Inverse Problems

Regularization

Gabriel Peyré

CEREMADE

DAUPHINE UNIVERSITÉ PARIS

Overview

- Variational Priors
- Partial Differential Equation Flow
- PDE Flow and Regularization Denoising
- Inverse Problems Variational Regularization
- Tomography Inversion

Smooth and Cartoon Priors

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

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

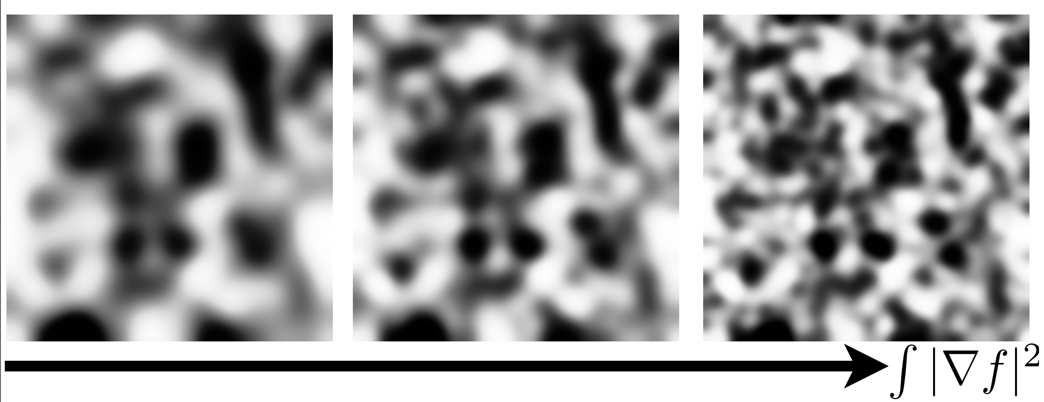
(real norm: $\mu\|f\| + \|f\|_{\text{Sob}}$)

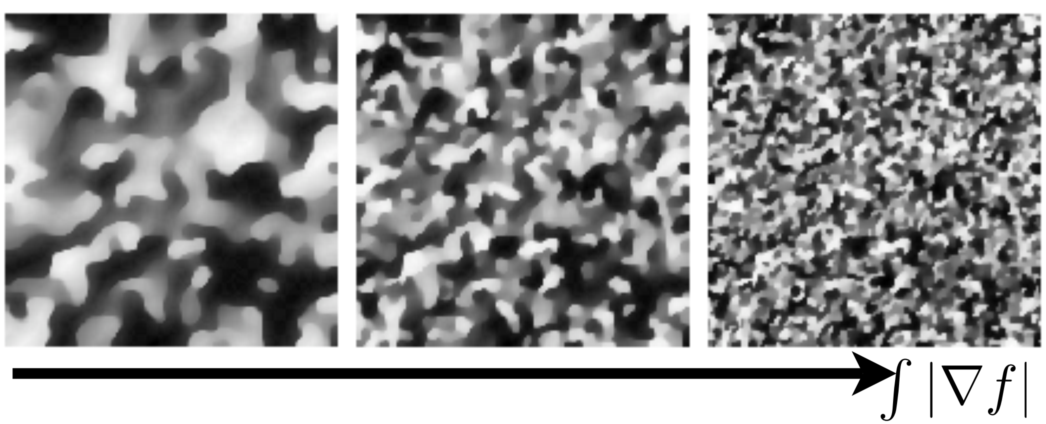
Total variation norm: $$J(f) = \|f\|_{\text{TV}} = \int \|\nabla_x f\| \mathrm{d}x$$

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

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

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

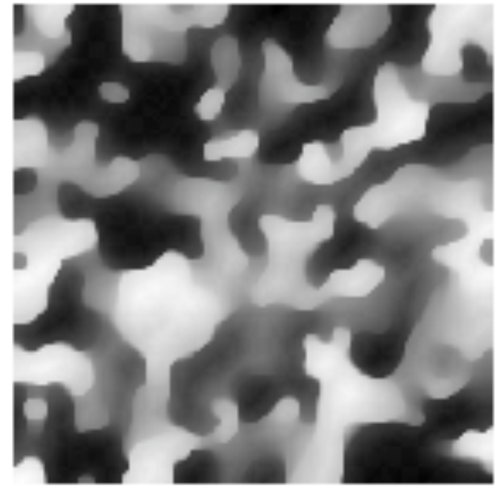




Natural Image Priors

"Typical" image drawn at random:

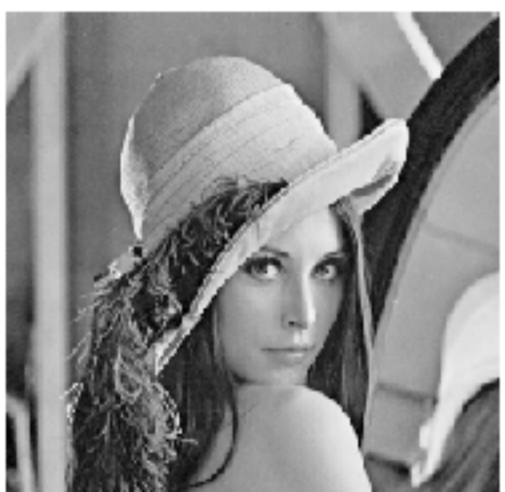
Small $\|f\|_{\mathrm{Sob}}$



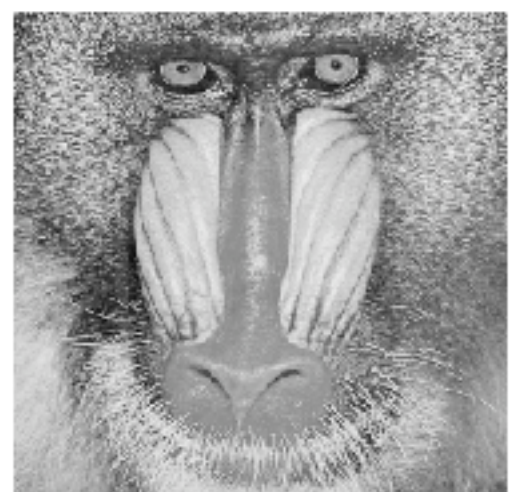
Small $\|f\|_{\mathrm{TV}}$

Natural images: structure + texture + noise + $\ldots$

TV=3988



TV=9387



Discrete Priors

Analog signal $f \in \mathrm{L}^2([0,1]^2) \longrightarrow$ discrete signal $f \in \mathbb{R}^N$.

Finite differences operators:

$$\delta_1 f[n_1, n_2] = f[n_1 + 1, n_2] - f[n_1, n_2]$$

$$\delta_2 f[n_1, n_2] = f[n_1, n_2 + 1] - f[n_1, n_2]$$

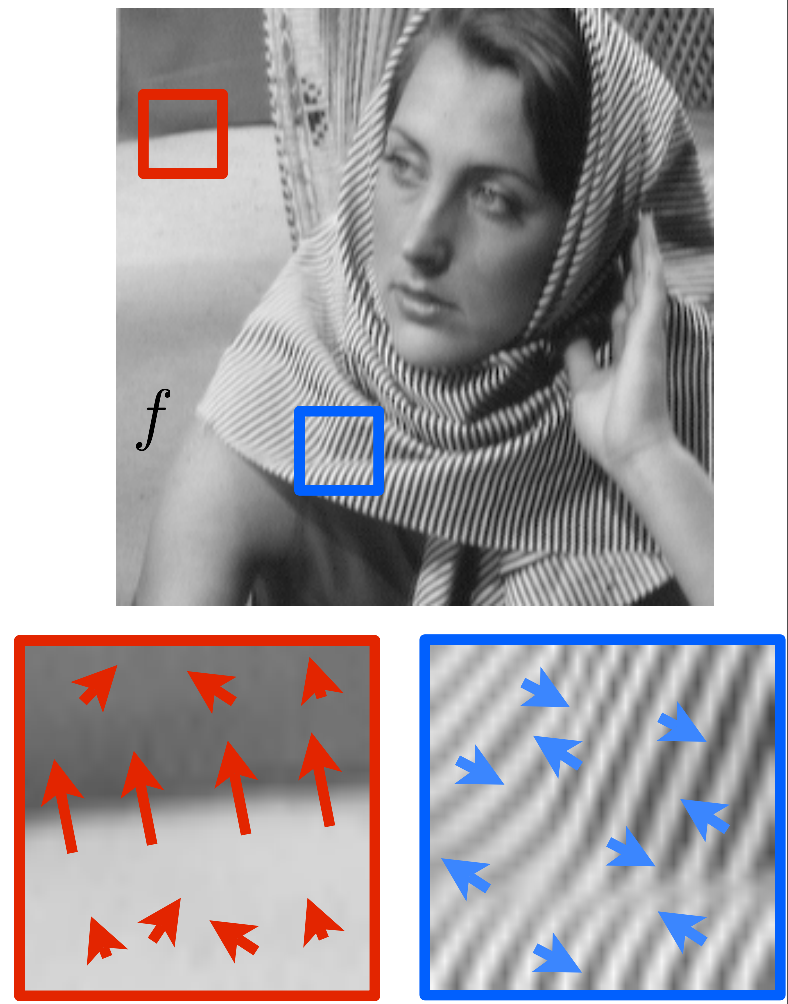
Discrete gradient:

$$\nabla f[n] = (\delta_1 f[n], \delta_2 f[n]) \in \mathbb{R}^{2 \times N}$$

Discrete energies:

$$J_{\mathrm{Sob}}(f) = \frac{1}{2} \sum_n (\delta_1 f[n])^2 + (\delta_2 f[n])^2$$

$$J_{\mathrm{TV}}(f) = \sum_n \sqrt{(\delta_1 f[n])^2 + (\delta_2 f[n])^2}$$



∇f ∇f

Discrete Differential Operators

Forward differences: $\delta_1 f[n_1, n_2] = f[n_1, n_2] - f[n_1 - 1, n_2]$

(Periodic boundary conditions)

