Inverse Problems Regularization

Gabriel Peyré

www.numerical-tours.com

CEREMADE

DAUPHINE UNIVERSITÉ PARIS

Overview

- Variational Priors

- Gradient Descent and PDE's

- Inverse Problems Regularization

Smooth and Cartoon Priors

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

Sobolev semi-norm: $$J(f) = \|f\|_{W^{1,2}}^2 = \int_{\mathbb{R}^2} \|\nabla f(x)\| \mathrm{d}x$$

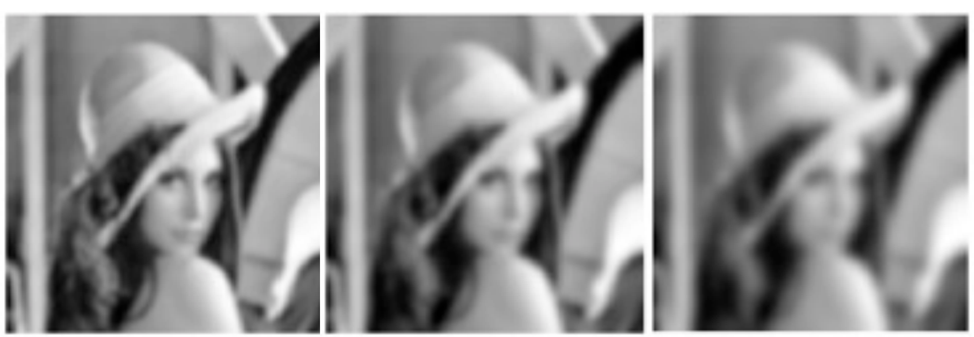
$\int |\nabla f|^2$

Smooth and Cartoon Priors

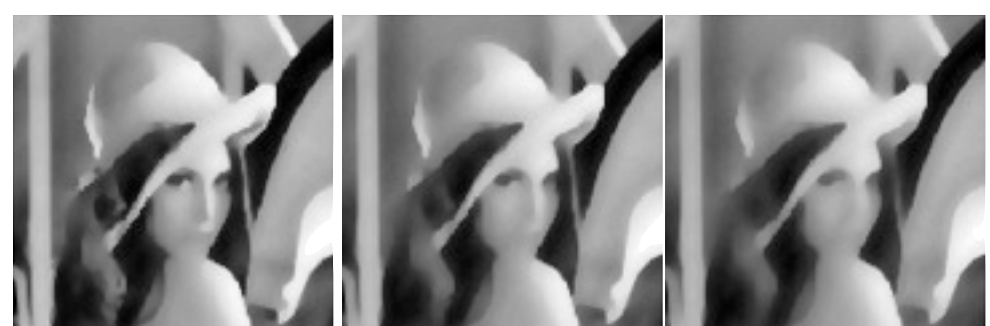
Prior model: energy $J(f) \in \mathbb{R}$ low for images of the model $f \in \Theta$.

Sobolev semi-norm: $$J(f) = \|f\|^2_{W^{1,2}} = \int_{\mathbb{R}^2} \|\nabla f(x)\| \mathrm{d}x$$

Total variation semi-norm: $$J(f) = \|f\|_{TV} = \int_{\mathbb{R}^2} \|\nabla f(x)\| \mathrm{d}x$$



$\int |\nabla f|^2$



$\int |\nabla f|$

Smooth and Cartoon Priors

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

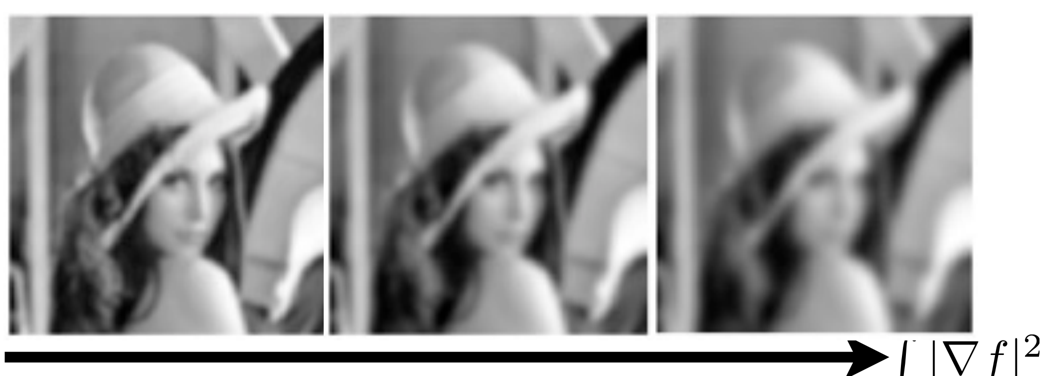
Sobolev semi-norm: $$J(f) = \|f\|^2_{W^{1,2}} = \int_{\mathbb{R}^2} \|\nabla f(x)\| \mathrm{d}x$$

Total variation semi-norm: $$J(f) = \|f\|_{TV} = \int_{\mathbb{R}^2} \|\nabla f(x)\| \mathrm{d}x$$

Co-area formula:

$$\|f\|_{\mathrm{TV}} = \int_{\mathbb{R}} \mathrm{length}(\mathcal{C}_t)\mathrm{d}t \qquad \text{Level set} \quad \mathcal{C}_t = \{x \setminus f(x) = t\}$$

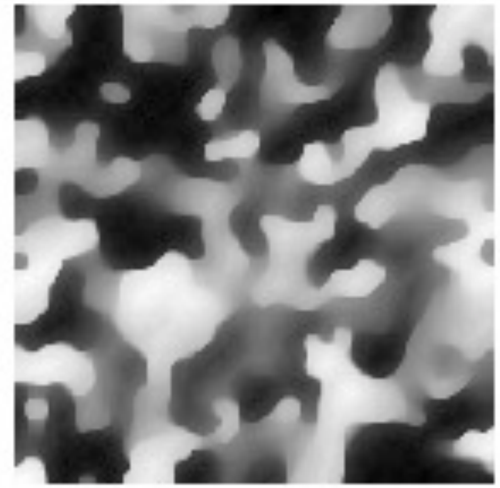
$\longrightarrow$ Extension to non-smooth functions $f \in \mathrm{BV}([0,1]^2)$



Natural Image Priors

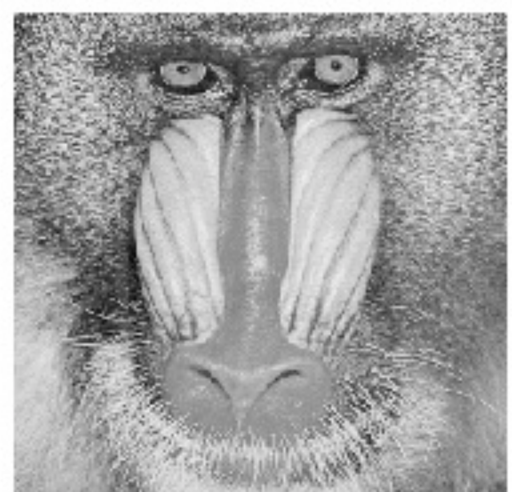
"Typical" image drawn at random:

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



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

Natural images: structure + texture + noise + ...



Discrete Priors

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

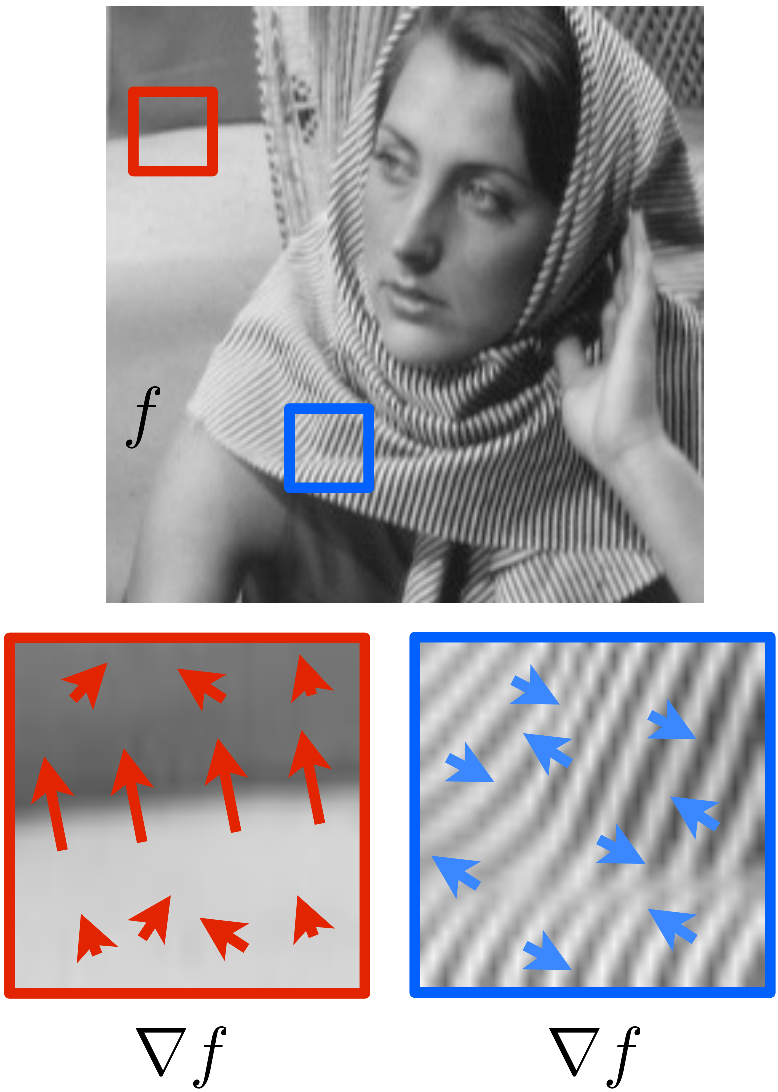
Finite differences operators:

