian filter of width $\mu > 0$.

effective measurements: $P \sim 1/\mu$.



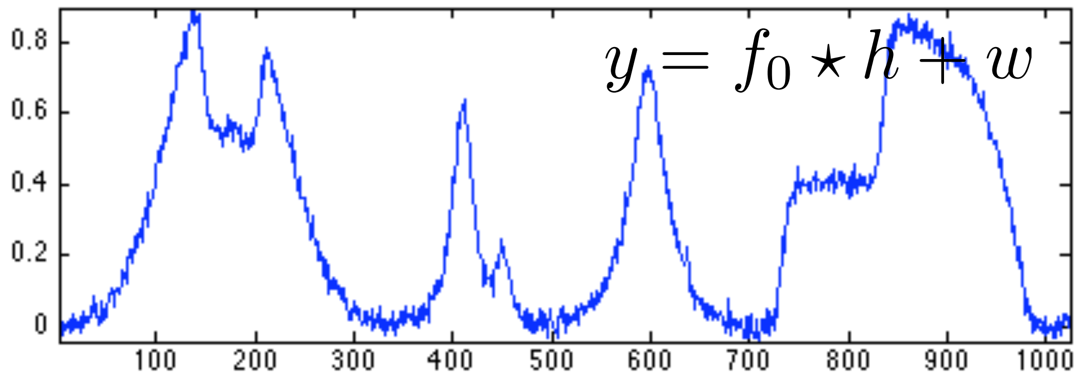
Sparse Wavelet De-blurring

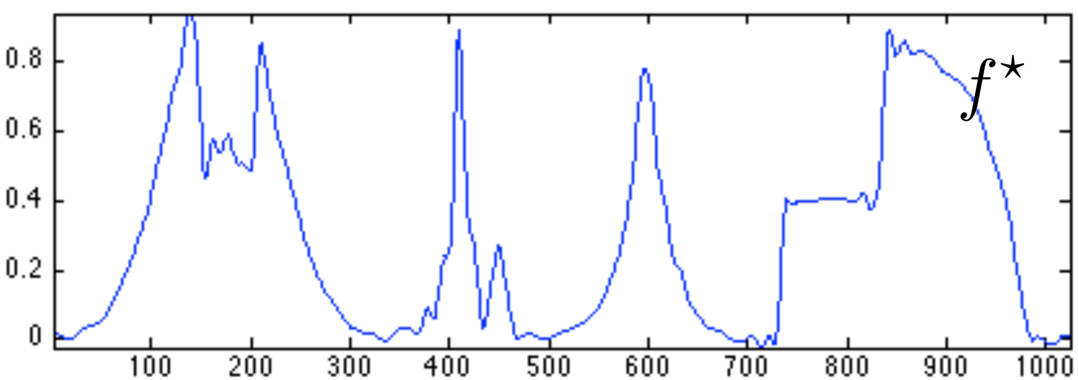
Sparsity of f_0 in wavelets $\{\psi_m\} = \{\psi_{j,n}\}_{m=(j,n)}$.

$$f^\star = \underset{f \in \mathbb{R}^N}{\text{argmin}} \ \frac{1}{2}\|f \star h - y\|^2 + \lambda \sum_m |\langle f,\, \psi_m \rangle|.$$

- *Inversion:* $\tilde f^{(k)} = f^{(k)} - \tau \Phi^*(\Phi f^{(k)} - y)$.
- *Denoising:* $f^{(k+1)} = \sum_m S^1_{\lambda\tau}(\langle \tilde f^{(k)},\, \psi_m\rangle)\psi_m$

$$\tau < \frac{2}{\max_\omega |\hat h(\omega)|^2}$$





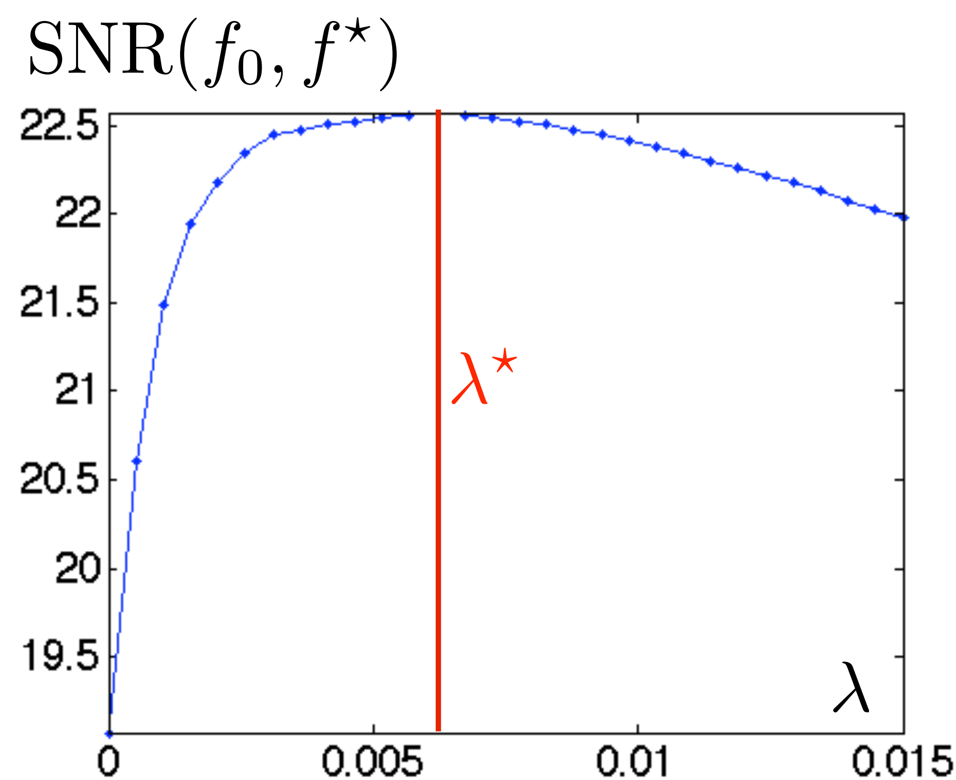


Image De-blurring

Original f_0

$y = h \star f_0 + w$

Image De-blurring

Original f_0

$y = h \star f_0 + w$

L^2 regularization:

$$f^\star = \underset{f \in \mathbb{R}^N}{\operatorname{argmin}} \; \|f \star h - y\|^2 + \lambda \|f\|^2$$

$$\hat{f}^\star(\omega) = \frac{\hat{h}(\omega)}{|\hat{h}(\omega)|^2 + \lambda} \hat{y}(\omega)$$

Image De-blurring

Original f_0

$y = h \star f_0 + w$

L^2 regularization:

$$f^\star = \underset{f \in \mathbb{R}^N}{\operatorname{argmin}} \|f \star h - y\|^2 + \lambda \|f\|^2$$

$$\hat{f}^\star(\omega) = \frac{\hat{h}(\omega)}{|\hat{h}(\omega)|^2 + \lambda} \hat{y}(\omega)$$

Sobolev regularization: $$f^\star = \underset{f \in \mathbb{R}^N}{\operatorname{argmin}} \|f \star h - y\|^2 + \lambda \|\nabla f\|^2$$

$$\hat{f}^\star(\omega) = \frac{\hat{h}(\omega)}{|\hat{h}(\omega)|^2 + \lambda |\omega|^2} \hat{y}(\omega)$$