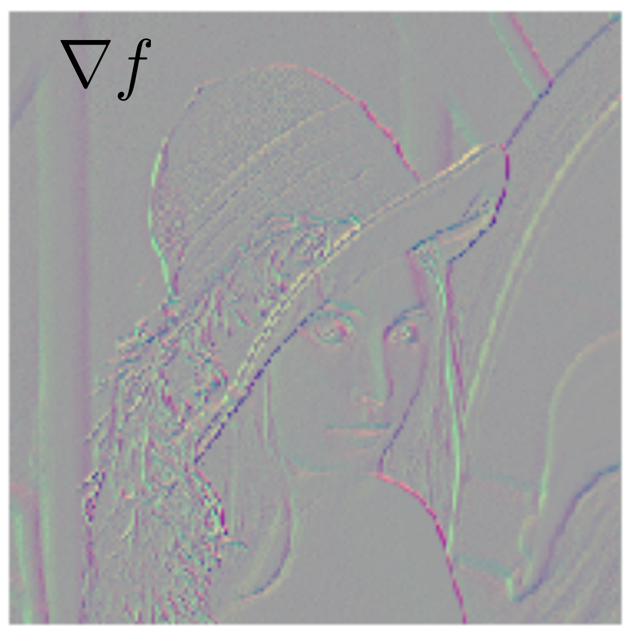
Backward differences (adjoint): $\tilde{\delta}_1 f[n_1, n_2] = f[n_1 + 1, n_2] - f[n_1, n_2]$

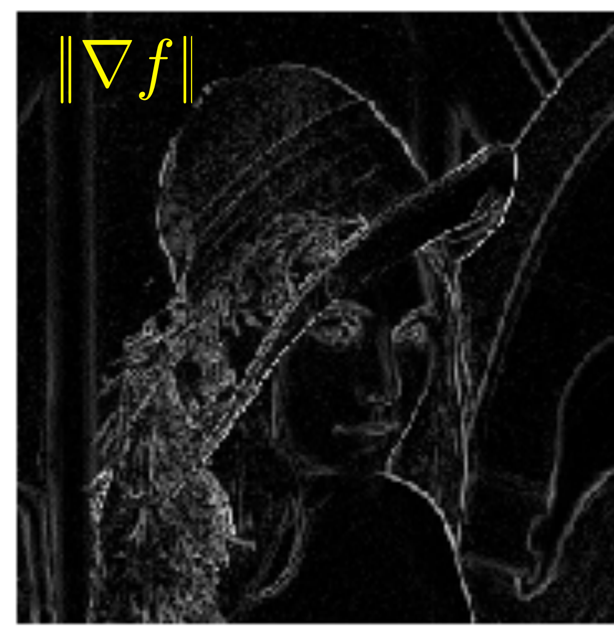
Adjoint: $\delta_1^* = -\tilde{\delta}_1$.

Gradient operator: $\nabla f[n] = (\delta_1 f[n], \delta_2 f[n]) \in \mathbb{R}^{2\times N}$

Divergence operator: $\mathrm{div}(v) = \tilde{\delta}_1 v_1[n] + \tilde{\delta}_2 v_2[n]$

$\nabla : \mathbb{R}^N \longrightarrow \mathbb{R}^{2\times N}$ $\xleftrightarrow{\mathrm{div} = -\nabla^*}$ div: $\mathrm{R}^{2\times N} \longrightarrow \mathbb{R}^N$





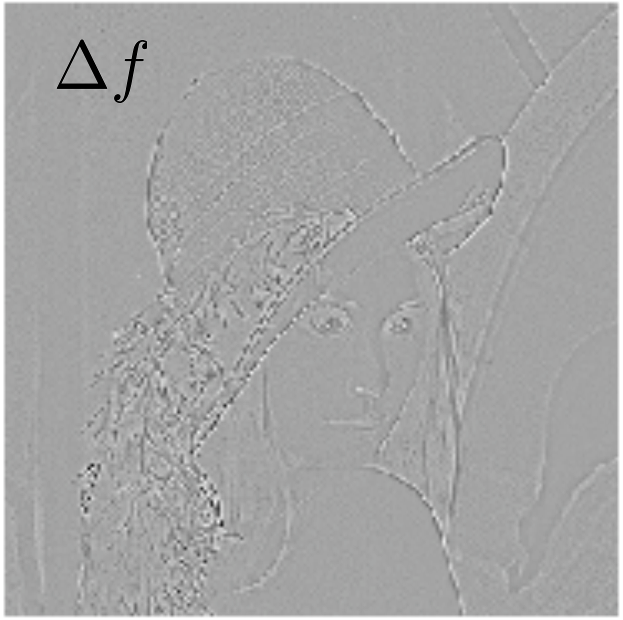
Laplacian Operator

Laplacian : $\Delta f = \mathrm{div}(\nabla f)$. $\quad \Delta f[n] = \sum_{p \in V_4(n)} f[p] - 4f[n]$

$$\frac{\partial^2 f}{\partial x_1^2}(x) + \frac{\partial^2 f}{\partial x_2^2} \approx N^2 \Delta f[n] \quad \text{for} \quad x = n/N.$$

Sobolev energy: $\quad J(f) = \sum_n \|\nabla f[n]\|^2 = \langle \nabla f, \nabla f \rangle = -\langle \Delta f, f \rangle$

Laplacian Δ semi-definite negative operator.



Overview

- Variational Priors
- **Partial Differential Equation Flow**
- PDE Flow and Regularization Denoising
- Inverse Problems Variational Regularization
- Tomography Inversion

Energy Flow and Gradient Descent

Discrete energy minimization: gradient descent

$$f^{(k+1)} = f^{(k)} - \tau \,\mathrm{grad}_{f^{(k)}} J \qquad\qquad f^{(0)} = f$$

Step size τ must be small enough.

Energy flow: $\quad t > 0 \longmapsto f_t \in \mathbb{R}^N$

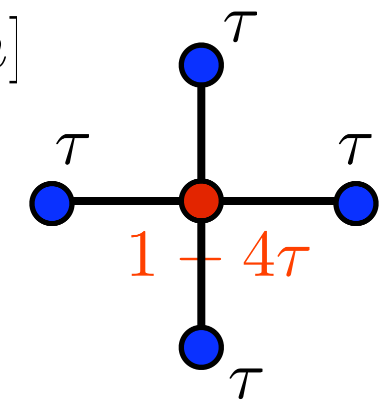
$$\frac{\partial f_t}{\partial t} = -\,\mathrm{grad}_{f_t} J$$

Discretization $f^{(k)}$ at time $t = k\tau$.

Heat Flow

Sobolev energy: $J(f) = \frac{1}{2} \sum_n \|\nabla f[n]\|^2$ $\quad$ $\text{grad}_f J = -\Delta f$

Discrete heat flow: $\frac{\partial f_t}{\partial t}[n] = -(\text{grad}_{f_t} J)[n] = \Delta f_t[n]$



Discrete times:

$$f^{(k+1)}[n] = f^{(k)}[n] + \tau \Big(\sum_{p \in V_4(n)} f[p] - 4f[n] \Big) = f \star h[n]$$

$$f^{(k)} = f \star h \star \ldots \star h = f \star^k h$$

Stable & convergent if $\tau < 1/4$.

Continuous Heat Flow

Sobolev energy: $$J(f) = \int \|\nabla_x f\|^2 \mathrm{d}x$$

Continuous heat flow: $$\frac{\partial f_t}{\partial t}(x) = -(\text{grad}_{f_t} J)(x) = \Delta f_t(x)$$

Explicit solution: $$f_t = f \star h_t \quad \text{where} \quad h_t(x) = \frac{1}{4\pi t} e^{-\frac{-\|x\|^2}{4t}}$$

Total Variation Gradient

TV energy: $J(f) = \|f\|_{\mathrm{TV}} = \sum_n \|\nabla f[n]\|.$

$\|\nabla f\|$

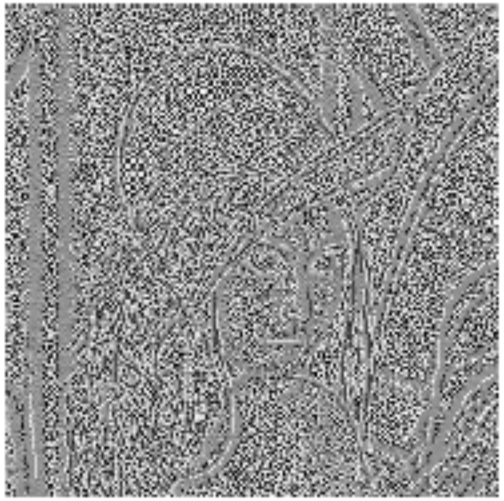
J non-differentiable at f such that $\exists\, n,\ \nabla f[n] = 0.$

