WHAT ENERGY TO MINIMIZE ?

Mila Nikolova

CMLA, ENS Cachan, CNRS, PRES UniverSud
61 Av. President Wilson, F-94230 Cachan, FRANCE

http://www.cmla.ens-cachan.fr/ nikolova
nikolova@cmla.ens-cachan.fr

Tutorial - EUSIPCO 07 (September) - Poznan

Outline

1. Energy minimization methods and applications
2. Regularity of minimizers
3. Non-smooth regularization
4. Non-smooth data-fidelity
5. Nonsmooth data-fidelity and regularization
6. Smooth energies
7. Non-convex regularization
8. Model distortions under Bayesian MAP
9. Computational issues
10. Open questions
11. References

1. Energy minimization methods and applications

1.1 The energy method

$\boldsymbol{u_o}$ (unknown) $\qquad$ $\boldsymbol{v}$ (data) $=$ Transform($\boldsymbol{u_o}$)$+\boldsymbol{n}$ (noise)

signal, nice picture, density map $\qquad$ *dirty blurred picture, degraded transformed map*

solution $\hat{\boldsymbol{u}}$ $\nearrow$ close to data

$\searrow$ coherent with priors

$$\hat{\boldsymbol{u}} = \underset{\boldsymbol{u}\in\boldsymbol{\Omega}}{\mathbf{argmin}} \quad \boldsymbol{\mathcal{F}_v(u)} \qquad \boldsymbol{\Omega} \text{ - constraints}$$

$$\underbrace{\boldsymbol{\mathcal{F}_v(u)}}_{\text{energy}} = \underbrace{\boldsymbol{\Psi(u,v)}}_{\text{data-fidelity}} + \underbrace{\boldsymbol{\beta\Phi(u)}}_{\text{prior}}$$

Related formulation:

$$\text{minimize } \Big\{\boldsymbol{\Phi(u)} \text{ s.t. } \boldsymbol{u} \in \boldsymbol{\Omega}, \quad \boldsymbol{\Psi(u,v)} \le \boldsymbol{\tau}\Big\} \qquad \text{“s.t.”} = \text{“subject to”}$$

$\beta = \beta(\boldsymbol{\tau}) > \mathbf{0}$ - Lagrange parameter

$\boldsymbol{\Psi}$—*based on the model relating* $\boldsymbol{u_0}$ *to* $\boldsymbol{v}$ *(deterministic and random phenomena)*

$\boldsymbol{\Phi}$—*“regularity” requirements (a priori information, expected or imposed features)*

1.2 Problems solved in this way

- **Denoising** $v = u_o + n$ (often $\boldsymbol{n} \sim$ independent identically distributed (i.i.d.) noise)
- **Segmentation** v(data) $=$ $\boldsymbol{u_o}$(sketch) $+$ $\boldsymbol{n}$(texture &noise)
- **Inverse problems**—deblurring, tomography; optical, seismic and nuclear imaging...

  direct problem $\boldsymbol{v} = \boldsymbol{A}(\boldsymbol{u_o}) \odot \boldsymbol{n}$, e.g. "$\odot$" = "+", $\boldsymbol{A}$ blur, Radon transform...

  the inverse problem is often ill-posed (Hadamard conditions) $\Rightarrow$ $A^{-1}v$ makes no sense
- **Zooming** given $\boldsymbol{v}$ find an $\hat{u}$ on a finer grid
- **Superresolution** given $v_1, \ldots, v_n$—low-resolution data, find a high-resolution $\hat{u}$
- **Coding and compression** find a (sparse) representation in a frame or a basis
- **Shrinkage estimators** restore noisy frame coefficients using knowledge on their distribution
- **Learning** given an input-output training sequence, find a function that explains the system
- **Motion estimation, color reproduction and many others**

1.3 Common reasons to define energies

$\boldsymbol{\Psi}$—based on the model for the observation instrument and the noise

usually $\{\hat{\boldsymbol{u}} \ : \ \boldsymbol{\Psi}(\hat{\boldsymbol{u}}, \boldsymbol{v}) = \mathbf{0}\}$ solves the observation equation

`History:` $\hat{\boldsymbol{u}} = \arg\min_{\boldsymbol{u}} \|\boldsymbol{A}\boldsymbol{u} - \boldsymbol{v}\|^{\mathbf{2}}$ unstable if $\boldsymbol{A}$ ill-conditioned, if $A = I$ then $\hat{u} = v$

$\boldsymbol{\mathcal{F}_v(u)} = \|\boldsymbol{A}\boldsymbol{u} - \boldsymbol{v}\|^{\mathbf{2}} + \boldsymbol{\beta}\|\boldsymbol{u}\|^{\mathbf{2}}$ *[Tikhonov &Arsenin 77]* (stable but smooth edges)