Image De-blurring

Original f_0

$y = h \star f_0 + w$

L^2 regularization:

$$f^\star = \underset{f \in \mathbb{R}^N}{\operatorname{argmin}} \|f \star h - y\|^2 + \lambda \|f\|^2$$

$$\hat{f}^\star(\omega) = \frac{\hat{h}(\omega)}{|\hat{h}(\omega)|^2 + \lambda} \hat{y}(\omega)$$

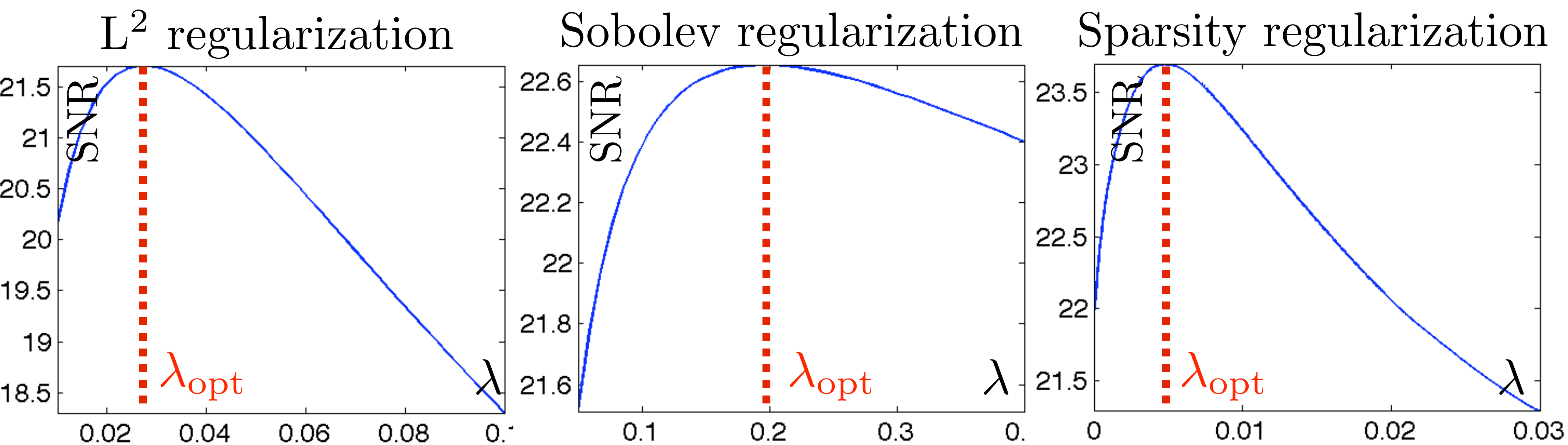
Sobolev regularization: $$f^\star = \underset{f \in \mathbb{R}^N}{\operatorname{argmin}} \|f \star h - y\|^2 + \lambda \|\nabla f\|^2$$

$$\hat{f}^\star(\omega) = \frac{\hat{h}(\omega)}{|\hat{h}(\omega)|^2 + \lambda |\omega|^2} \hat{y}(\omega)$$

Sparsity regularization:

$$f^\star = \underset{f \in \mathbb{R}^N}{\operatorname{argmin}} \frac{1}{2}\|f \star h - y\|^2 + \lambda \sum_m |\langle f,\, \psi_m \rangle|.$$

Comparison of Regularizations



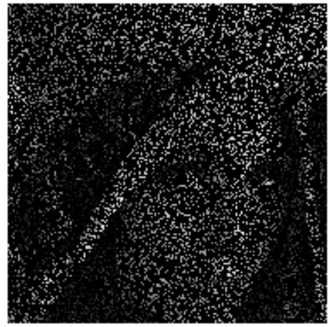
L^2	Sobolev	Sparsity	Invariant
SNR=21.7dB	SNR=22.7dB	SNR=23.7dB	SNR=24.7dB

Overview

- Inverse Problems Regularization
- Sparse Prior in Orthogonal Bases
- Sparse Regularization in Orthogonal Bases
- Examples of applications:
  - *Sparse Diracs*: Seismic Deconvolution
  - *Sparse in 1D Wavelets*: Deblurring
  - ***Sparse in 2D Wavelets*: Inpainting**
- Redundant Representations
- Sparse Regularization in Redundant Dictionaries

Inpainting Problem

Inpainting: set $\Omega \subset \{0, \ldots, N-1\}$ of missing pixels, $P = N - |\Omega|$.

 $\xrightarrow{\Phi}$ 

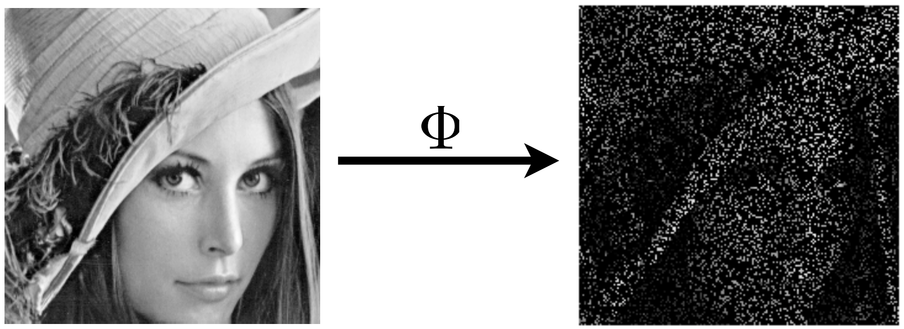
$$(\Phi f)(x) = \begin{cases} 0 & \text{if} \quad x \in \Omega, \\ f(x) & \text{if} \quad x \notin \Omega. \end{cases}$$

Measures: $\quad y = \Phi f_0 + w \qquad\qquad \Phi = \mathrm{diag}_m(\delta_{\Omega^c}[m])$

Recovery: $\quad f^\star = \underset{f \in \mathbb{R}^N}{\mathrm{argmin}} \; \frac{1}{2}\|\Phi f - y\|^2 + \lambda \sum_m |\langle f,\, \psi_m \rangle|.$

Inpainting Problem

Inpainting: set $\Omega \subset \{0, \ldots, N-1\}$ of missing pixels, $P = N - |\Omega|$.



$$(\Phi f)(x) = \begin{cases} 0 & \text{if} \quad x \in \Omega, \\ f(x) & \text{if} \quad x \notin \Omega. \end{cases}$$