If $\nabla f[n] \neq 0$, $\quad (\mathrm{grad}_f\, J)[n] = -\mathrm{div}\left(\frac{\nabla f}{\|\nabla f\|}\right)[n]$

Generalized (sub) gradient:

$\mathrm{grad}_f\, J$ not unique if $\nabla f[n] = 0$,

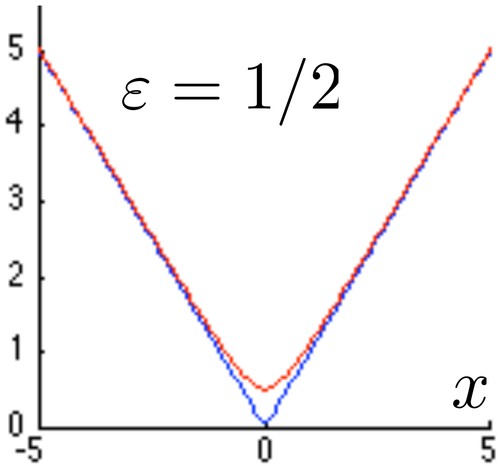
$g = \mathrm{div}(v) \in \mathrm{grad}_f\, J$ satisfies $\|v[n]\| \leqslant 1.$

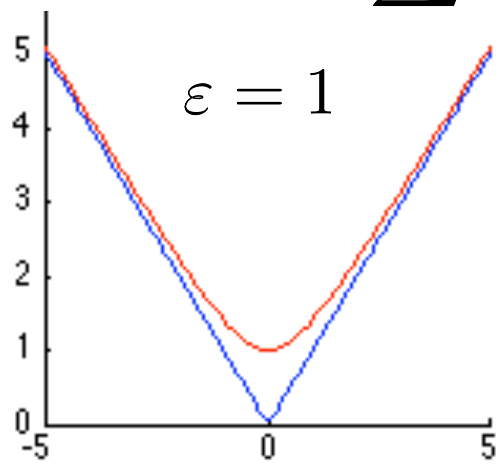


$\mathrm{div}\left(\frac{\nabla f}{\|\nabla f\|}\right)$

Regularized Total Variation

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



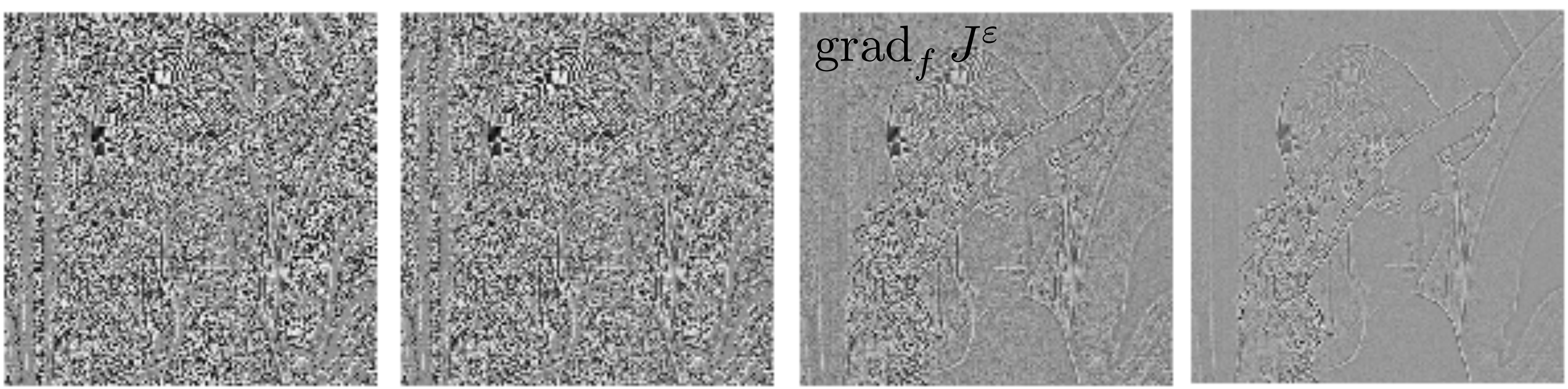


$x \mapsto |x|$

$x \mapsto \sqrt{\varepsilon^2 + x^2}$

Regularized gradient: $\text{grad}_f\, J^\varepsilon = -\text{div}\left(\frac{\nabla f}{\sqrt{\varepsilon^2 + \|\nabla f\|^2}} \right)$

$\text{grad}_f^\varepsilon \sim -\Delta/\varepsilon \quad \text{when} \quad \varepsilon \rightarrow +\infty$



$\varepsilon = 10^{-9}$ $\varepsilon = 10^{-2}$ $\varepsilon = 10^{-1}$ $\varepsilon = 1/2$

Total Variation Flow

Regularized TV-flow: $$\frac{\partial f_t}{\partial t} = \mathrm{div}\left(\frac{\nabla f_t}{\sqrt{\varepsilon^2 + \|\nabla f_t\|^2}}\right)$$

Small ε: $J^\varepsilon \approx J$ but instabilities.

TV flow smooth less the edges than heat diffusion.

$f_t \to$ constant when $t \to +\infty$.

Overview

- Variational Priors
- Partial Differential Equation Flow
- **PDE Flow and Regularization Denoising**
- Inverse Problems Variational Regularization
- Tomography Inversion

Gradient Flow for Denoising

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

Denoising using gradient flow: $\dfrac{\partial f_t}{\partial t} = -\operatorname{grad}_{f_t} J \qquad f_{t=0} = f$

Estimator $\tilde{f} = f_{t_0}$ for a well chose $t = t_0$.

$f_t \to$ constant when $t \to +\infty$.

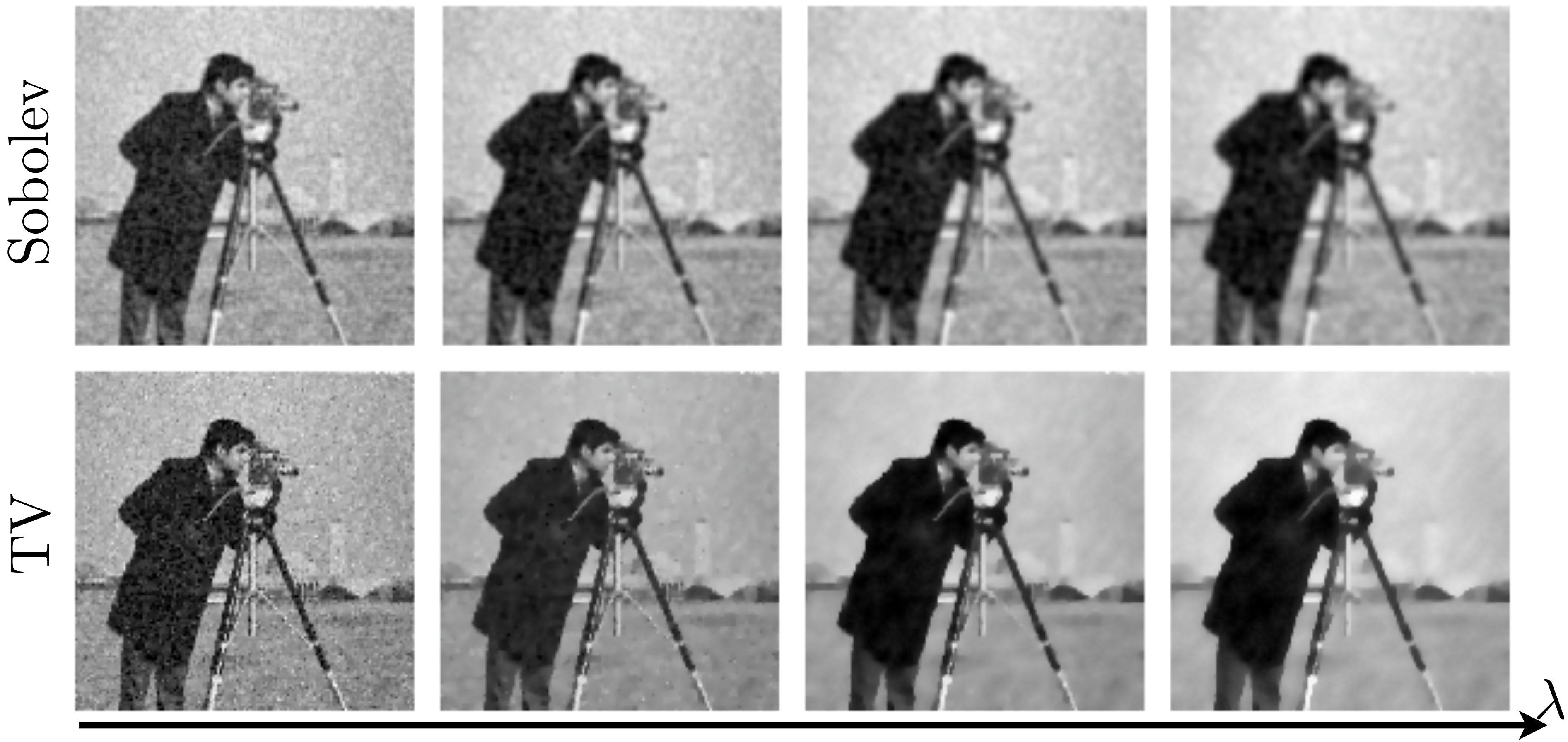
Regularization for Denoising

Denoising using regularization: $$f^\star_\lambda = \underset{g \in \mathbb{R}^N}{\operatorname{argmin}} \frac{1}{2}\|f - g\|^2 + \lambda J(g)$$