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

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

Discrete gradient:

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



Discrete Priors

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

Finite differences operators:

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

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

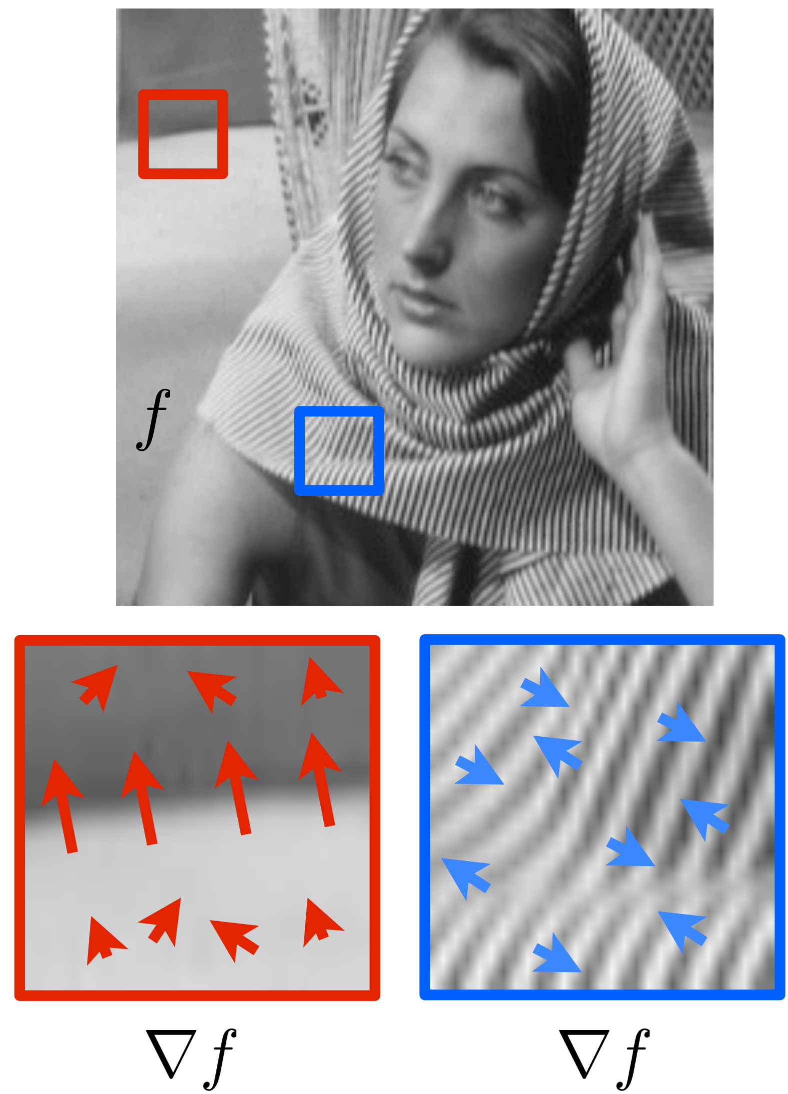
Discrete gradient:

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

Discrete energies:

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

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



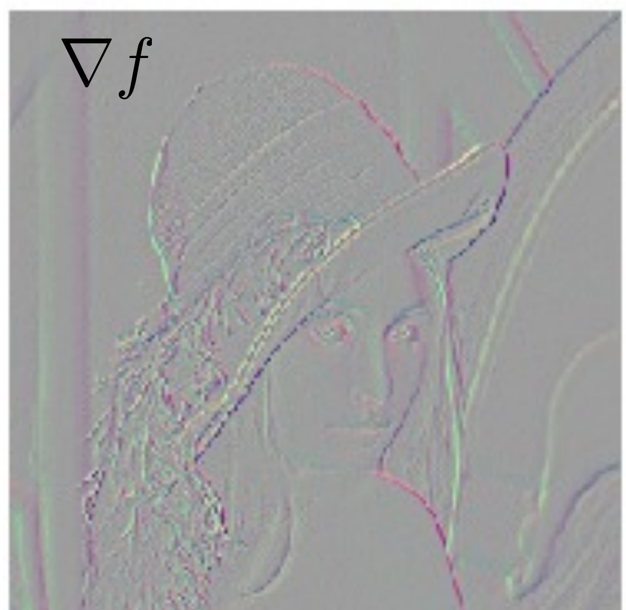
Discrete Differential Operators

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



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)

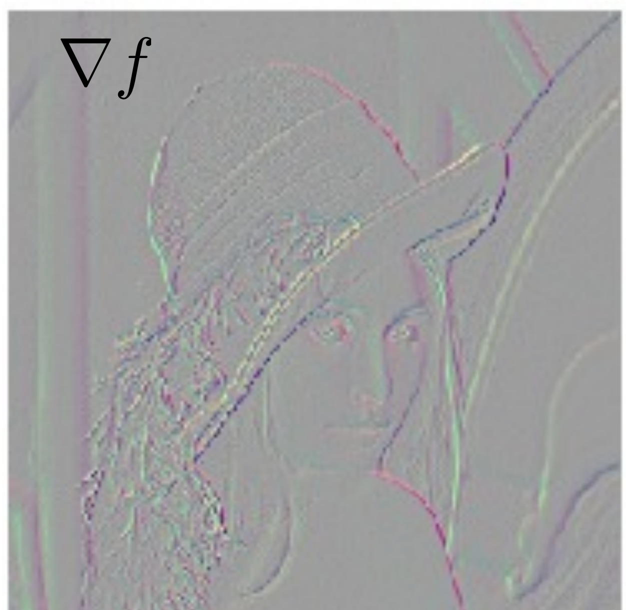
Backward differences (adjoint): $\tilde{\delta}_1 f[n_1, n_2] = f[n_1 + 1, n_2] - f[n_1, n_2]$

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

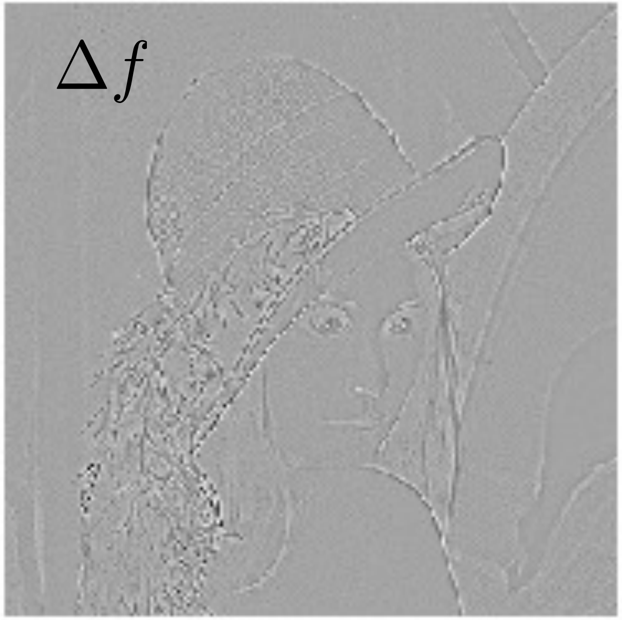
$$\nabla : \mathbb{R}^N \longrightarrow \mathbb{R}^{2\times N} \quad \xleftrightarrow{\operatorname{div} = -\nabla^*} \quad \operatorname{div}: \mathrm{R}^{2\times N} \longrightarrow \mathbb{R}^N$$



Laplacian Operator

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

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



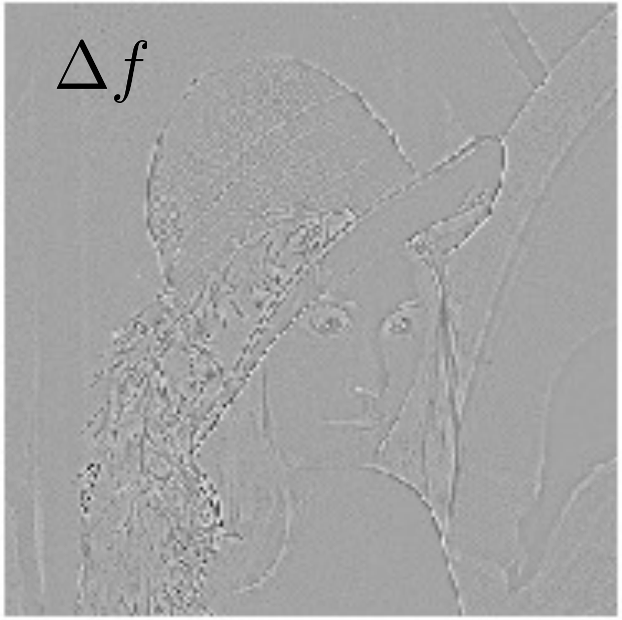
Laplacian Operator

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

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

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

Laplacian Δ semi-definite negative operator.