Measures: $y = \Phi f_0 + w$ $\qquad \Phi = \text{diag}_m(\delta_{\Omega^c}[m])$

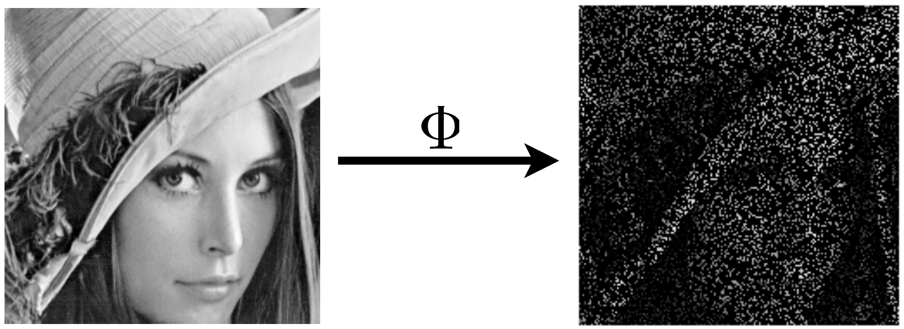
Recovery: $$f^\star = \underset{f \in \mathbb{R}^N}{\text{argmin}} \frac{1}{2}\|\Phi f - y\|^2 + \lambda \sum_m |\langle f, \psi_m \rangle|.$$

Algorithm: $\tau < 2$, simple expression for $\tau = 1$,

- *Inversion:* $\tilde{f}^{(k)}[m] = \begin{cases} f^{(k)}[m] & \text{if} \quad x \in \Omega, \\ y[m] & \text{if} \quad x \notin \Omega. \end{cases}$.
- *Denoising:* $f^{(k+1)} = \sum_m S^1_\lambda(\langle \tilde{f}^{(k)}, \psi_m \rangle)\psi_m$

Inpainting Problem

Inpainting: set $\Omega \subset \{0, \ldots, N-1\}$ of missing pixels, $P = N - |\Omega|$.



$$(\Phi f)(x) = \begin{cases} 0 & \text{if} \quad x \in \Omega, \\ f(x) & \text{if} \quad x \notin \Omega. \end{cases}$$

Measures: $y = \Phi f_0 + w$ $\qquad\qquad \Phi = \mathrm{diag}_m(\delta_{\Omega^c}[m])$

Recovery: $$f^\star = \underset{f \in \mathbb{R}^N}{\mathrm{argmin}} \; \frac{1}{2}\|\Phi f - y\|^2 + \lambda \sum_m |\langle f, \psi_m \rangle|.$$

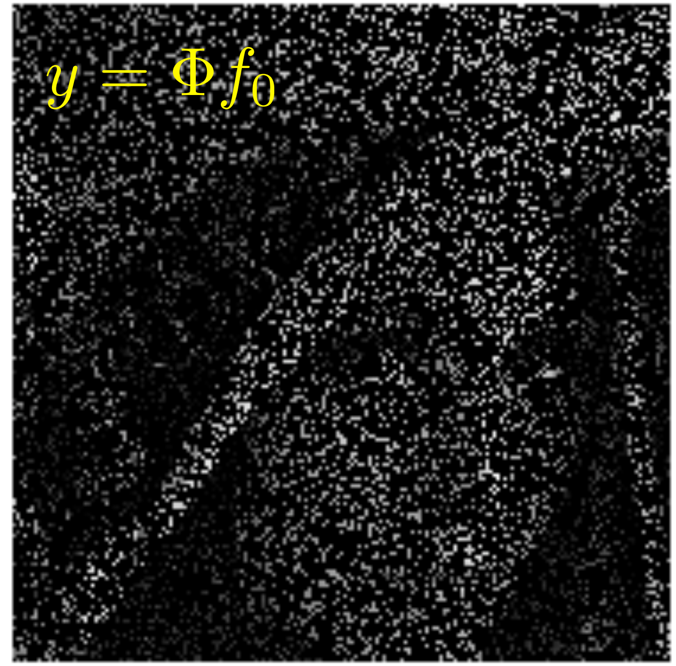
Algorithm: $\tau < 2$, simple expression for $\tau = 1$,

- *Inversion:* $\tilde{f}^{(k)}[m] = \begin{cases} f^{(k)}[m] & \text{if} \quad x \in \Omega, \\ y[m] & \text{if} \quad x \notin \Omega. \end{cases}$.
- *Denoising:* $f^{(k+1)} = \sum_m S^1_\lambda(\langle \tilde{f}^{(k)}, \psi_m \rangle)\psi_m$

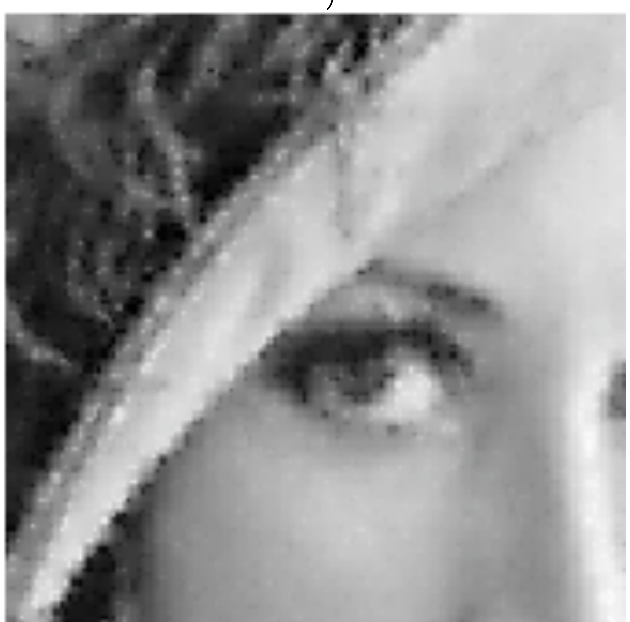
Extension: translation invariant wavelet frame denoising.

$\longrightarrow$ does not minimize a variational functional.