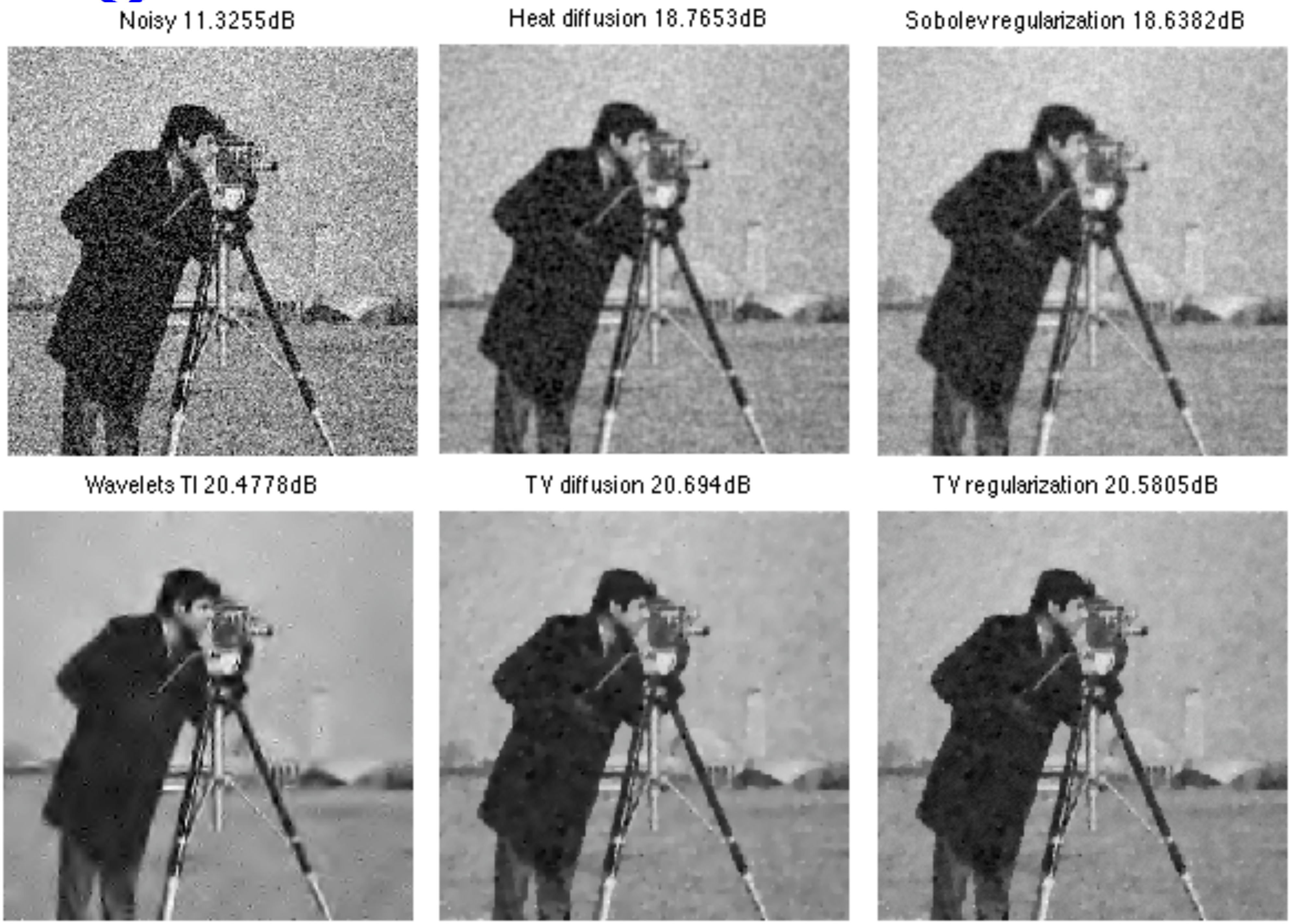
Estimator $\tilde{f} = f^\star_\lambda$ for a well chosen λ.

PDE flow: $$\frac{\partial f_t}{\partial t} = f_t - f - \lambda \operatorname{grad}_{f_t} J \qquad f_{t=0} \text{ arbitrary}$$

$f_t \to f^\star_\lambda$ when $t \to +\infty$.



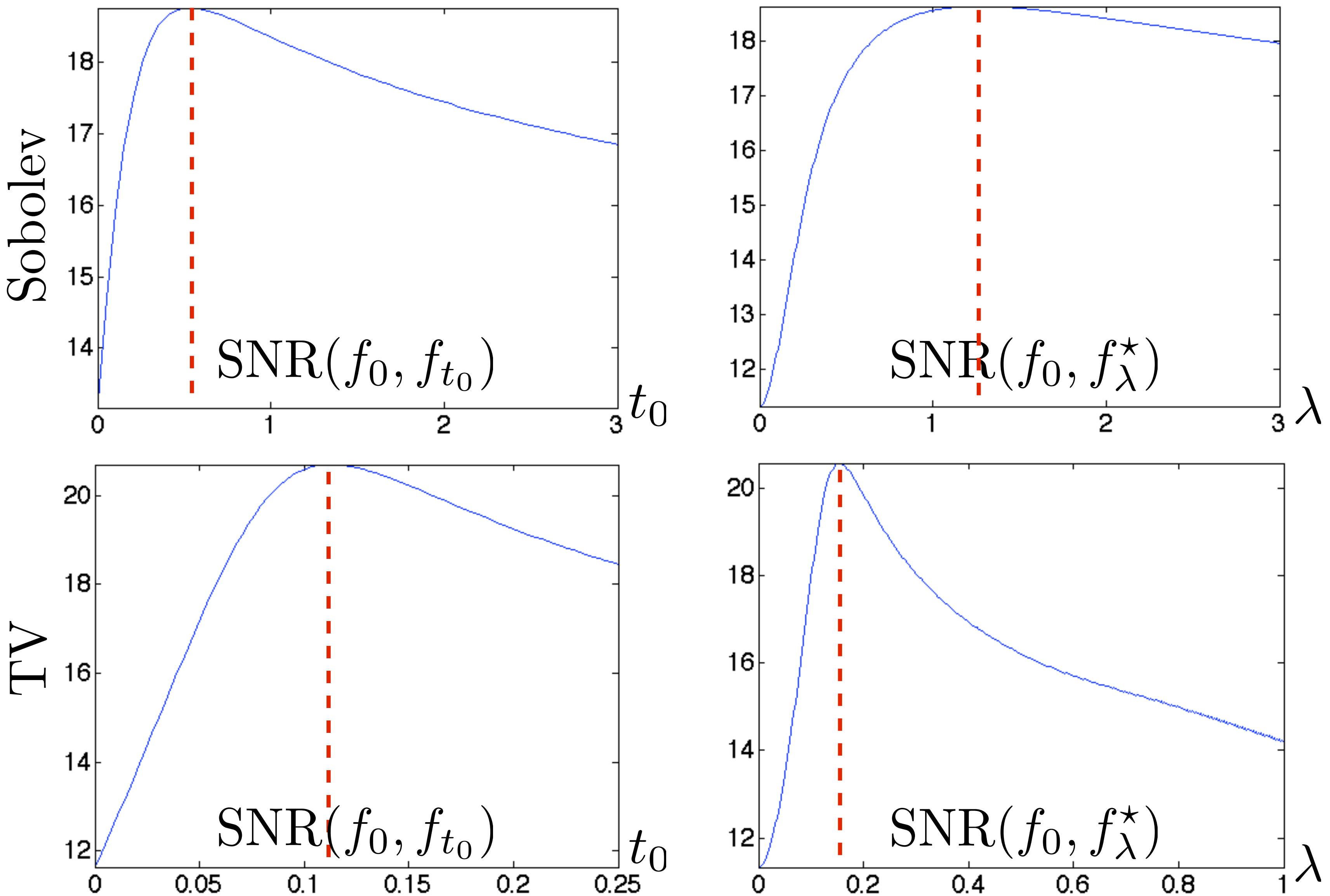
Comparison of the Methods



Optimal Parameter Selection

"Oracle" choice of t_0 or λ: minimize $\|\tilde{f} - f_0\|$.

$\longrightarrow$ t_0 and λ depends on both σ and the regularity of f_0 (e.g. $J(f_0)$).



Overview

- Variational Priors
- Partial Differential Equation Flow
- PDE Flow and Regularization Denoising
- **Inverse Problems Variational Regularization**
- Tomography Inversion

Inverse Problems

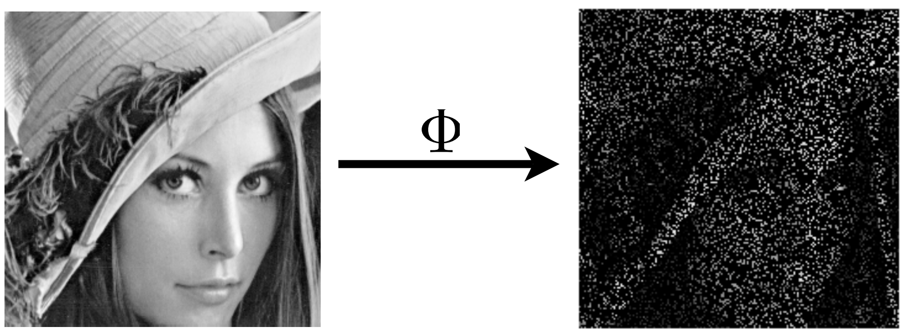
Recovering f_0 from P noisy measurements $y = \Phi f_0 + w \in \mathbb{R}^P$.