Gradient: Images vs. Functionals

Function: $\tilde{f} : x \in \mathbb{R}^2 \mapsto f(x) \in \mathbb{R}$

$$\nabla \tilde{f}(x) = (\partial_1 \tilde{f}(x), \partial_2 \tilde{f}(x)) \in \mathbb{R}^2$$

$$\tilde{f}(x + \varepsilon) = \tilde{f}(x) + \langle \nabla f(x), \, \varepsilon \rangle_{\mathbb{R}^2} + O(\|\varepsilon\|_{\mathbb{R}^2}^2)$$

Gradient: Images vs. Functionals

Function: $\tilde{f} : x \in \mathbb{R}^2 \mapsto f(x) \in \mathbb{R}$

$$\nabla \tilde{f}(x) = (\partial_1 \tilde{f}(x), \partial_2 \tilde{f}(x)) \in \mathbb{R}^2$$

$$\tilde{f}(x+\varepsilon) = \tilde{f}(x) + \langle \nabla f(x),\, \varepsilon \rangle_{\mathbb{R}^2} + O(\|\varepsilon\|^2_{\mathbb{R}^2})$$

Discrete image: $f \in \mathbb{R}^N, N = n^2$

$$f[i_1, i_2] = \tilde{f}(i_1/n, i_2/n) \qquad \nabla f[i] \approx \nabla \tilde{f}(i/n)$$

Gradient: Images vs. Functionals

$$\textit{Function:} \qquad \tilde{f} : x \in \mathbb{R}^2 \mapsto f(x) \in \mathbb{R}$$

$$\nabla \tilde{f}(x) = (\partial_1 \tilde{f}(x), \partial_2 \tilde{f}(x)) \in \mathbb{R}^2$$

$$\tilde{f}(x+\varepsilon) = \tilde{f}(x) + \langle \nabla f(x),\, \varepsilon \rangle_{\mathbb{R}^2} + O(\|\varepsilon\|^2_{\mathbb{R}^2})$$

$$\textit{Discrete image:} \qquad f \in \mathbb{R}^N, N = n^2$$

$$f[i_1, i_2] = \tilde{f}(i_1/n, i_2/n) \qquad \nabla f[i] \approx \nabla \tilde{f}(i/n)$$

$$\textit{Functional:} \qquad J : f \in \mathbb{R}^N \mapsto J(f) \in \mathbb{R}$$

$$J(f+\eta) = J(f) + \langle \nabla J(f),\, \eta \rangle_{\mathbb{R}^N} + O(\|\eta\|^2_{\mathbb{R}^N})$$

$$\nabla J : \mathbb{R}^N \mapsto \mathbb{R}^N$$

Gradient: Images vs. Functionals

Function: $\tilde{f} : x \in \mathbb{R}^2 \mapsto f(x) \in \mathbb{R}$

$$\nabla \tilde{f}(x) = (\partial_1 \tilde{f}(x), \partial_2 \tilde{f}(x)) \in \mathbb{R}^2$$

$$\tilde{f}(x+\varepsilon) = \tilde{f}(x) + \langle \nabla f(x),\, \varepsilon \rangle_{\mathbb{R}^2} + O(\|\varepsilon\|^2_{\mathbb{R}^2})$$

Discrete image: $f \in \mathbb{R}^N, N = n^2$

$$f[i_1, i_2] = \tilde{f}(i_1/n, i_2/n) \qquad \nabla f[i] \approx \nabla \tilde{f}(i/n)$$

Functional: $J : f \in \mathbb{R}^N \mapsto J(f) \in \mathbb{R}$

$$J(f+\eta) = J(f) + \langle \nabla J(f),\, \eta \rangle_{\mathbb{R}^N} + O(\|\eta\|^2_{\mathbb{R}^N})$$

$$\nabla J : \mathbb{R}^N \mapsto \mathbb{R}^N$$

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

$$\nabla J(f) = (\nabla^* \circ \nabla) f = -\Delta f$$

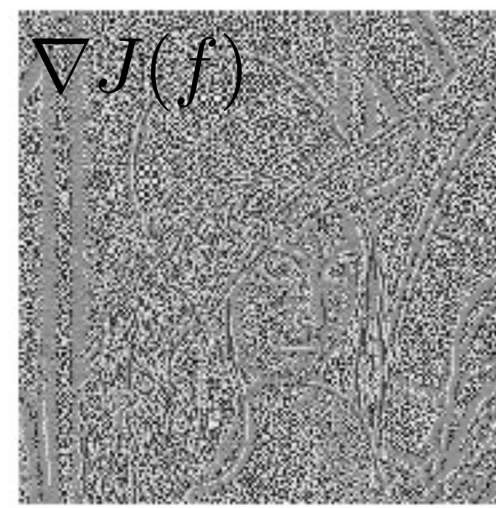
Total Variation Gradient

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

If $\exists n, \nabla f[n] = 0$, J not differentiable at f.

If $\forall\, n, \nabla f[n] \neq 0$,

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



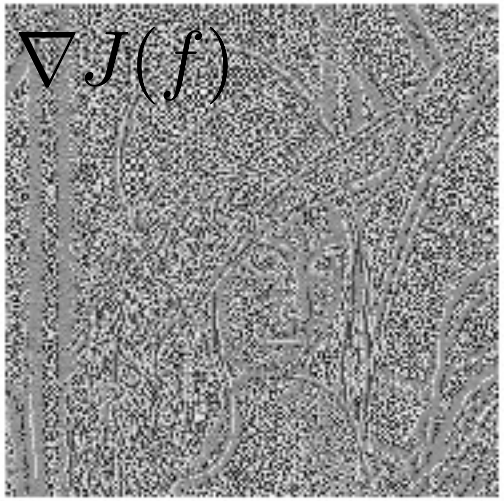
Total Variation Gradient

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

If $\exists n, \nabla f[n] = 0$, J not differentiable at f.

If $\forall\, n, \nabla f[n] \neq 0$,

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



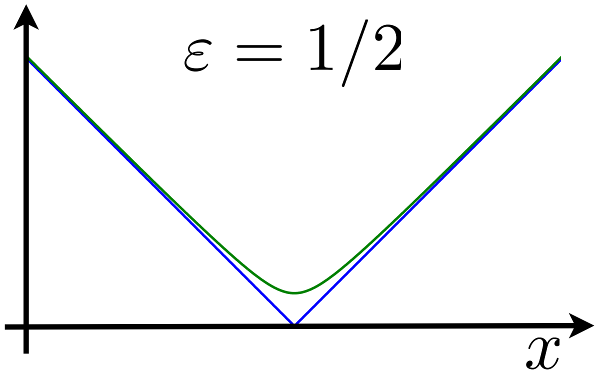
Sub-differential:

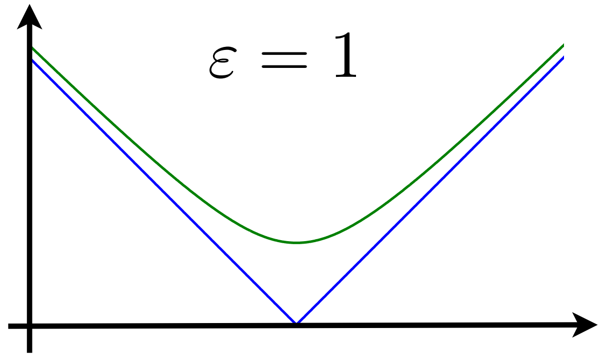
$$\partial J(f) = \{-\,\mathrm{div}(\alpha)\ ;\ \|\alpha[n]\| \leqslant 1 \quad \text{and} \quad \alpha \in \mathcal{C}_{\nabla f}\}$$

$$\mathcal{C}_u = \left\{\alpha \in \mathbb{R}^{2\times N} \setminus (u[n]=0) \Rightarrow (\alpha[n] = u[n]/\|u[n]\|)\right\}$$