In most cases $\boldsymbol{\Psi}(\boldsymbol{u}, \boldsymbol{v}) = \|\boldsymbol{A}\boldsymbol{u} - \boldsymbol{v}\|^{\mathbf{2}}$, $\boldsymbol{A}$ linear (favored by feasibility)

Ψ is a degradation model, cannot compensate for the loss of information, good prior is needed

$\boldsymbol{\Phi}$—model for the unknown $\boldsymbol{u}$ (statistics, smoothness, edges, textures, expected features)

- Bayesian approach $\boldsymbol{\Phi}(\boldsymbol{u}) = \sum_{\boldsymbol{i}} \boldsymbol{\varphi}(\|\boldsymbol{G_i}\boldsymbol{u}\|)$
- Variations, PDE $\boldsymbol{\Phi}(\boldsymbol{u}) = \int_{\boldsymbol{\Omega}} \boldsymbol{\varphi}(\|\boldsymbol{\nabla}\boldsymbol{u}\|)\boldsymbol{dx}$

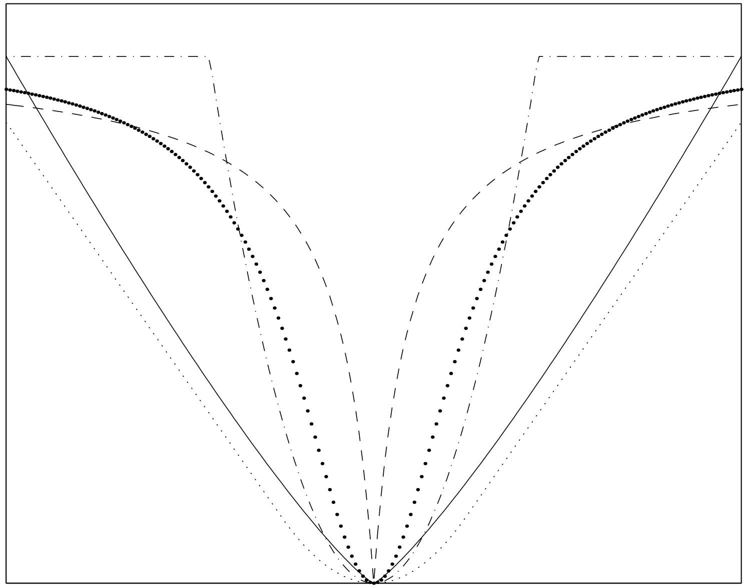
$\varphi : \mathrm{R}_+ \to \mathrm{R}_+$— *potential function (PF)*

Permanent (difficult) requirement—that φ enables the restoration of sharp edges in $\hat{\boldsymbol{u}}$

Customary functions φ where $\alpha > 0$—parameter

Convex PFs	
$\varphi(\lvert t\rvert)$ smooth at zero	$\varphi(\lvert t\rvert)$ nonsmooth at zero
$\varphi(t) = t^\alpha,\ 1<\alpha\leq 2$	$\varphi(t) = t$
$\varphi(t) = \sqrt{\alpha + t^2}$	
Nonconvex PFs	
$\varphi(\lvert t\rvert)$ smooth at zero	$\varphi(\lvert t\rvert)$ nonsmooth at zero
$\varphi(t) = \min\{\alpha t^2, 1\}$	$\varphi(t) = t^\alpha,\ 0 < \alpha < 1$
$\varphi(t) = \dfrac{\alpha t^2}{1+\alpha t^2}$	$\varphi(t) = \dfrac{\alpha t}{1+\alpha t}$
$\varphi(t) = \log(\alpha t^2 + 1)$	$\varphi(t) = \log(\alpha t + 1)$
$\varphi(t) = 1 - \exp(-\alpha t^2)$	$\varphi(0)=0,\ \varphi(t)=1$ if $t\neq 0$

$\varphi(\lvert t\rvert)$ nonsmooth at $0 \ \Leftrightarrow\ \varphi'(0) > 0$



Systematic assumptions

$\boldsymbol{\varphi}$ **is increasing on** $\mathbb{R}_+$ **with** $\boldsymbol{\varphi(0) = 0}$

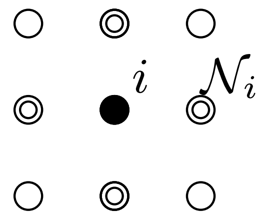
edge-preserving $\boldsymbol{\varphi}$ **satisfy** $\displaystyle\boldsymbol{\lim_{t\to\infty} \frac{\varphi'(t)}{t} = 0}$ ($\approx$ linear growth at infinity)

Bayesian Maximum a Posteriori (MAP)