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

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

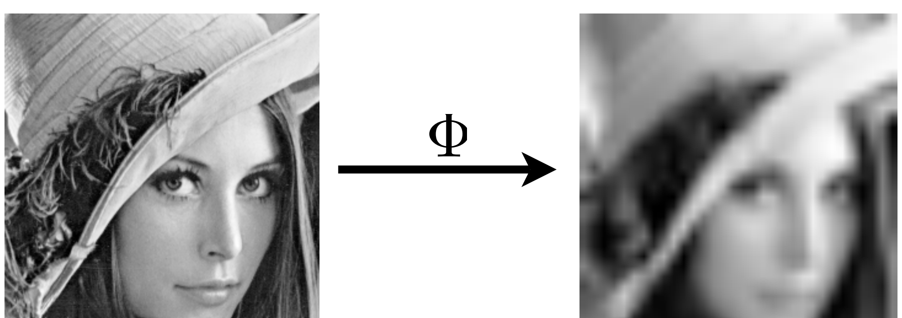
Denoising: $\Phi = \mathrm{Id}_N$, $P = N$.

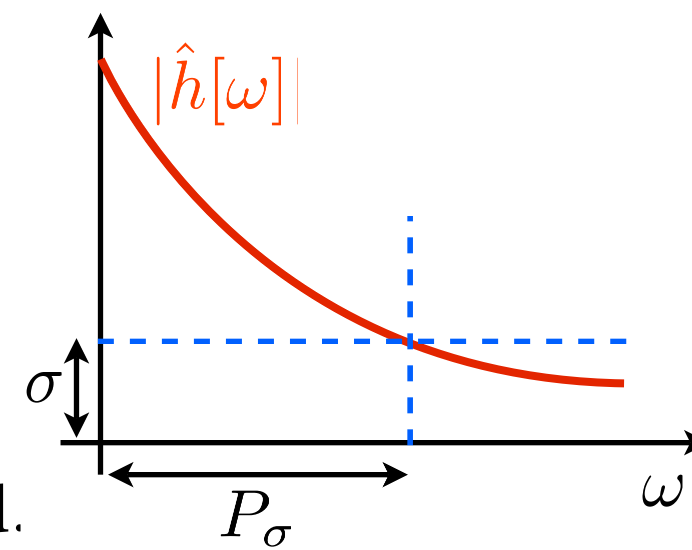
Inpainting: set $\Omega \subset \{0, \ldots, N-1\}$ of missing pixels, $P = N - |\Omega|$.



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

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





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

Inverse Problem Regularization

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

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

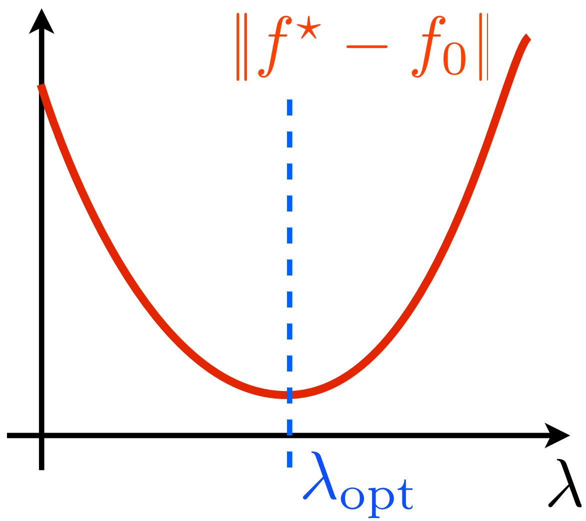
Regularized inverse: $$f^\star = \underset{f \in \mathbb{R}^N}{\operatorname{argmin}} \underbrace{\boxed{\frac{1}{2}\|y - \Phi f\|^2}}_{\text{Data fitting}} + \underbrace{\boxed{\lambda J(f)}}_{\text{Regularity}}$$

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

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

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

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



Pseudo-Inverse

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

Noise free: $f^\star = \underset{\Phi f = y}{\text{argmin}}\ \|f\| = \Phi^+ y \quad \text{where} \quad \Phi^+ = \Phi^*(\Phi\Phi^*)^{-1}$

Example: $y = \Phi f_0 = \varphi \star f_0$; $\quad \hat{f}^\star[\omega] = \begin{cases} \hat{y}[\omega]/\hat{\varphi}[\omega] & \text{if} \quad \hat{\varphi}[\omega] \neq 0, \\ 0 & \text{if} \quad \hat{\varphi}[\omega] = 0. \end{cases}$

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

$$f^\star = (\Phi^*\Phi + \lambda \mathrm{Id}_N)^{-1}\Phi^* y$$

Example: $y = \Phi f_0 + w = \varphi \star f_0 + w$, $\quad \hat{f}^\star[\omega] = \dfrac{\hat{\varphi}[\omega]^*}{|\varphi[\omega]|^2 + \lambda}\hat{y}[\omega]$

General case: use conjugate gradient.

Sobolev Regularization

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

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

Quadratic regularization:

$$f^\star = (\Phi^* \Phi - \lambda \Delta)^{-1} \Phi^* y \qquad\qquad \hat{f}^\star[\omega] = \frac{\hat{\varphi}[\omega]^*}{|\varphi[\omega]|^2 - \lambda \hat{\Delta}[\omega]} \hat{y}[\omega]$$

Example: $y = \Phi f_0 + w = \varphi \star f_0 + w$,

Continuous: $\hat{\Delta}(\omega) = -\|\omega\|^2$,

Discrete: $\hat{\Delta}[\omega] = -\sin\left(\frac{\omega_1 \pi}{N_1}\right)^2 - \sin\left(\frac{\omega_2 \pi}{N_2}\right)^2$

General case: gradient descent:

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

Convergence: $\tau < 2 / \|\Phi^* \Phi - \lambda \Delta\|$

where $\quad \|A\| = \max_{\|x\|=1} \|Ax\| = \lambda_{\max}(A)$

Sobolev Inpainting

Noiseless measurements: $$y(x) = \begin{cases} 0 & \text{if} \quad x \in \Omega, \\ f_0(x) & \text{if} \quad x \notin \Omega. \end{cases}$$

Orthogonal projector $\Phi^* = \Phi$.

Noiseless $w = 0$: $$f^\star = \underset{\Phi f = y}{\operatorname{argmin}} \|\nabla f\|^2 = \sum_n \|\nabla f[n]\|^2$$

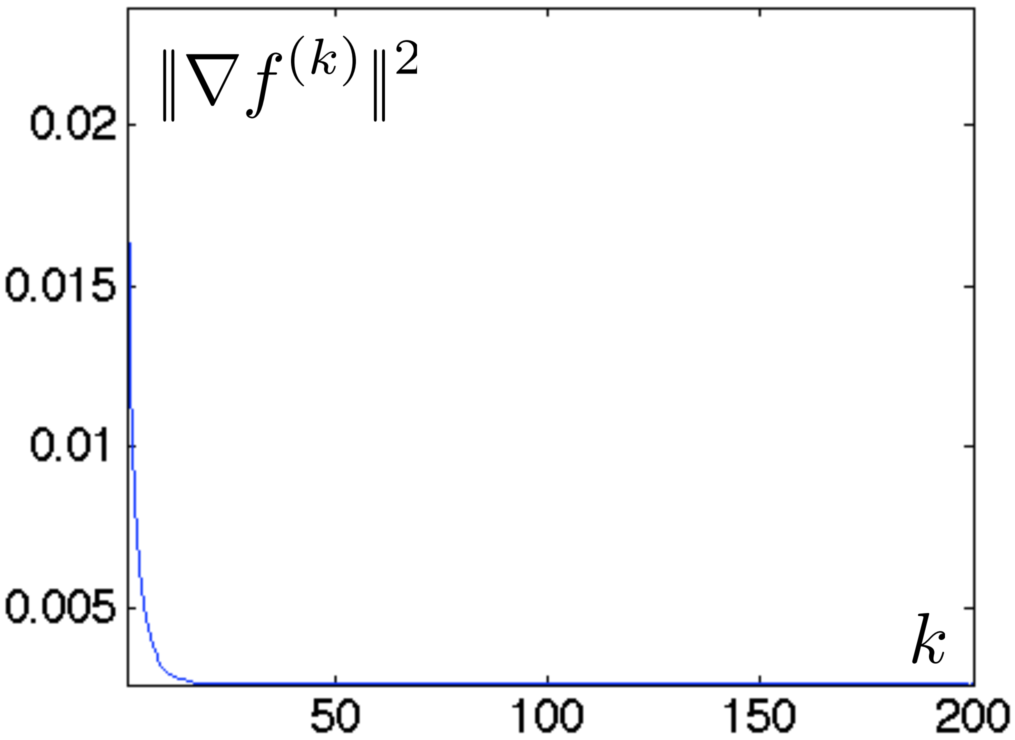
Projected gradient descent: $$f^{(k+1)} = \operatorname{Proj}_{\Phi f = y}\left(f^{(k)} + \tau \Delta f^{(k)} \right)$$