Regularized Total Variation

$$\|u\|_\varepsilon = \sqrt{\|u\|^2 + \varepsilon^2} \qquad J_\varepsilon(f) = \sum_n \|\nabla f[n]\|_\varepsilon$$



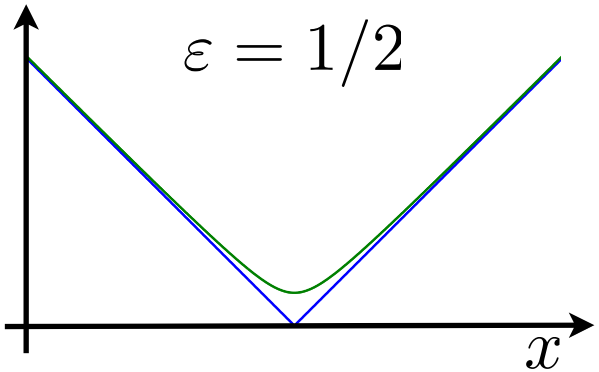


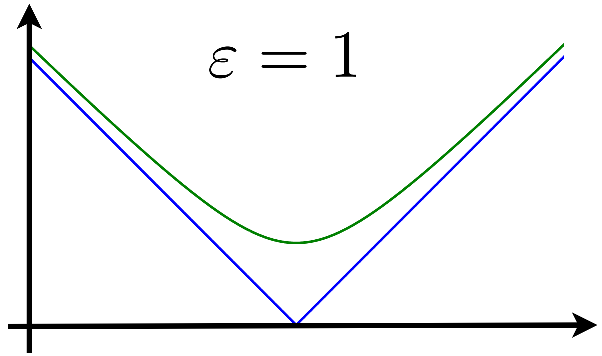
$\sqrt{x^2 + \varepsilon^2}$

$|x|$

Regularized Total Variation

$$\|u\|_\varepsilon = \sqrt{\|u\|^2 + \varepsilon^2} \qquad J_\varepsilon(f) = \sum_n \|\nabla f[n]\|_\varepsilon$$



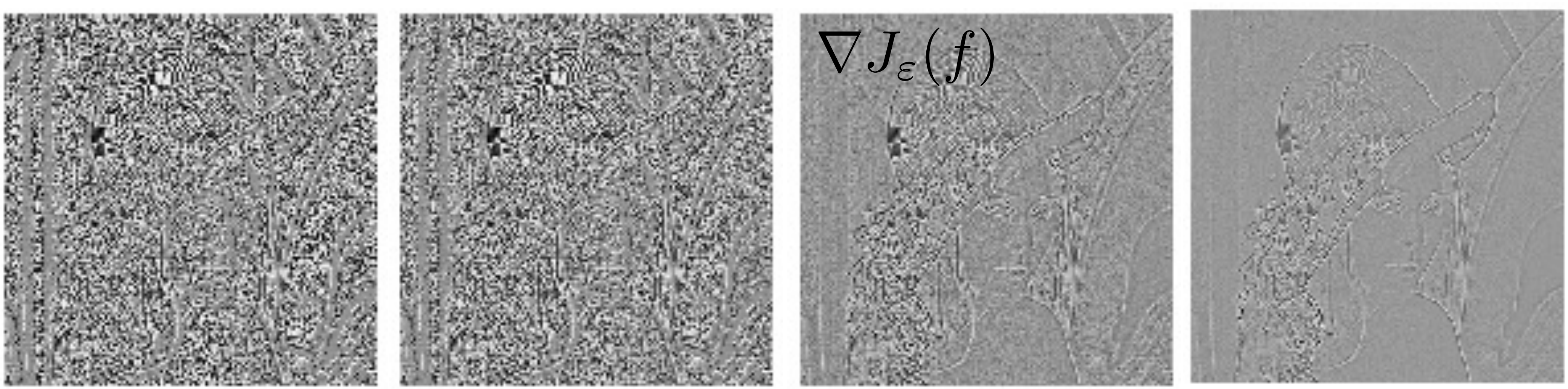


$\sqrt{x^2 + \varepsilon^2}$

$|x|$

Regularized gradient:
$$\nabla J_\varepsilon(f) = -\mathrm{div}\left(\frac{\nabla f}{\|\nabla f\|_\varepsilon}\right)$$

$$\nabla J_\varepsilon \sim -\Delta/\varepsilon \quad \text{when} \quad \varepsilon \to +\infty$$



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

Overview

- Variational Priors
- **Gradient Descent and PDE's**
- Inverse Problems Regularization

Gradient Descent

Discrete energy minimization: gradient descent

$$f^{(k+1)} = f^{(k)} - \tau_k \nabla J(f^{(k)}) \qquad\qquad f^{(0)} \text{ is given.}$$

Gradient Descent

Discrete energy minimization: gradient descent

$$f^{(k+1)} = f^{(k)} - \tau_k \nabla J(f^{(k)}) \qquad f^{(0)} \text{ is given.}$$

Theorem: If f is convex, C^1, ∇f is L-Lipschitz, and $0 < \tau < 2/L$, then $f^{(k)} \overset{k \to +\infty}{\longrightarrow} f^\star$ a solution of $\min_f J(f)$.

Gradient Descent

Discrete energy minimization: gradient descent

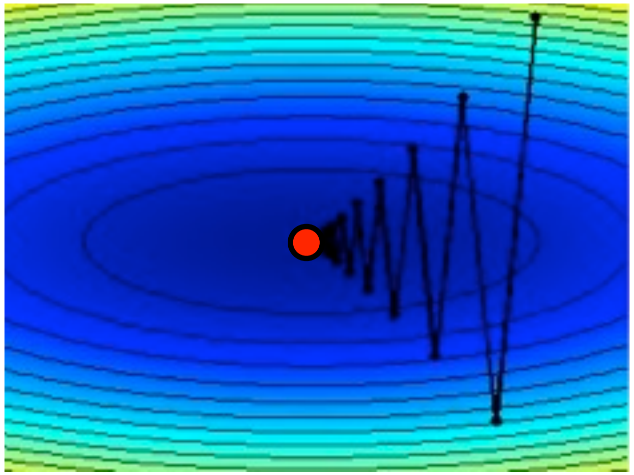
$$f^{(k+1)} = f^{(k)} - \tau_k \nabla J(f^{(k)}) \qquad\qquad f^{(0)} \text{ is given.}$$

Theorem: If f is convex, C^1, ∇f is L-Lipschitz, and $0 < \tau < 2/L$, then $f^{(k)} \overset{k\to+\infty}{\longrightarrow} f^\star$ a solution of $\min_f J(f)$.

Optimal step size: $\tau_k = \underset{\tau\in\mathbb{R}^+}{\operatorname{argmin}}\ J(f^{(k)} - \tau \nabla J(f^{(k)}))$

Proposition: One has

$$\langle \nabla J(f^{(k+1)}), \nabla J(f^{(k)}) \rangle = 0$$



Gradient Flows and PDE's

Fixed step size $\tau_k = \tau$: $$\frac{f^{(k+1)} - f^{(k)}}{\tau} = -\nabla J(f^{(k)})$$

Gradient Flows and PDE's

Fixed step size $\tau_k = \tau$: $$\frac{f^{(k+1)} - f^{(k)}}{\tau} = -\nabla J(f^{(k)})$$

Denote $f_t = f^{(k)}$ for $t = k\tau$, one obtains formally as $\tau \to 0$:

$$\forall\, t > 0, \quad \frac{\partial f_t}{\partial t} = -\nabla J(f_t) \quad \text{and} \quad f_0 = f^{(0)}$$

Gradient Flows and PDE's

Fixed step size $\tau_k = \tau$: $$\frac{f^{(k+1)} - f^{(k)}}{\tau} = -\nabla J(f^{(k)})$$

Denote $f_t = f^{(k)}$ for $t = k\tau$, one obtains formally as $\tau \to 0$:

$$\forall\, t > 0, \quad \frac{\partial f_t}{\partial t} = -\nabla J(f_t) \quad \text{and} \quad f_0 = f^{(0)}$$

Sobolev flow: $J(f) = \int \|\nabla f(x)\| \mathrm{d}x$

Heat equation: $$\frac{\partial f_t}{\partial t} = \Delta f_t$$

Explicit solution:

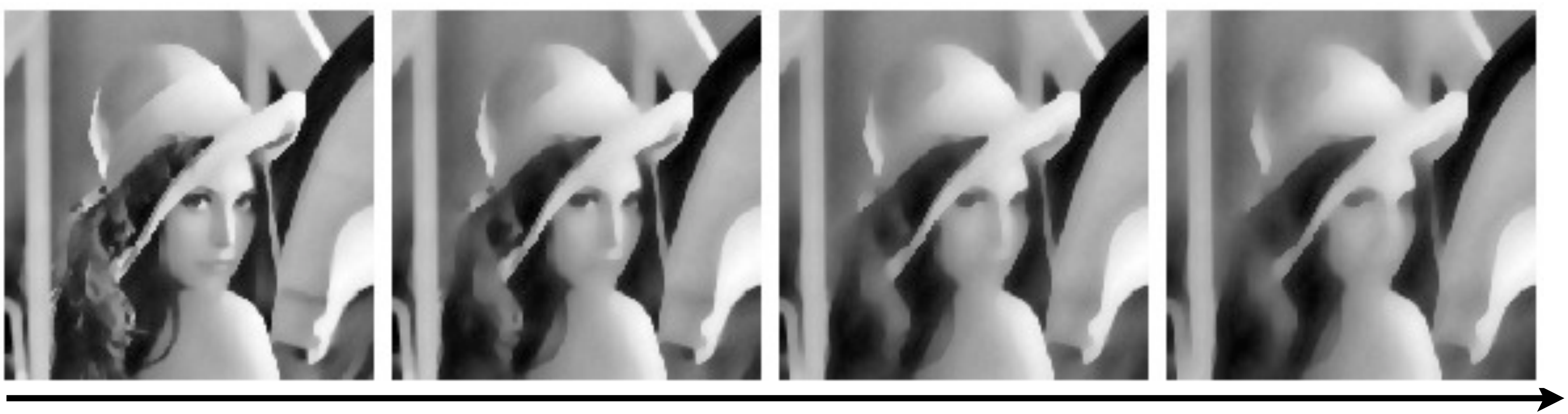
$$f_t = f \star h_t \quad \text{where} \quad h_t(x) = \frac{1}{4\pi t} e^{-\frac{-\|x\|^2}{4t}}$$

Total Variation Flow

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

TV flow smooth less the edges than heat diffusion.

$$f_t \to \text{constant when } t \to +\infty.$$



t

Application: Denoising

Noisy observations: $y = f + w$, $w \sim \mathcal{N}(0, \mathrm{Id}_N)$.

Denoising using gradient flow:

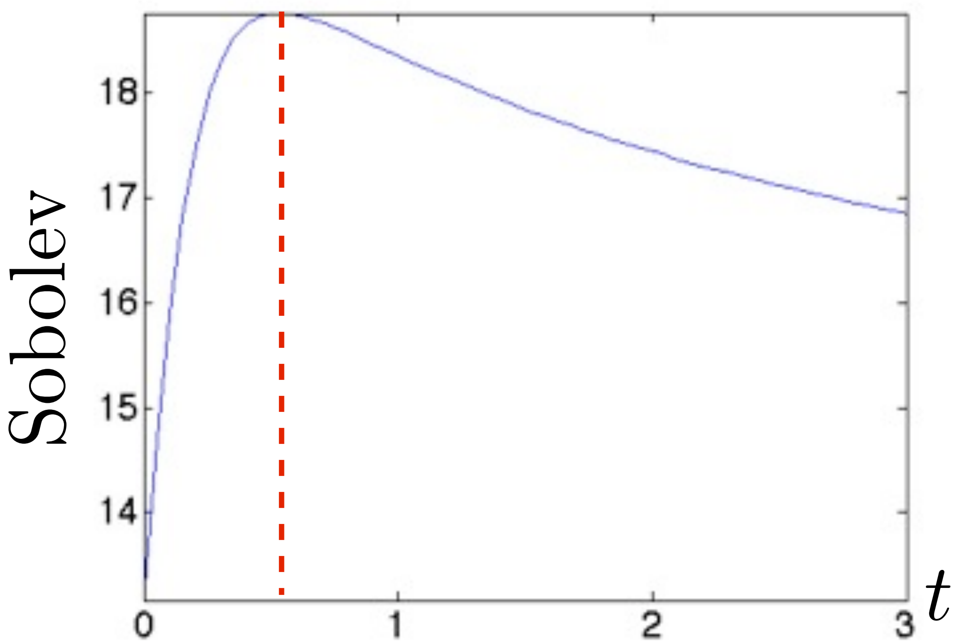
$$\frac{\partial f_t}{\partial t} = -\nabla J(f_t) \quad \text{and} \quad f_{t=0} = y$$

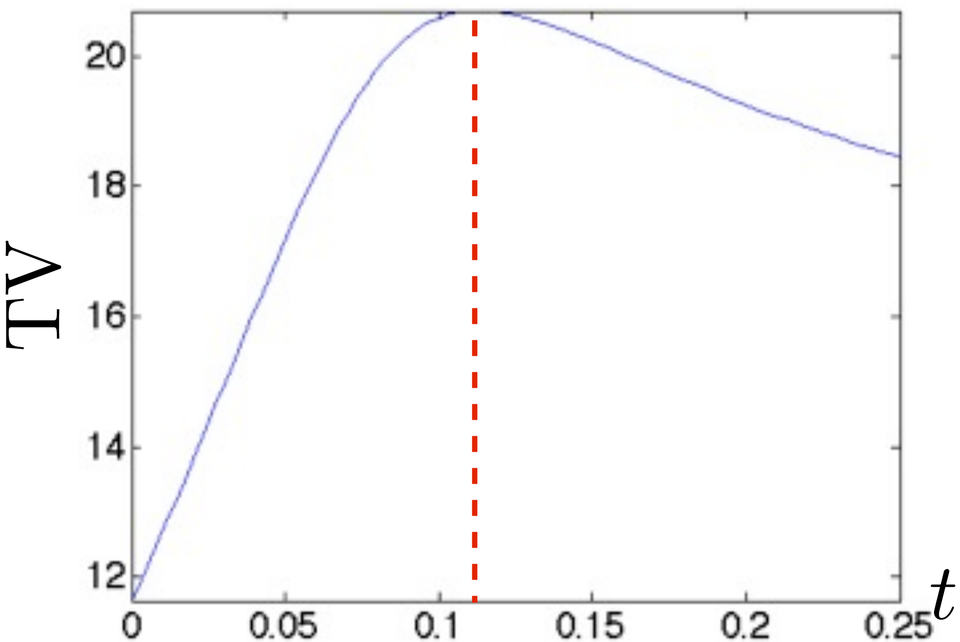
Optimal Parameter Selection

Optimal choice of t: minimize $\|f_t - f\|$

$\longrightarrow$ not accessible in practice.

$$\mathrm{SNR}(f_t, f) = -20\log_{10}\left(\frac{\|f - f_t\|}{\|f\|}\right)$$





Overview

- Variational Priors
- Gradient Descent and PDE's
- **Inverse Problems Regularization**

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.

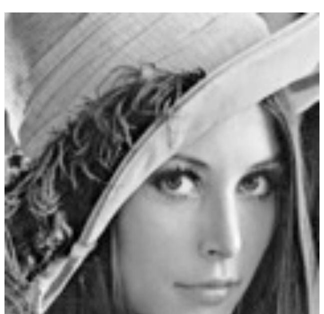
Denoising: $\Phi = \mathrm{Id}_N$, $P = N$.

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

$\xrightarrow{\Phi}$

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

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



$\xrightarrow{\Phi}$

Inverse Problems

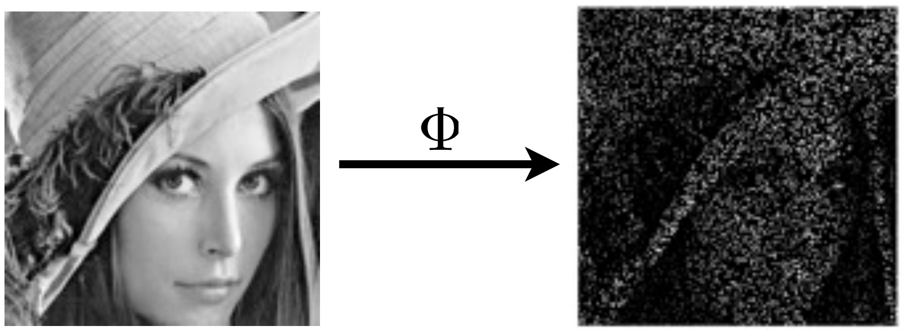
Recovering f_0 from P noisy measurements $y = \Phi f_0 + w \in \mathbb{R}^P$.

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

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

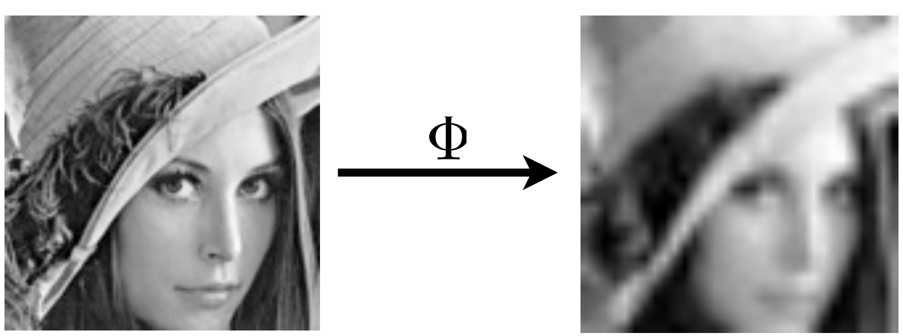
Denoising: $\Phi = \mathrm{Id}_N$, $P = N$.