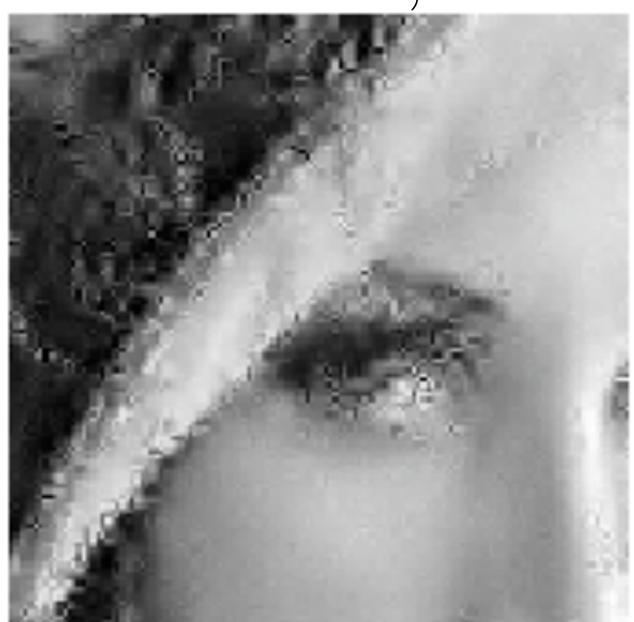
Inpainting Results



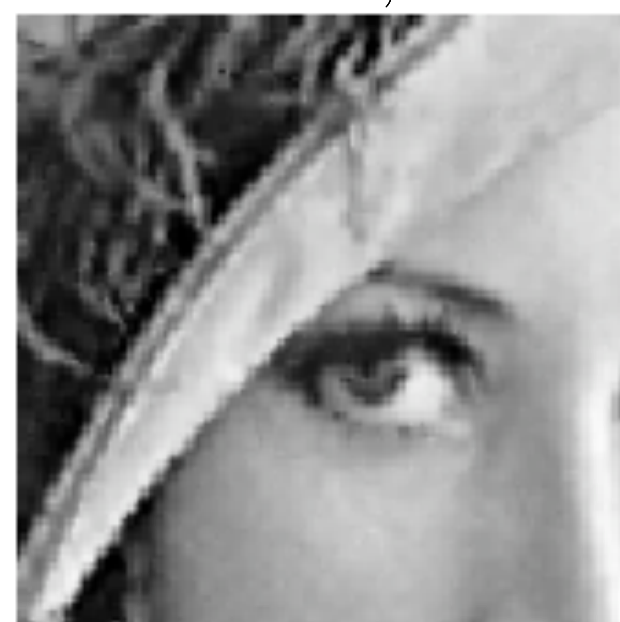
Sobolev, 20.8dB



Wavelets orth, 16.6dB



Wavelets TI, 23.6dB



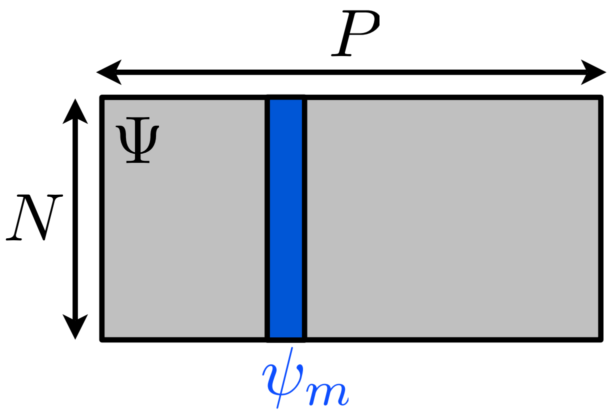
Overview

- Inverse Problems Regularization
- Sparse Prior in Orthogonal Bases
- Sparse Regularization in Orthogonal Bases
- Examples of applications:
  - *Sparse Diracs*: Seismic Deconvolution
  - *Sparse in 1D Wavelets*: Deblurring
  - *Sparse in 2D Wavelets*: Inpainting
- **Redundant Representations**
- Sparse Regularization in Redundant Dictionaries

Sparsity in Redundant Dictionaries

Redundant dictionary of $\mathbb{R}^N$: $\{\psi_m\}_{m=0}^{P-1}$, $P \geqslant N$.

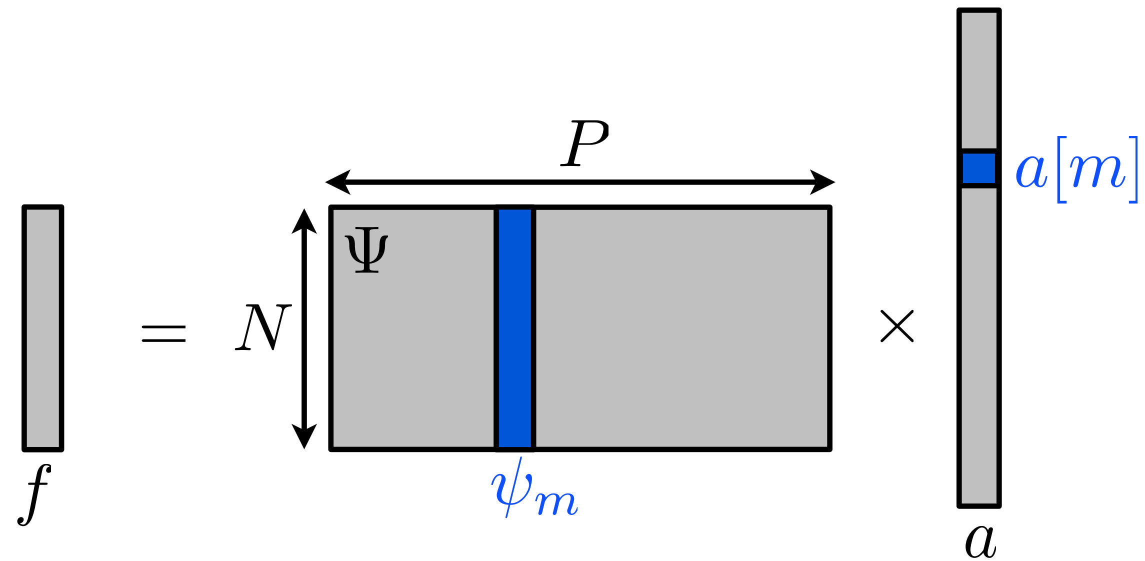
Examples: translation invariant wavelet frame,
union of ortho-bases, {Fourier, Wavelets}.



Sparsity in Redundant Dictionaries

Redundant dictionary of $\mathbb{R}^N$: $\{\psi_m\}_{m=0}^{P-1}$, $P \geqslant N$.

Examples: translation invariant wavelet frame,
union of ortho-bases, {Fourier, Wavelets}.



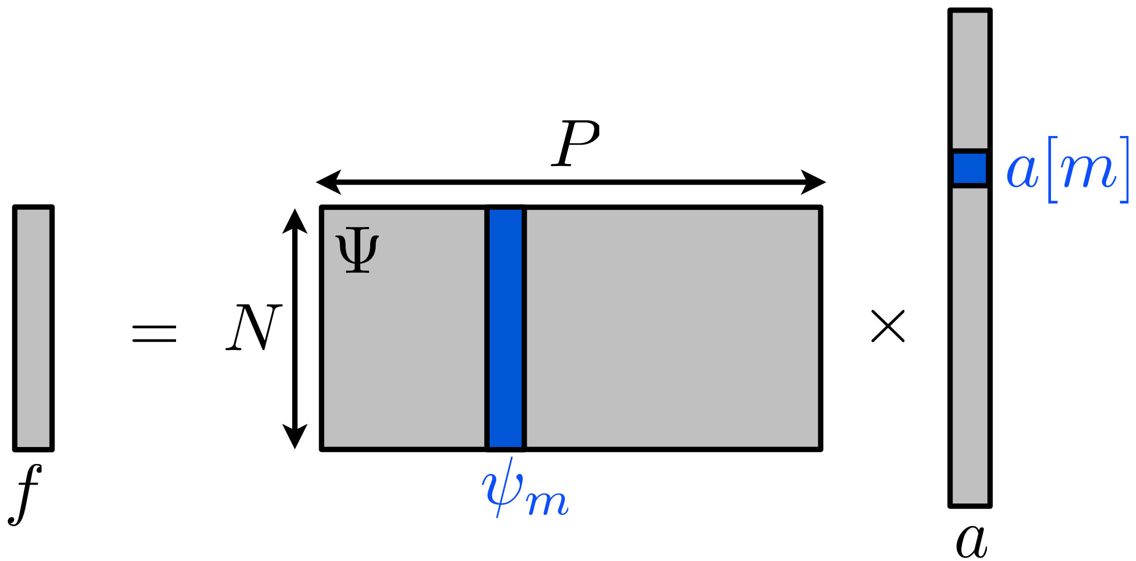
Signal decomposition: $$f = \Psi a = \sum_m a[m]\psi_m$$

$\longrightarrow$ non-unique if $P > N$.