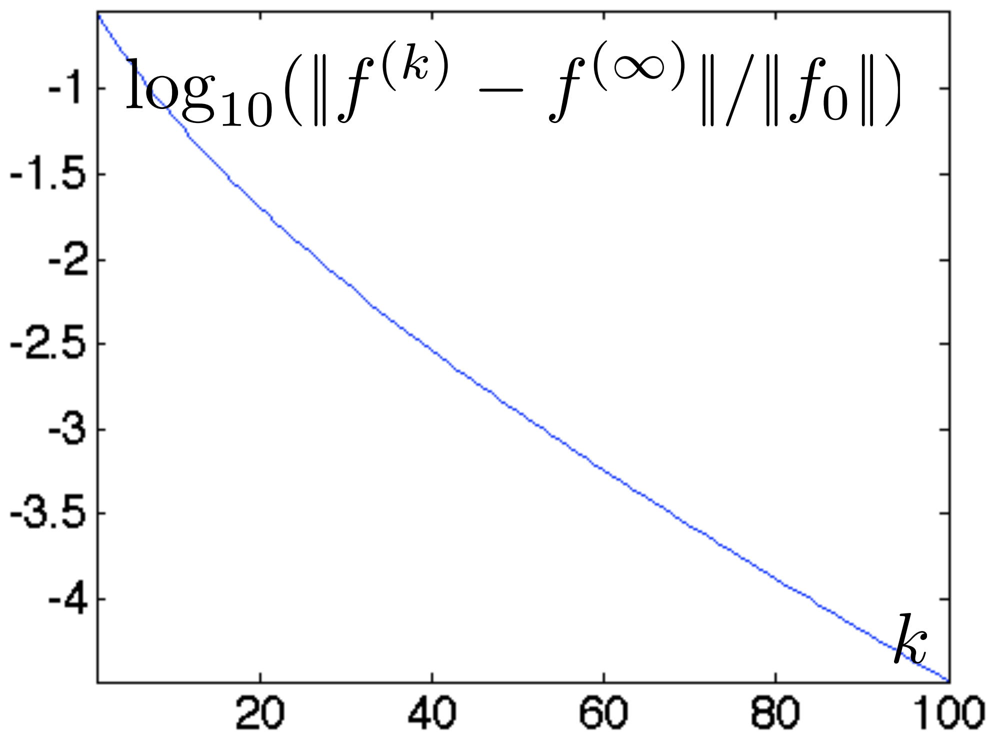
Convergence: $\tau < 2/\|\Delta\| = 1/4$.

Inpainting projected gradient descent:

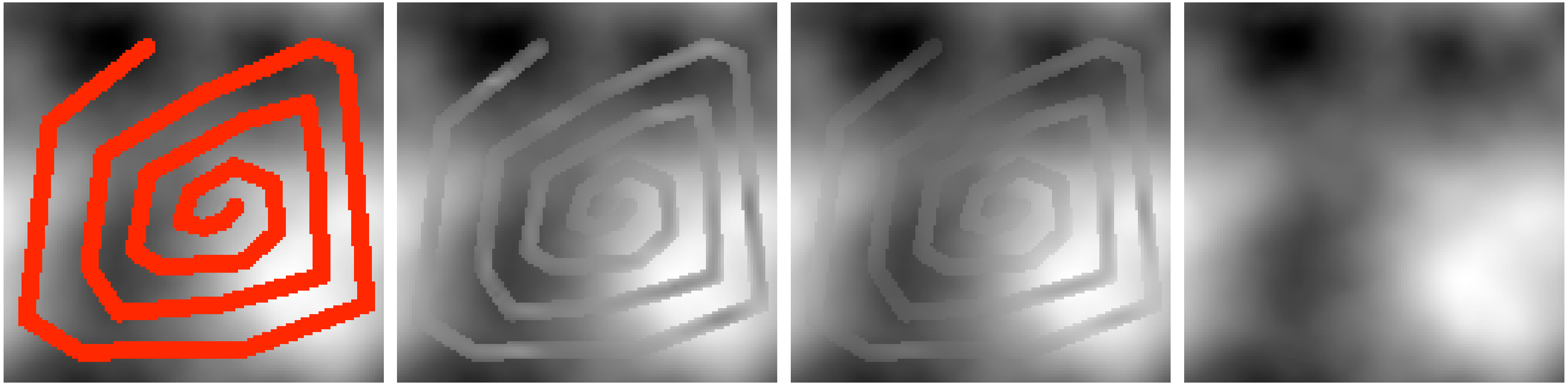
- Imposing values: $\tilde{f}^{(k)}(x) = \begin{cases} f^{(k)}(x) & \text{if} \quad x \in \Omega, \\ y(x) & \text{if} \quad x \notin \Omega. \end{cases}$
- Denoising: $f^{(k+1)} = \tilde{f}^{(k)} + \tau \Delta \tilde{f}^{(k)}$.

Example of Gradient Descent

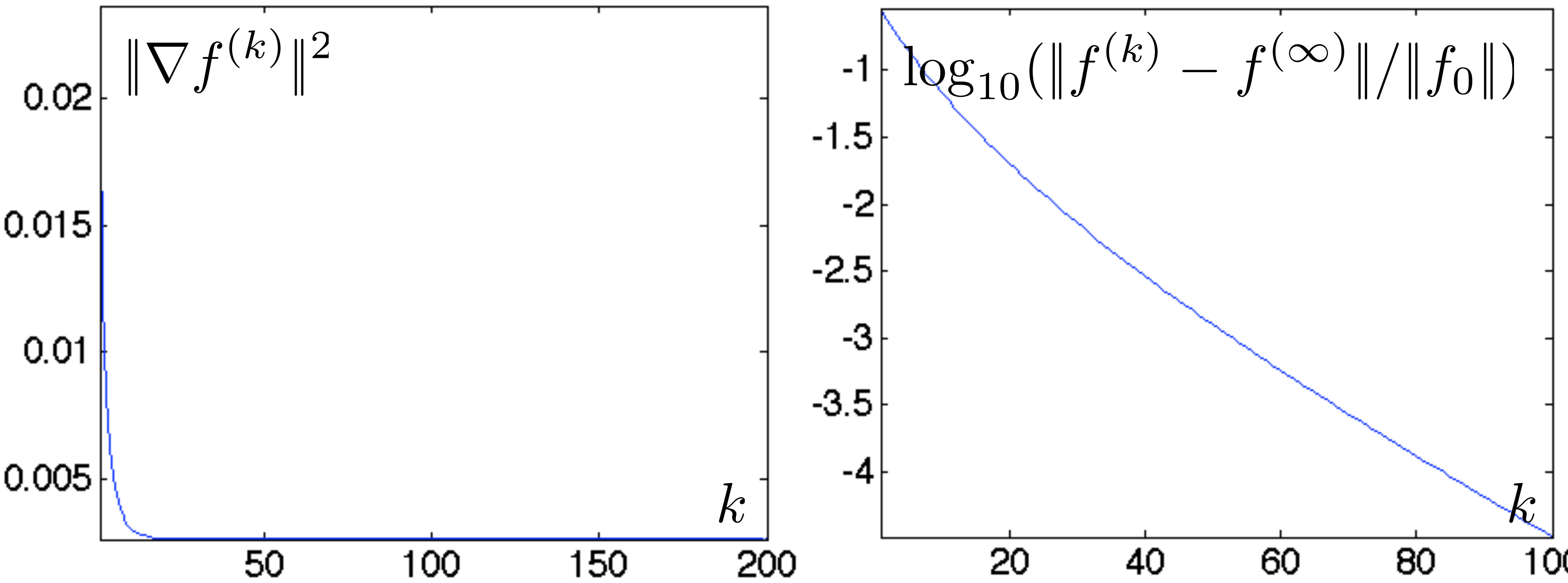




Example of Gradient Descent



Measurements y | Iter. #1 | Iter. #3 | Iter. #50



Example of Sobolev Inpainting

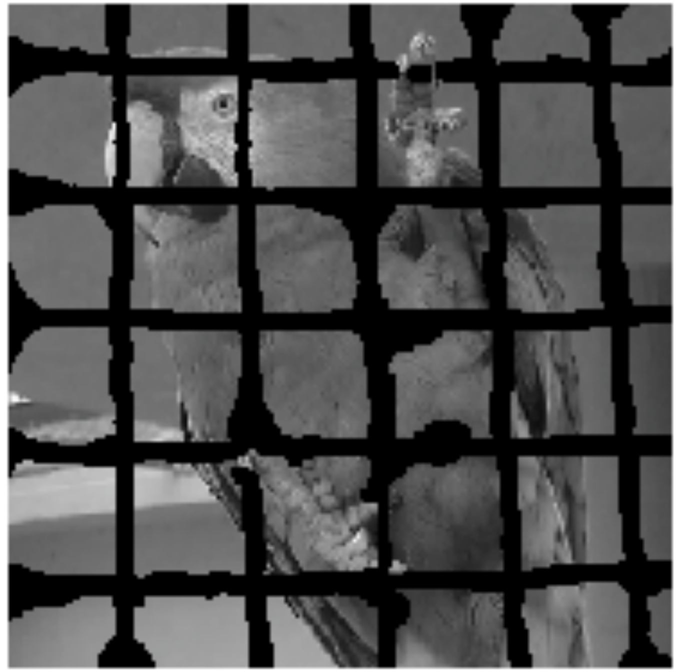

Original f_0


Observation $y = \Phi f_0$

Inpainted, SNR=18.2dB

Object Removal Inpaiting

Image f_0



Observations y

Inpainting $f^\star$

$f^\star$ is very different from f_0 !