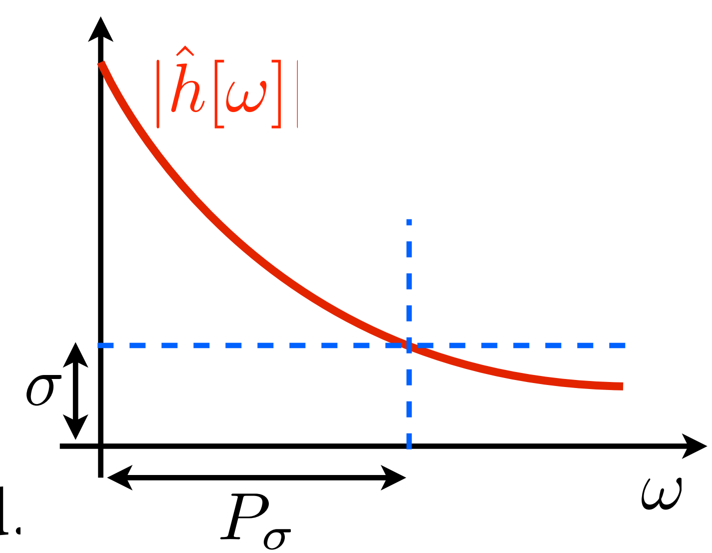
Inpainting: set $\Omega \subset \{0, \ldots, N-1\}$ of missing pixels, $P = N - |\Omega|$.



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

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





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:

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

Inverse Problem Regularization

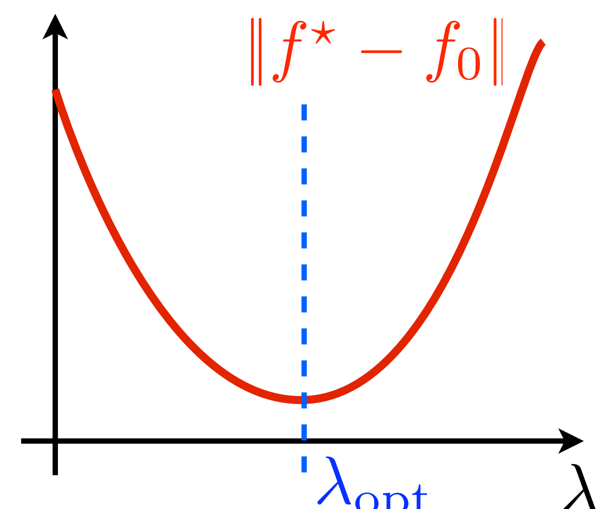
Noisy measurements $y = \Phi f_0 + w \in \mathbb{R}^P$, $w[n] \sim \mathcal{N}(0, \sigma)$.

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.



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 $f_0 \in \Theta$.

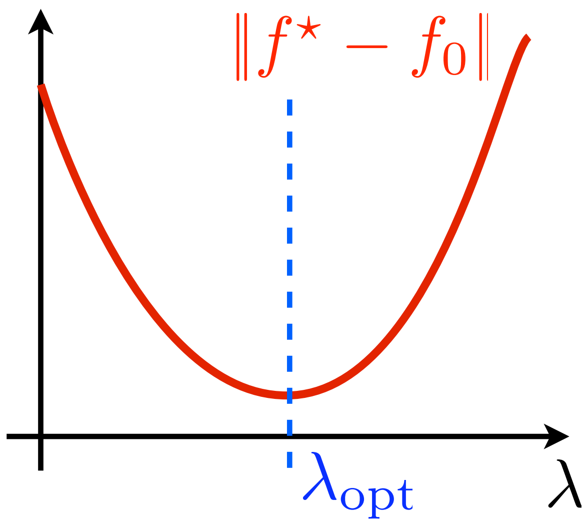
Regularized inverse:
$$f^\star = \underset{f \in \mathbb{R}^N}{\operatorname{argmin}} \underbrace{\boxed{\frac{1}{2}\|y - \Phi f\|^2}}_{\text{Data fitting}} + \underbrace{\boxed{\lambda J(f)}}_{\text{Regularity}}$$

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

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



Sobolev Regularization

Sobolev prior: $J(f) = \frac{1}{2}\|\nabla f\|^2$ (assuming $1 \notin \ker(\Phi)$)

$$f^\star = \underset{f \in \mathbb{R}^N}{\operatorname{argmin}}\, \mathcal{E}(f) = \|y - \Phi f\|^2 + \lambda \|\nabla f\|^2$$

Sobolev Regularization

Sobolev prior: $J(f) = \frac{1}{2}\|\nabla f\|^2$ (assuming $1 \notin \ker(\Phi)$)

$$f^\star = \underset{f \in \mathbb{R}^N}{\operatorname{argmin}}\, \mathcal{E}(f) = \|y - \Phi f\|^2 + \lambda \|\nabla f\|^2$$

Proposition: $\nabla \mathcal{E}(f^\star) = 0 \iff (\Phi^* \Phi - \lambda \Delta) f^\star = \Phi^* y$

$\longrightarrow$ Large scale linear system.

Sobolev Regularization

Sobolev prior: $J(f) = \frac{1}{2}\|\nabla f\|^2$ (assuming $1 \notin \ker(\Phi)$)

$$f^\star = \underset{f \in \mathbb{R}^N}{\text{argmin}}\, \mathcal{E}(f) = \|y - \Phi f\|^2 + \lambda \|\nabla f\|^2$$

Proposition: $\nabla \mathcal{E}(f^\star) = 0 \iff (\Phi^* \Phi - \lambda \Delta) f^\star = \Phi^* y$

$\longrightarrow$ Large scale linear system.

Gradient descent:

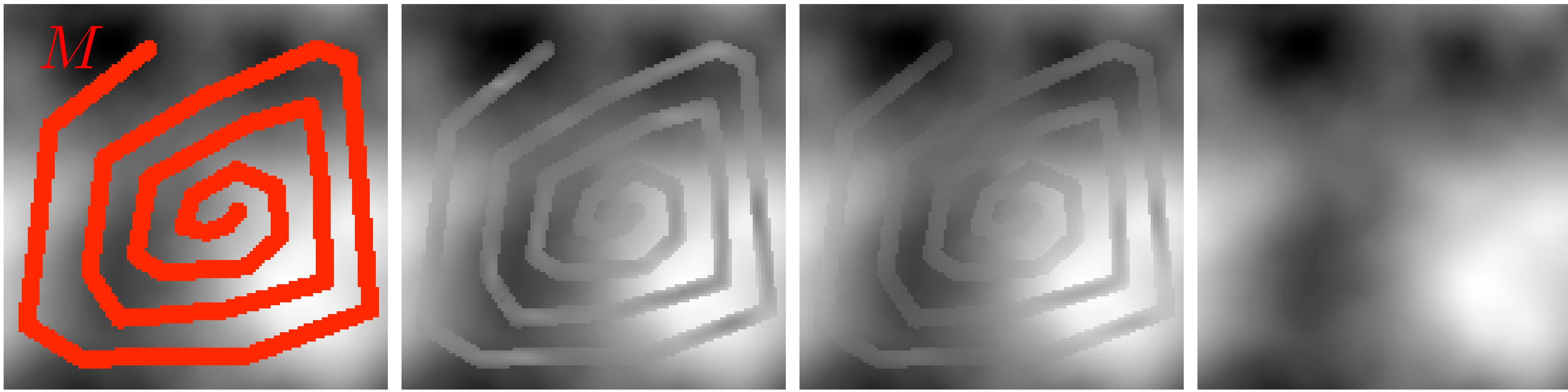
$$f^{(k+1)} = f^{(k)} - \tau \left(\Phi^*(\Phi f^{(k)} - y) - \lambda \Delta f^{(k)} \right)$$

Convergence: $\tau < 2/\|\Phi^*\Phi - \lambda\Delta\|$ where $\|A\| = \lambda_{\max}(A)$

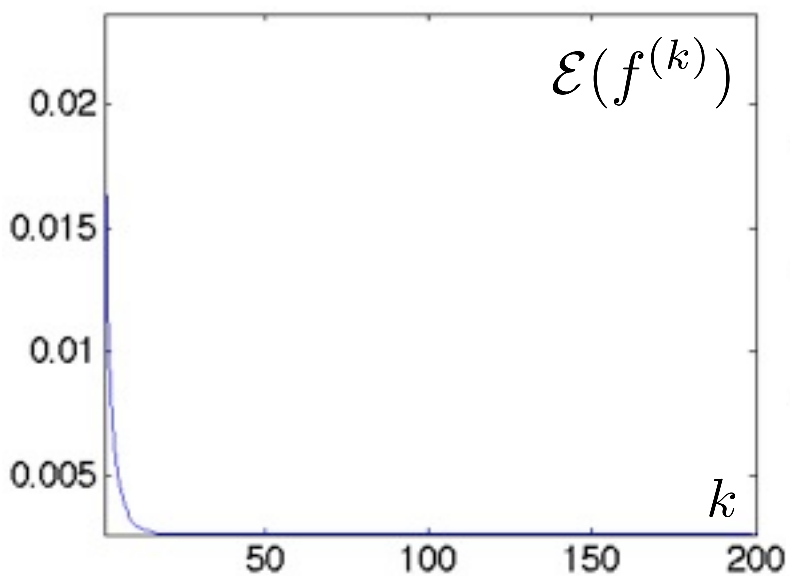
$\longrightarrow$ Slow convergence.