Sparsity in Redundant Dictionaries

Redundant dictionary of $\mathbb{R}^N$: $\{\psi_m\}_{m=0}^{P-1}$, $P \geqslant N$.

Examples: translation invariant wavelet frame,
union of ortho-bases, {Fourier, Wavelets}.



Signal decomposition: $\quad f = \Psi a = \sum_m a[m]\psi_m$

$\longrightarrow$ non-unique if $P > N$.

Best M-term approximation: $f_M = \Psi a_M, \quad a_M = \underset{\|a\|_0 \leqslant M}{\text{argmin}} \|f - \Psi a\|$

$\longrightarrow$ NP-hard to compute for general Ψ.

Sparse L1 Redundant Approximation

Greedy algorithms: (Orthogonal) Matching Pursuit [Zhang, Davis, Mallat].

Convex relaxation: Basis Pursuit [Chen,Donoho].

Best ℓ^1 approximation: $f_\varepsilon^\star = \Psi a_\varepsilon^\star$, $\quad a_\varepsilon^\star = \underset{\|f - \Psi a\| \leqslant \varepsilon}{\text{argmin}} \; \|a\|_1$

$$f_\lambda^\star = \Psi a_\lambda^\star, \qquad a_\lambda^\star = \underset{a \in \mathbb{R}^P}{\text{argmin}} \; \frac{1}{2}\|f - \Psi a\|^2 + \lambda \|a\|_1$$

Equivalent inverse problem:

$f = \Psi a + w$, $w =$approximation residual.

Applications: denoising, image separation.

Sparse L1 Redundant Approximation

Greedy algorithms: (Orthogonal) Matching Pursuit [Zhang, Davis, Mallat].

Convex relaxation: Basis Pursuit [Chen,Donoho].

Best ℓ^1 approximation: $\quad f_\varepsilon^\star = \Psi a_\varepsilon^\star, \quad a_\varepsilon^\star = \underset{\|f - \Psi a\| \leqslant \varepsilon}{\text{argmin}} \; \|a\|_1$

$$f_\lambda^\star = \Psi a_\lambda^\star, \qquad a_\lambda^\star = \underset{a \in \mathbb{R}^P}{\text{argmin}} \; \frac{1}{2}\|f - \Psi a\|^2 + \lambda \|a\|_1$$

Equivalent inverse problem:

$f = \Psi a + w$, $w =$approximation residual.

Applications: denoising, image separation.

Iterative thresholding:

$$a^{(k+1)} = S_{\lambda\tau}\left(a^{(k)} + \tau \Psi^*(f - \Psi a^{(k)})\right)$$

If $\tau \leqslant 2/\|\Psi^* \Psi\|$, $\quad a^{(k)} \to a_\lambda^\star$

Image Separation

Decompose $f = f_1^\star + f_2^\star + w$, $(f_1^\star, f_2^\star)$ morphological components, w noise.

Differente morphologies: sparse in different dictionaries Ψ_1, Ψ_2.

Image Separation

Decompose $f = f_1^\star + f_2^\star + w$, $(f_1^\star, f_2^\star)$ morphological components, w noise.

Differente morphologies: sparse in different dictionaries Ψ_1, Ψ_2.

Isotropic: wavelet frame.

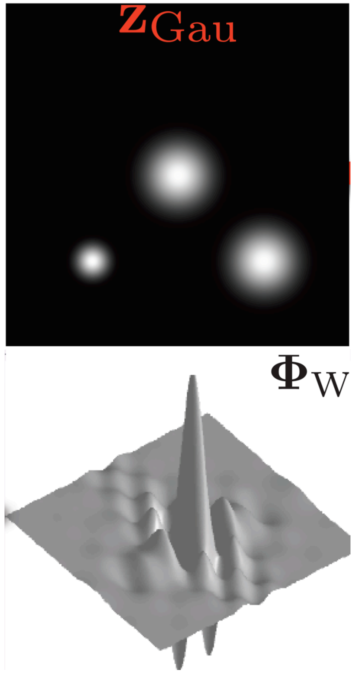


Image Separation

Decompose $f = f_1^\star + f_2^\star + w$, $(f_1^\star, f_2^\star)$ morphological components, w noise.

Differente morphologies: sparse in different dictionaries Ψ_1, Ψ_2.

Isotropic: wavelet frame.

Texture: local cosine frame.

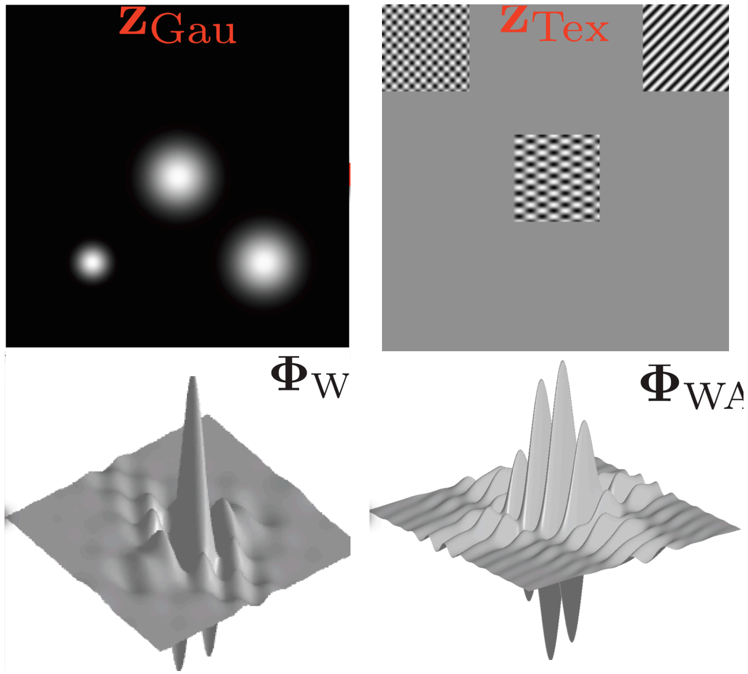


Image Separation

Decompose $f = f_1^\star + f_2^\star + w$, $(f_1^\star, f_2^\star)$ morphological components, w noise.

Differente morphologies: sparse in different dictionaries Ψ_1, Ψ_2.

Isotropic: wavelet frame.

Texture: local cosine frame.

Lines: ridgelets frame.