L^2 error and SNR are not good measures of quality.

$\longrightarrow$ Pereceptual metric, visual inspection.

Total Variation Regularization

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

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

Remark: $f^\star$ not necessarily unique.

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

Convergence: τ should be small enough.

$\longrightarrow$ τ of the order of ε.

TV Inpainting

Noiseless, $w = 0$: $\quad f^\star = \underset{\Phi f = y}{\text{argmin}}\; J^\varepsilon(f)$

Projected gradient descent:

$$f^{(k+1)} = f^{(k)} - \tau\, \text{grad}_{f^{(k)}}\, J^\varepsilon$$

Convergence: τ should be small enough ($\sim \varepsilon$).

Example: inpainting with projected gradient descent:

- Imposing values: $\tilde{f}^{(k)}(x) = \begin{cases} f^{(k)}(x) & \text{if} \quad x \in \Omega, \\ y(x) & \text{if} \quad x \notin \Omega. \end{cases}$
- Denoising: $f^{(k+1)} = \tilde{f}^{(k)} + \tau \text{div}\left(\frac{\nabla \tilde{f}^{(k)}}{\sqrt{\varepsilon^2 + \|\nabla \tilde{f}^{(k)}\|^2}} \right)$

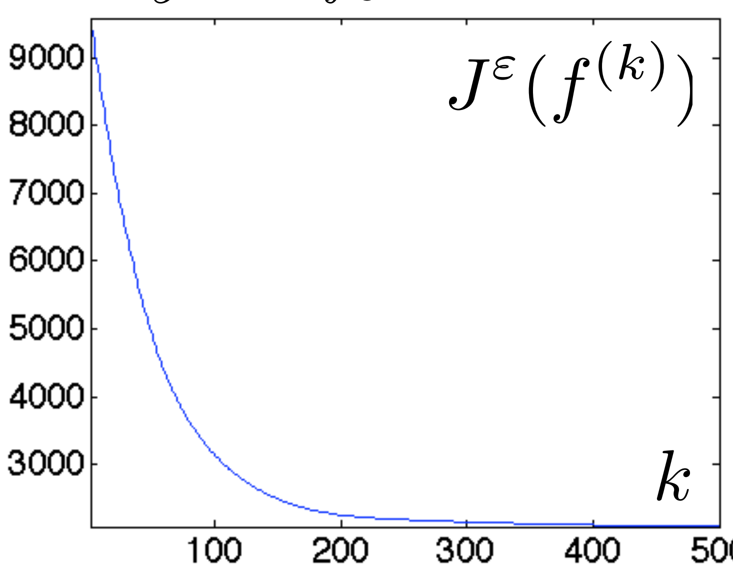
Inpainting: Sobolev vs. TV

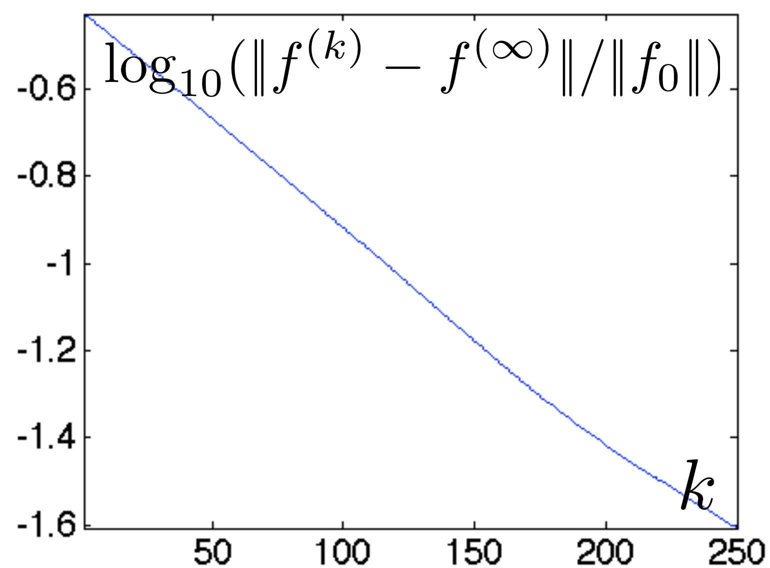
SNR: Sobolev > TV. Visual: Sobolev < TV.

Observation $y = \Phi f_0$

Sobolev, SNR=19.4dB

Sobolev, SNR=19dB





Overview

- Variational Priors
- Partial Differential Equation Flow
- PDE Flow and Regularization Denoising
- Inverse Problems Variational Regularization
- Tomography Inversion

Tomography and Radon Transform

Ray $\Delta_{t,\theta}$: $\quad x \cdot \tau_\theta = x_1 \cos\theta + x_2 \sin\theta = t$

Tomography projection:

$$\forall\, \theta \in [0,\pi), \forall\, t \in \mathbb{R}, \quad p_\theta(t) = \int_{\Delta_{t,\theta}} f(x)\, ds = \iint f(x)\, \delta(x \cdot \tau_\theta - t)\, dx$$

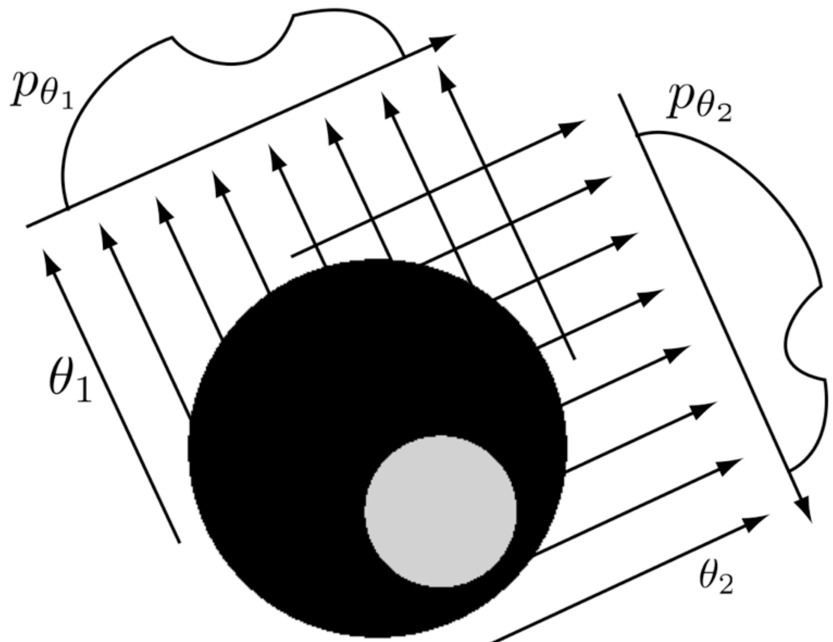
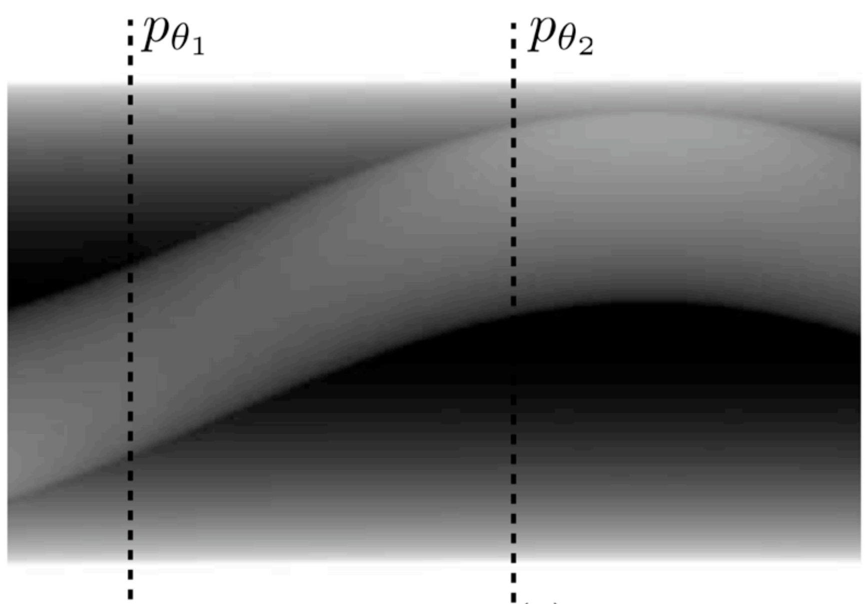


Image f



Radon $p_\theta(t)$

Partial Tomographic Measurements

Fourier slice theorem:

$$\forall\, \theta \in [0, \pi)\ ,\ \forall\, \xi \in \mathbb{R} \quad \hat{p}_\theta(\xi) = \hat{f}(\xi \cos\theta, \xi \sin\theta)$$

Back-projection (pseudo-inverse):

$$f(x) = \frac{1}{2\pi} \int_0^\pi p_\theta \star h(x \cdot \tau_\theta)\, d\theta \qquad \text{with} \quad \hat{h}(\xi) = |\xi|.$$