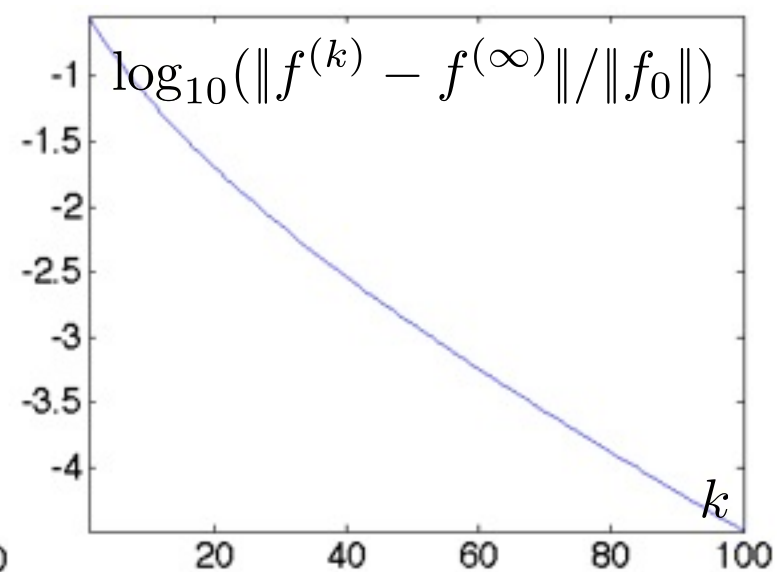
Example: Inpainting

Mask M, $\quad \Phi = \mathrm{diag}_i(1_{i \in M}) \qquad (\Phi f)[i] = \begin{cases} 0 & \text{if} \quad i \in M, \\ f[i] & \text{otherwise.} \end{cases}$



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





Conjugate Gradient

Symmetric linear system:

$$Ax = b \iff \min_{x \in \mathbb{R}^n} \mathcal{E}(x) = \frac{1}{2}\langle Ax,\, x\rangle - \langle x,\, b\rangle$$

Conjugate Gradient

Symmetric linear system:

$$Ax = b \iff \min_{x \in \mathbb{R}^n} \mathcal{E}(x) = \frac{1}{2}\langle Ax,\, x\rangle - \langle x,\, b\rangle$$

Intuition: $\quad x^{(k+1)} = \operatorname{argmin} \mathcal{E}(x)$

$$\text{s.t.} \quad x - x^{(k)} \in \operatorname{span}(\nabla\mathcal{E}(x^{(0)}), \ldots, \nabla\mathcal{E}(x^{(k)}))$$

Proposition: $\quad \forall\, \ell < k, \quad \langle \nabla\mathcal{E}(x^k),\, \nabla\mathcal{E}(x^\ell)\rangle = 0$

Conjugate Gradient

Symmetric linear system:

$$Ax = b \iff \min_{x \in \mathbb{R}^n} \mathcal{E}(x) = \frac{1}{2}\langle Ax,\, x\rangle - \langle x,\, b\rangle$$

Intuition: $\quad x^{(k+1)} = \operatorname{argmin} \mathcal{E}(x)$

$$\text{s.t.} \quad x - x^{(k)} \in \operatorname{span}(\nabla\mathcal{E}(x^{(0)}), \ldots, \nabla\mathcal{E}(x^{(k)}))$$

Proposition: $\quad \forall\, \ell < k, \quad \langle \nabla\mathcal{E}(x^k),\, \nabla\mathcal{E}(x^\ell)\rangle = 0$

Initialization: $\quad x^{(0)} \in \mathbb{R}^N, r^{(0)} = b - Ax^{(0)}, p^{(0)} = r^{(0)}$

Iterations: $\quad v^{(k)} = \nabla\mathcal{E}(x^{(k)}) = Ax^{(k)} - b$

$$x^{(k+1)} = x^{(k)} - r^{(k)} d^{(k)} \;\left|\; \begin{aligned} d^{(k)} &= \nabla\mathcal{E}(x^{(k)}) + \frac{\|v^{(k)}\|}{\|v^{(k-1)}\|} d^{(k-1)} \\ r^{(k)} &= \frac{\langle \nabla\mathcal{E}(x^{(k)}),\, d^{(k)}\rangle}{\langle Ad^{(k)},\, d^{(k)}\rangle} \end{aligned}\right.$$

Total Variation Regularization

$$\|u\|_\varepsilon = \sqrt{\|u\|^2 + \varepsilon^2} \qquad\qquad J_\varepsilon(f) = \sum_n \|\nabla f[n]\|_\varepsilon$$

TV_ε *regularization:* $\qquad$ (assuming $1 \notin \ker(\Phi)$)

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

Total Variation Regularization

$$\|u\|_\varepsilon = \sqrt{\|u\|^2 + \varepsilon^2} \qquad\qquad J_\varepsilon(f) = \textstyle\sum_n \|\nabla f[n]\|_\varepsilon$$

TV_ε regularization: (assuming $1 \notin \ker(\Phi)$)

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

Gradient descent: $f^{(k+1)} = f^{(k)} - \tau_k \nabla\mathcal{E}(f^{(k)})$

$$\nabla\mathcal{E}(f) = \Phi^*(\Phi f - y) + \lambda \nabla J_\varepsilon(f)$$

$$\nabla J_\varepsilon(f) = -\text{div}\left(\frac{\nabla f}{\|\nabla f\|_\varepsilon}\right)$$

Convergence: requires $\tau \sim \varepsilon$.

Total Variation Regularization

$$\|u\|_\varepsilon = \sqrt{\|u\|^2 + \varepsilon^2} \qquad\qquad J_\varepsilon(f) = \sum_n \|\nabla f[n]\|_\varepsilon$$

TV_ε regularization: (assuming $1 \notin \ker(\Phi)$)

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

Gradient descent: $f^{(k+1)} = f^{(k)} - \tau_k \nabla \mathcal{E}(f^{(k)})$

$$\nabla \mathcal{E}(f) = \Phi^*(\Phi f - y) + \lambda \nabla J_\varepsilon(f)$$

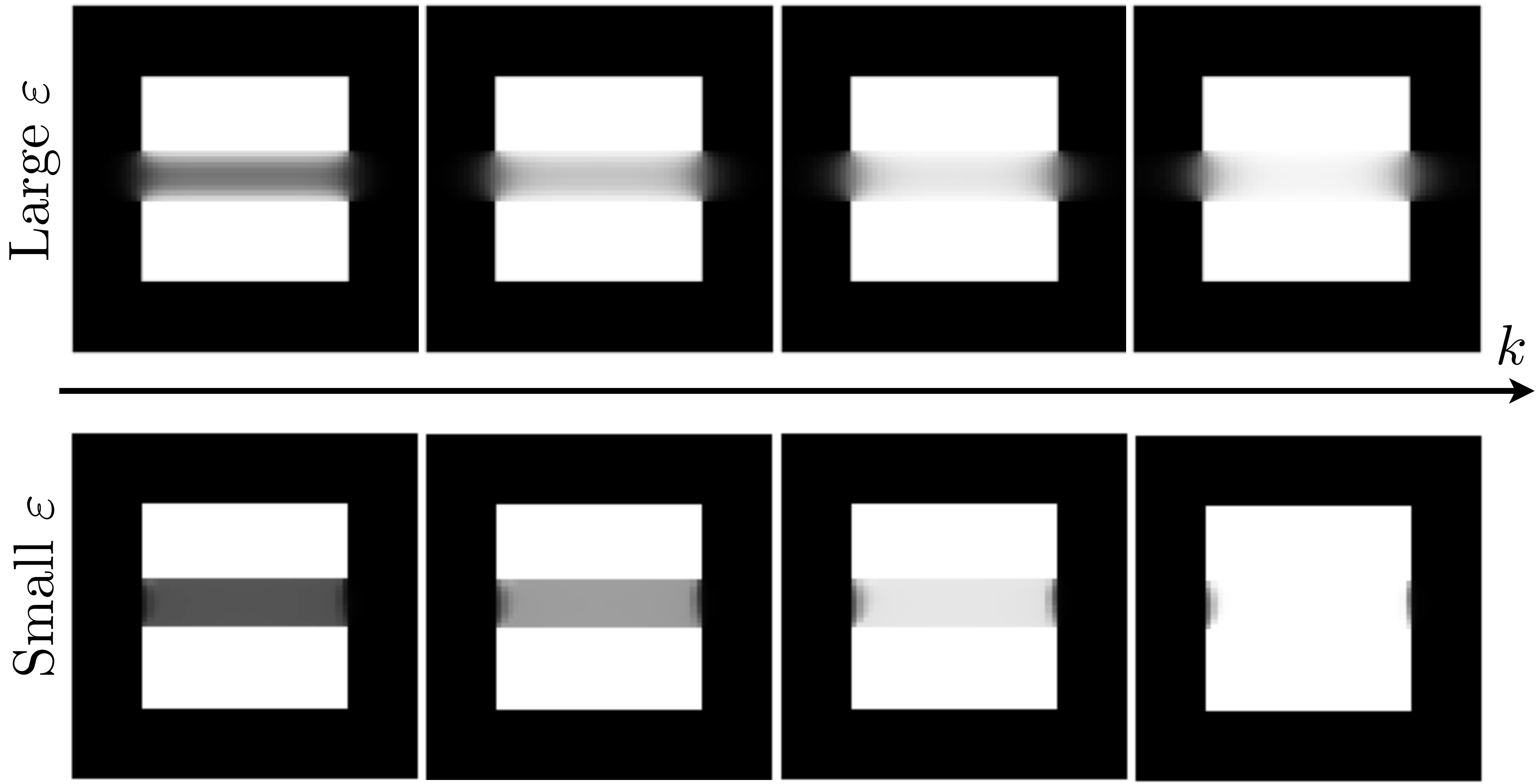
$$\nabla J_\varepsilon(f) = -\mathrm{div}\left(\frac{\nabla f}{\|\nabla f\|_\varepsilon} \right)$$