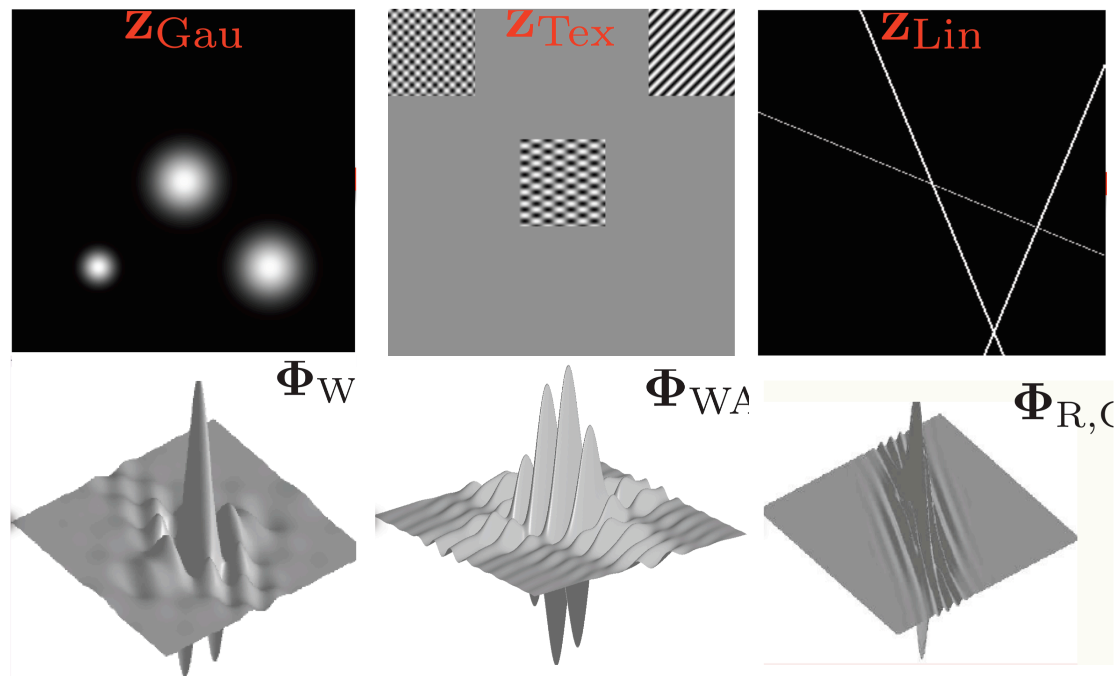


Image Separation

Decompose $f = f_1^\star + f_2^\star + w$, $(f_1^\star, f_2^\star)$ morphological components, w noise.

Differente morphologies: sparse in different dictionaries Ψ_1, Ψ_2.

Isotropic: wavelet frame.

Texture: local cosine frame.

Lines: ridgelets frame.

Cartoon: curvelet frame.

Other: adaptivity (bandlets), learning, ...

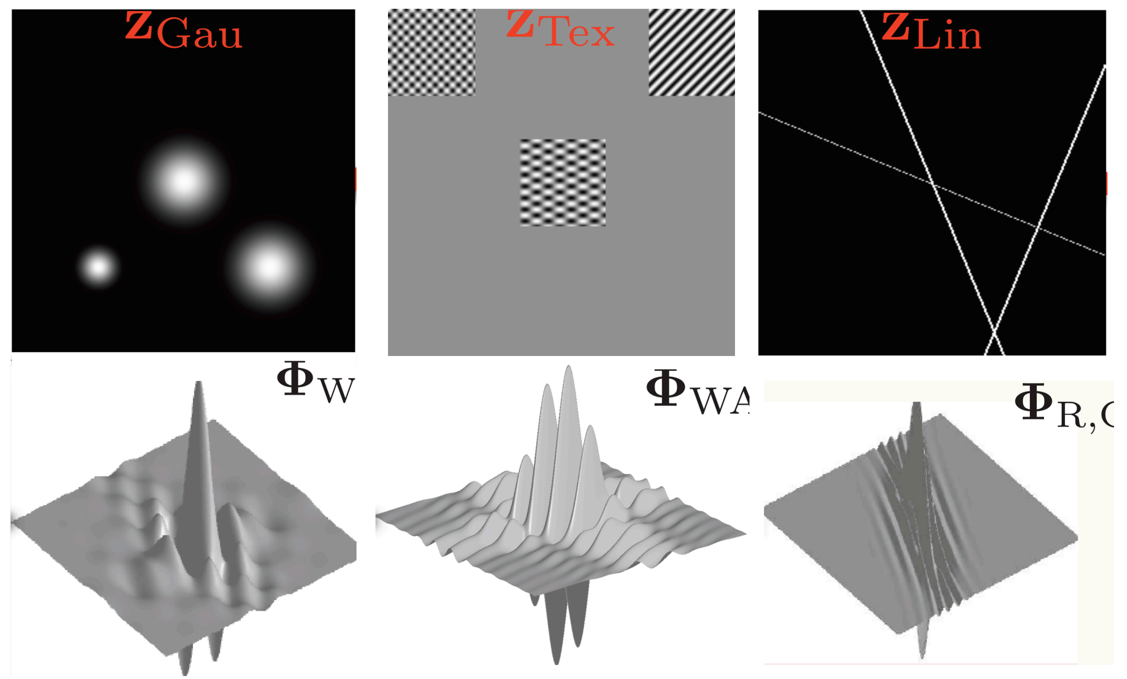


Image Separation

Decompose $f = f_1^\star + f_2^\star + w$, $(f_1^\star, f_2^\star)$ morphological components, w noise.

Differente morphologies: sparse in different dictionaries Ψ_1, Ψ_2.

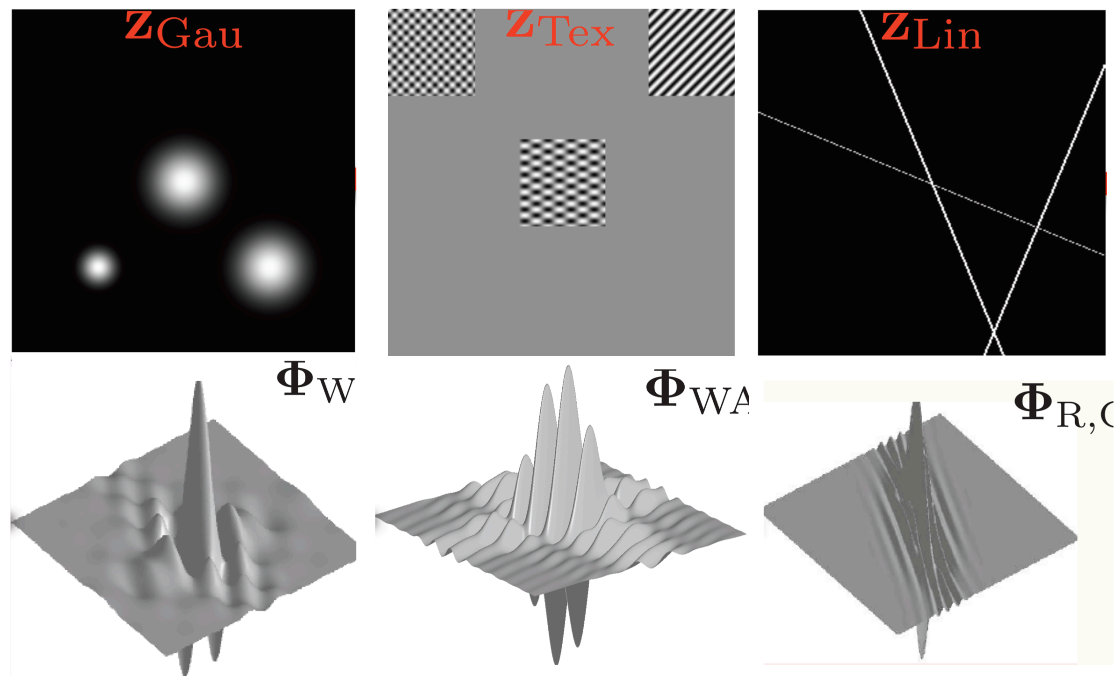
Isotropic: wavelet frame.