Limited number of orientations $\{\theta_k = \pi/k\}_{0 \leqslant k < K}$, $Rf = (p_{\theta_k})_{0 \leqslant k < K}$

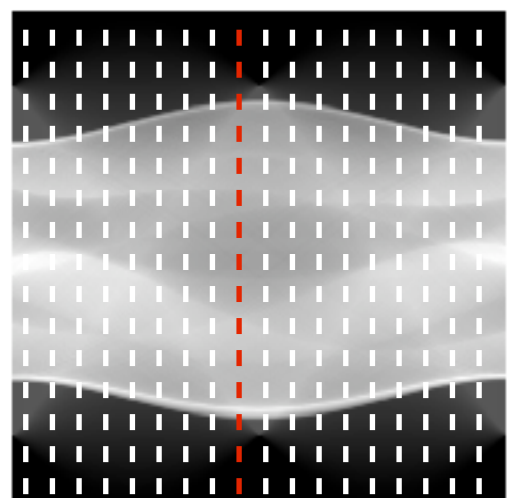
$$\text{Knowing } Rf \iff \text{knowing } \{\hat{f}(\xi \cos(\theta_k), \xi \sin(\theta k))\}_k$$

Simple discrete set-up:

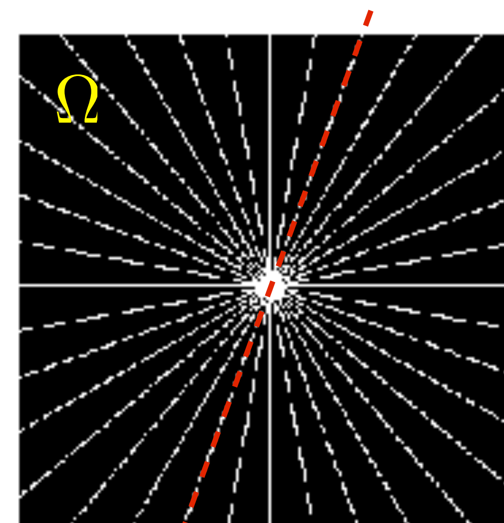
partial Fourier measurements

$$\Phi f = (\hat{f}[\omega])_{\omega \in \Omega} \in \mathbb{R}^P$$

$\rightarrow$ Inpainting missing frequencies



θ_k



θ_k

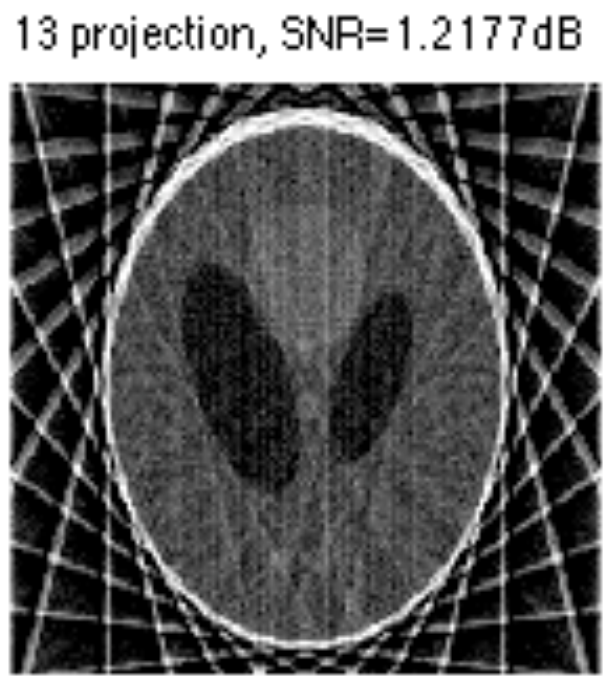
Radon Pseudo-Inverse

Noiseless measures: $\Phi f[\omega] = y[\omega] = \hat{f}[\omega]$ for $\omega \in \Omega$.

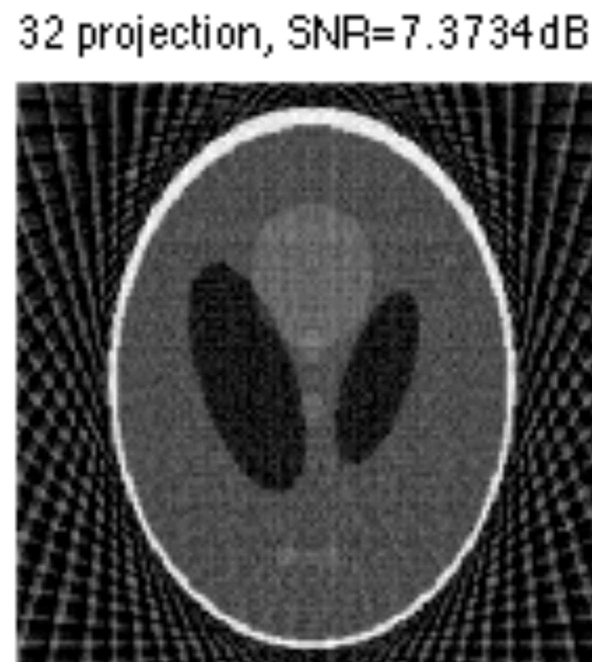
Peuso-inverse $f^+ = R^+ y$:

$$\hat{f}^+[\omega] = \begin{cases} y[\omega] & \text{if} \quad \omega \in \Omega, \\ 0 & \text{if} \quad \omega \notin \Omega. \end{cases}$$

Signal f



$\mathrm{R}^+ Rf$, $K = 13$



$\mathrm{R}^+ Rf$, $K = 32$

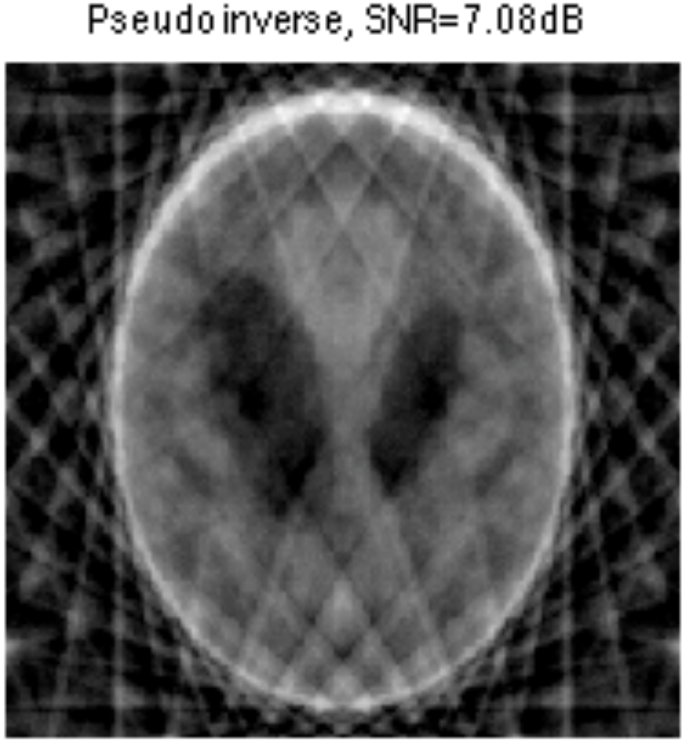
Regularized Inversion

Noisy measurements: $\forall\, \omega \in \Omega,\ y[\omega] = \hat{f}[\omega] + w[\omega]$.

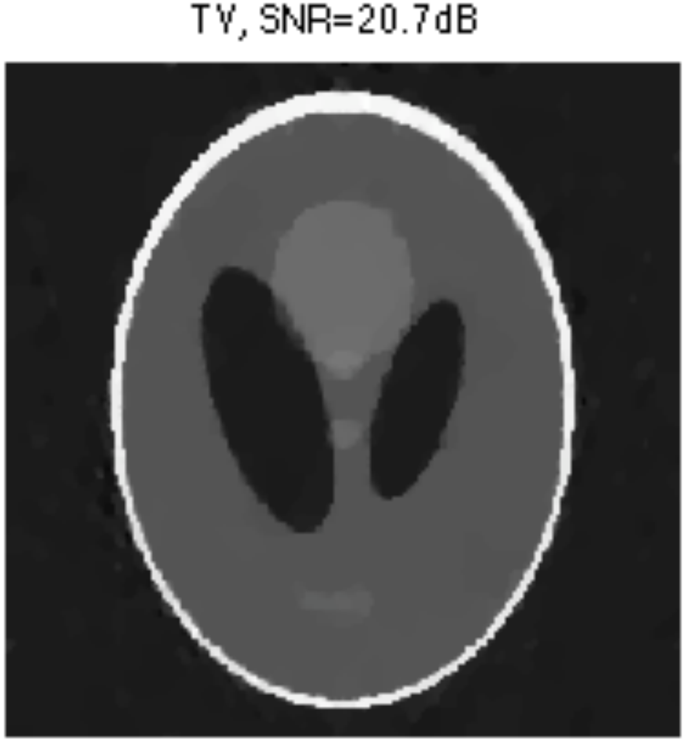
Noise $w[\omega] \sim \mathcal{N}(0, \sigma)$, white noise.

TV regularization:

$$f^+ = \underset{f}{\operatorname{argmin}}\ \lambda \|f\|_{\mathrm{TV}} + \frac{1}{2} \sum_{\omega \in \Omega} |y[\omega] - \hat{f}[\omega]|^2.$$



f^+



$f^\star$