Convergence: requires $\tau \sim \varepsilon$.

Newton descent:

$$f^{(k+1)} = f^{(k)} - H_k^{-1} \nabla \mathcal{E}(f^{(k)}) \quad \text{where} \quad H_k = \partial^2 \mathcal{E}_\varepsilon(f^{(k)})$$

TV vs. Sobolev Converge



Inpainting: Sobolev vs. TV

Observations y

Sobolev

Total variation

Projected Gradient Descent

Noiseless problem: $$f^\star \in \underset{f}{\operatorname{argmin}}\ J^\varepsilon(f) \quad \text{s.t.} \quad f \in \mathcal{H} \quad (\star)$$

Contraint: $\mathcal{H} = \{f \,;\, \Phi f = y\}$.

Projected gradient descent:

$$f^{(k+1)} = \operatorname{Proj}_{\mathcal{H}}\left(f^{(k)} - \tau_k \nabla J_\varepsilon(f^{(k)})\right)$$

$$\operatorname{Proj}_{\mathcal{H}}(f) = \underset{\Phi g = y}{\operatorname{argmin}}\ \|g - f\|^2 = f + \Phi^*(\Phi^*\Phi)^{-1}(y - \Phi f)$$

Inpainting: $$\operatorname{Proj}_{\mathcal{H}}(f)[i] = \begin{cases} f[i] & \text{if} \quad i \in M, \\ y[i] & \text{otherwise.} \end{cases}$$

Proposition: If ∇J_ε is L-Lipschitz and $0 < \tau_k < 2/L$,

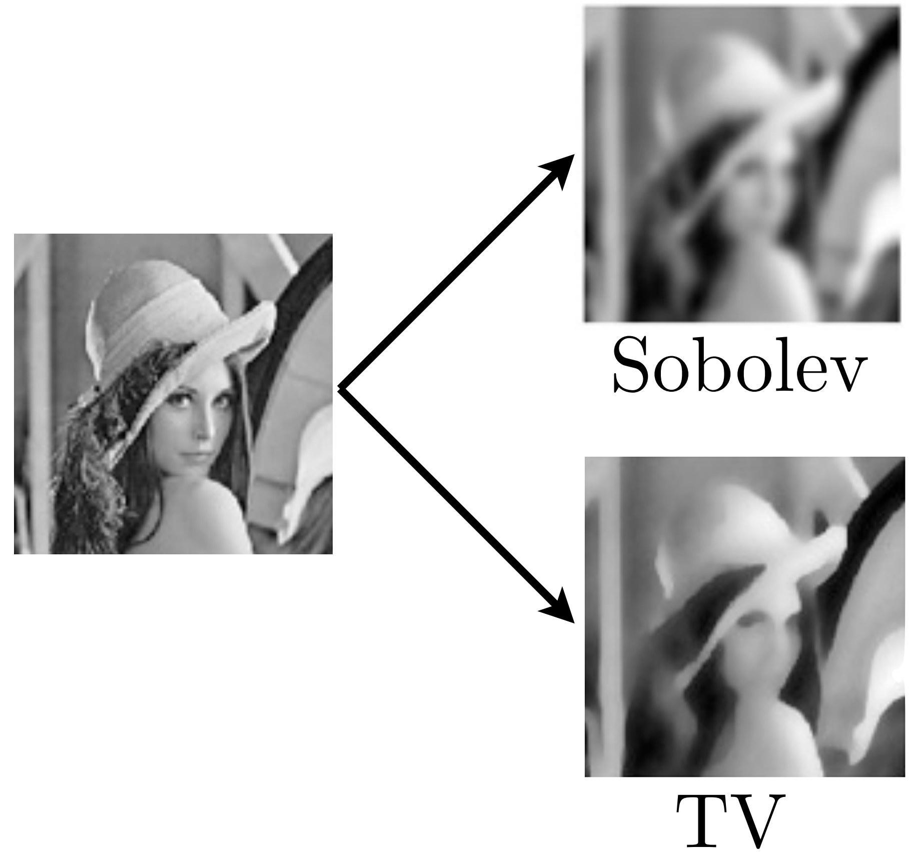
$$f^{(k)} \overset{k \to +\infty}{\longrightarrow} f^\star \qquad \text{a solution of } (\star).$$

Conclusion

Priors: Non-quadratic

$\Downarrow$

better edge recovery.

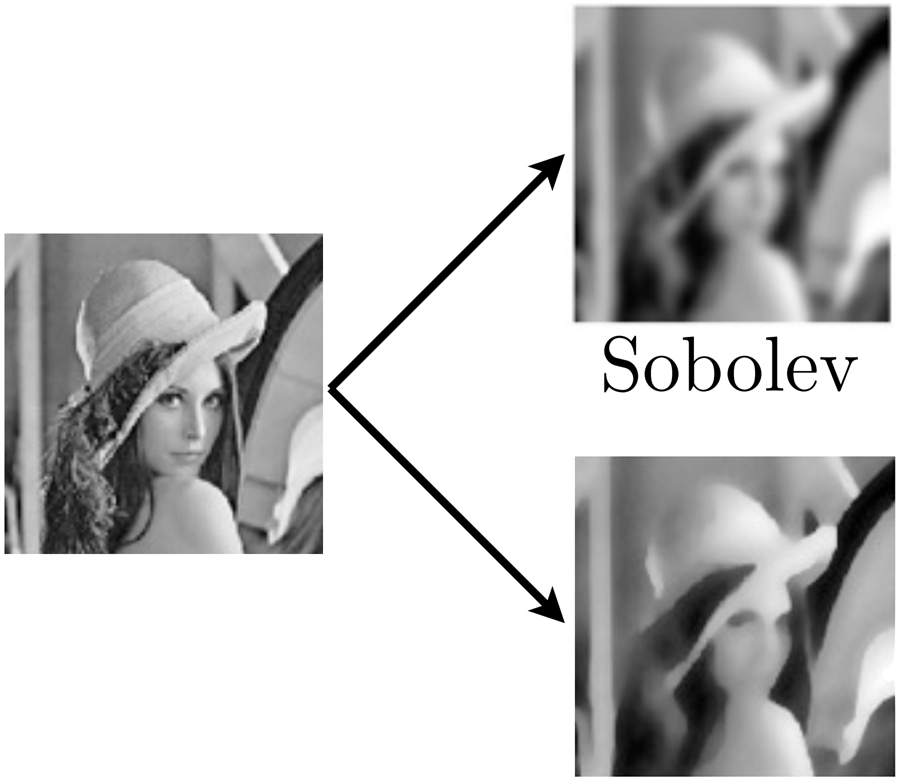


Conclusion

Priors: Non-quadratic

$$\Downarrow$$

better edge recovery.



Sobolev

TV

Variational regularization:

$$\Updownarrow$$

Optimization

– Gradient descent.
– Projected gradient.
– Newton.
– Conjugate gradient.

$\longrightarrow$ *Non-smooth optimization ?*