Texture: local cosine frame.

Lines: ridgelets frame.

Cartoon: curvelet frame.

Other: adaptivity (bandlets), learning, ...



Highly redundant union dictionary:

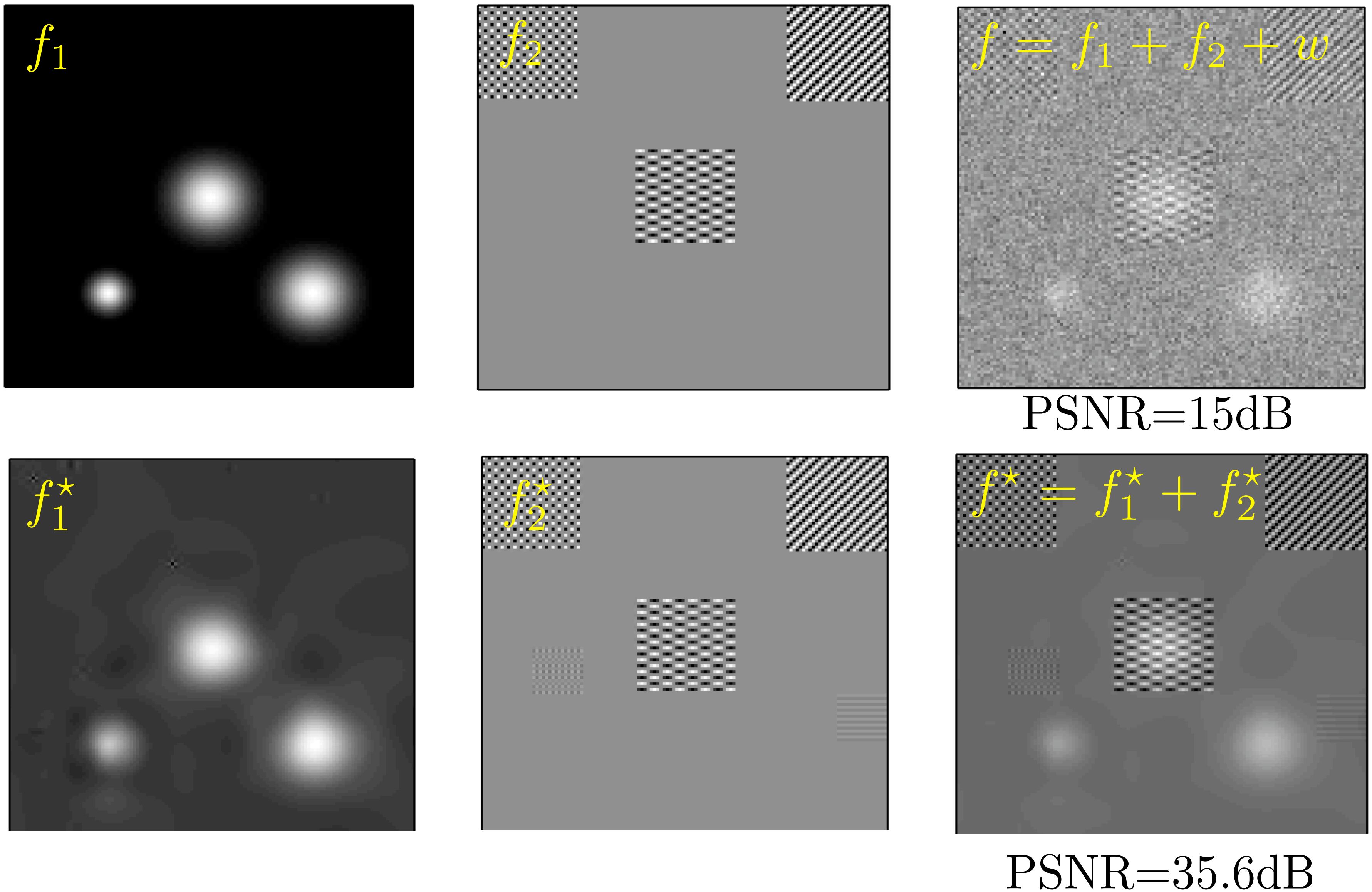
$$\Psi = [\Psi_1, \Psi_2] = \{\psi_m^1\}_{m=0}^{P_1-1} \cup \{\psi_m^2\}_{m=0}^{P_2-1}$$

$$(a_1^\star, a_2^\star) = \underset{(a_1,a_2)\in\mathbb{R}^P}{\operatorname{argmin}} \frac{1}{2}\|f - (\Psi_1 a_1 + \Psi_2 a_2)\|^2 + \lambda(\|a_1\|_1 + \|a_2\|_1)$$

Optimized components: $f_i^\star = \Psi_i a_i^\star$.

Examples of Decompositions

$$\Psi_1 = \{\text{redundant wavelet tight frame}\}$$
$$\Psi_2 = \{\text{redundant local cosine frame}\}$$



Results from [Fadili, Starck], see also MCA method.

Cartoon+Texture Separation

Cartoon sketch: sparse in

$$\Psi_1 = \{\text{curvelet tight frame}\}$$

Oscillating textures: sparse in

$$\Psi_2 = \{\text{redundant local cosine frame}\}$$

See also: Meyer's model, [Aujol et al.]

Overview

- Inverse Problems Regularization
- Sparse Prior in Orthogonal Bases
- Sparse Regularization in Orthogonal Bases
- Examples of applications:
  - *Sparse Diracs*: Seismic Deconvolution
  - *Sparse in 1D Wavelets*: Deblurring
  - *Sparse in 2D Wavelets*: Inpainting
- Redundant Representations
- Sparse Regularization in Redundant Dictionaries

Sparse Synthesis Regularization

Inverse problem: $\quad y = \Phi f_0 + w.$

Sparse synthesis regularization: looking for $f^\star = \Psi a^\star$ where

$$a^\star = \underset{a \in \mathbb{R}^P}{\operatorname{argmin}} \frac{1}{2}\|y - \Phi\Psi a\|^2 + \lambda \|a\|_1$$

$\longrightarrow$ sparse ℓ^1 approximation of y in the dictionary $\{\Phi\psi_m\}_m$.

Sparse Synthesis Regularization

Inverse problem: $\quad y = \Phi f_0 + w.$

Sparse synthesis regularization: looking for $f^\star = \Psi a^\star$ where

$$a^\star = \underset{a \in \mathbb{R}^P}{\operatorname{argmin}} \frac{1}{2} \|y - \Phi \Psi a\|^2 + \lambda \|a\|_1$$

$\longrightarrow$ sparse ℓ^1 approximation of y in the dictionary $\{\Phi \psi_m\}_m$.

Iterative thresholding: $\quad a^{(k+1)} = S_{\lambda\tau}\left(a^{(k)} + \Psi^* \Phi^* (y - \Phi \Psi a^{(k)})\right)$

$$\tau \leqslant 2/\|\Psi^* \Phi^* \Phi \Psi\| \qquad a^{(k)} \longrightarrow a^\star$$

Sparse Analysis Regularization

Sparse synthesis regularization: looking for $f^\star = \Psi a^\star$ where

$$a^\star = \underset{a \in \mathbb{R}^P}{\operatorname{argmin}} \frac{1}{2} \|y - \Phi \Psi a\|^2 + \lambda \|a\|_1$$

$\longrightarrow$ $f^\star$ has a sparse approximation in the $\{\psi_m\}_m$.

Sparse analysis regularization:

$$\bar{f}^\star = \underset{f \in \mathbb{R}^N}{\operatorname{argmin}} \frac{1}{2} \|y - \Phi f\|^2 + \lambda \sum_m |\langle f,\, \psi_m \rangle| = \frac{1}{2} \|y - \Phi f\|^2 + \lambda \|\Psi^* f\|_1$$

$\longrightarrow$ The correlations of $\bar{f}^\star$ with $\{\psi_m\}_m$ are sparse.

If $\{\psi_m\}_m$ is orthogonal, $\bar{f}^\star = f^\star$.

Solving Sparse Analysis

Proximal iteration:

$$f^{(k+1)} = P_{\lambda\tau_1}(f^{[k]}) \quad \text{where} \quad f^{[k]} = f^{(k)} + \Phi^*(y - \Phi f^{(k)})$$

$$\text{where} \quad P_{\lambda\tau}(f^{[k]}) = \underset{f \in \mathbb{R}^N}{\text{argmin}} \, \frac{1}{2} \|f^{[k]} - f\|^2 + \lambda\tau_1 \sum_m |\langle f, \psi_m \rangle|$$

Theorem: $P_{\lambda\tau_1}(f^{[k]}) = f^{[k]} - \Psi a^\star$ where $a^\star = \underset{\|a\|_\infty \leqslant \lambda\tau_1}{\text{argmin}} \|f^{[k]} - \Psi a\|^2$

Solving Sparse Analysis

Proximal iteration:

$$f^{(k+1)} = P_{\lambda\tau_1}(f^{[k]}) \quad \text{where} \quad f^{[k]} = f^{(k)} + \Phi^*(y - \Phi f^{(k)})$$

$$\text{where} \quad P_{\lambda\tau}(f^{[k]}) = \underset{f \in \mathbb{R}^N}{\operatorname{argmin}} \frac{1}{2}\|f^{[k]} - f\|^2 + \lambda\tau_1 \sum_m |\langle f,\, \psi_m \rangle|$$

Theorem: $P_{\lambda\tau_1}(f^{[k]}) = f^{[k]} - \Psi a^\star$ where $a^\star = \underset{\|a\|_\infty \leqslant \lambda\tau_1}{\operatorname{argmin}} \|f^{[k]} - \Psi a\|^2$

Initialize $f^{(0)}$, set $\ell = 0$, repeat until convergence:

1) *Inverse problem descent:*

$$f^{[k]} = f^{(k)} + \lambda\tau_1\Phi^*(y - \Phi f^{(k)})$$

2) *Denoising:* Initialize $a^{(0)}$, set $\ell = 0$, repeat:

a) $\tilde{a}^{(\ell)} = a^{(\ell)} + \tau_2\Psi^*(f^{[k]} - \Psi a^{(\ell)})$

b) $a^{(\ell+1)}[m] = \dfrac{\tilde{a}^{(\ell)}[m]}{\max(|\tilde{a}^{(\ell)}[m]|, \lambda\tau_1)}$ $\qquad \tau_2 \leqslant \dfrac{2}{\|\Psi^*\Psi\|}$

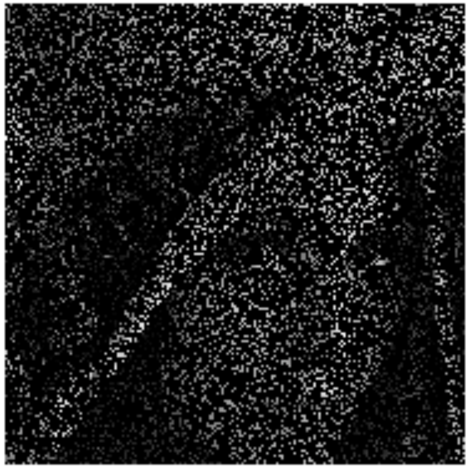
c) $\ell \leftarrow \ell + 1$

3) Set $f^{(k+1)} = f^{[k]} - \Psi a^{(\ell+1)}$, $k \leftarrow k + 1$

Conclusion

Linear data acquisition: ill-posed inverse problem $y = \Phi f_0 + w$.

Blurring, sub-sampling



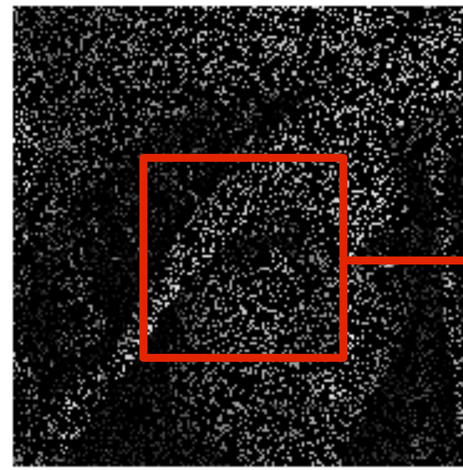
Missing pixels

Conclusion

Linear data acquisition: ill-posed inverse problem $y = \Phi f_0 + w$.

Regularization: linear Sobolev prior, non-linear sparse prior.

Blurring, sub-sampling



Missing pixels

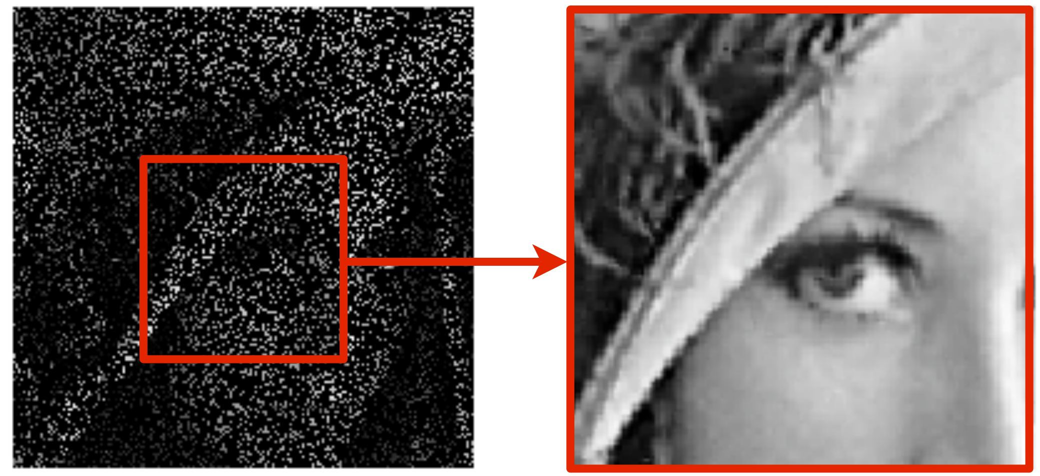
Inpainted

Conclusion

Linear data acquisition: ill-posed inverse problem $y = \Phi f_0 + w$.

Regularization: linear Sobolev prior, non-linear sparse prior.

Blurring, sub-sampling



Missing pixels

Inpainted

Sparse decomposition: redundant dictionary, image separation.