- $U, \ V$ random variables, events $U = u, \ V = v$ *[Besag 89, Tenorio 01]*
- Likelihood $= f_{V|U}(v|u)$ (physical considerations on data-acquisition)

  e.g. $V = AU + N$, $\{N_i\}$ i.i.d. $\sim f_N \quad \Rightarrow \quad f_{V|U}(v|u) = \prod_i f_N\big(\langle a_i, u\rangle - v_i\big)$

  $f_N = \mathcal{N}(0, \sigma^2) \ \Rightarrow \ f_{V|U}(v|u) = \frac{1}{Z}\exp\{-\frac{\|Au-v\|^2}{2\sigma^2}\}$



- Prior $=$ $\boldsymbol{f_U(u) \propto \exp\{-\lambda\Phi(u)\}}$ (Gibbsian form)
  - *Markov models* —local characteristics— $\boldsymbol{f_U(u_i \big| u_j, j \neq i) = f_U(u_i \big| u_j, j \in \underbrace{\mathcal{N}_i}_{\text{the neighbors of } \{i\}})}$

    Hammersley-Clifford theorem $\Rightarrow \Phi(u) = \dfrac{1}{2}\sum_i \sum_{j \in \mathcal{N}_i} \boldsymbol{\varphi}(u_i - u_j)$

    powerful tool for modeling
  - *Wavelet expansions* — coefficients $x_i = \langle w_i, u\rangle$ are i.i.d. $\sim f_{X_i}(t) = e^{\left(-\lambda_i\varphi(t)\right)}\frac{1}{Z}$
- Posterior (Bayesian rule) $f_{U|V}(u|v) = f_{V|U}(v|u) f_U(x) \frac{1}{Z}$

  MAP $\quad \boldsymbol{\hat{u} = \arg\max_u f_{U|V}(u|v)}$

  $$= \arg\min_u \big(-\ln f_{V|U}(v|u) - \ln f_U(u)\big)$$

  $$= \boldsymbol{\arg\min_u\big(\ \Psi(u, v) \ + \ \beta\Phi(u)\big)}$$

Variational approach

Euler-Lagrange: $\frac{\boldsymbol{A^*(A\hat{u}-v)}}{\boldsymbol{\beta}} = \text{div}\left(\frac{\boldsymbol{\varphi'(|\nabla\hat{u}|)}}{\boldsymbol{2|\nabla\hat{u}|}}\boldsymbol{\nabla\hat{u}}\right)$ *[Weickert 98, Aubert &Kornpr. 06]*

$\boldsymbol{A = I}$, if $\boldsymbol{u_t \approx \frac{u-v}{\beta}}$

$\Rightarrow$ anisotropic diffusion with initial condition $\boldsymbol{u_0 = v}$, time step $\boldsymbol{\beta}$ *[Scherzer &Weickert 00]*

Crucial step: $\boldsymbol{A^*Au - \frac{\beta}{2}\left(\frac{\varphi'(|\nabla u|)}{|\nabla u|}u_{tt} + \varphi''(|\nabla u|)u_{nn}\right) = A^*v}$

- *homogeneous regions:* $A^*Au - \frac{\beta}{2}\varphi''(0)\Delta u \approx A^*v$ (smoothing in all directions)
- *near edges:* $|\nabla u|$ can be preserved large if $\lim\limits_{t\to\infty}\varphi''(t) = 0$ so that $\varphi''(|\nabla u|)u_{nn} \approx 0$

Qualitative result

The contribution of A along u_{tt} and u_{nn} difficult to characterize

---

Both approaches lead to similar energies

$\boxed{\textbf{1.4 Minimizer approach}}$

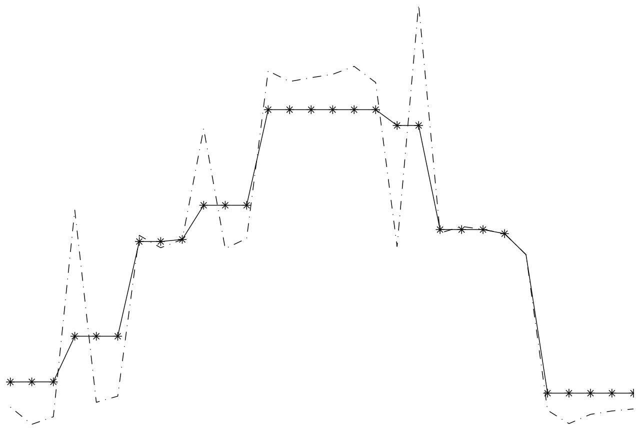
(the core of this tutorial)

Notice that $\hat{\boldsymbol{u}}$ is an implicit function of $\boldsymbol{v}$ and of the shape of $\boldsymbol{\mathcal{F}_v}$

Analyze the main features exhibited by the (local) minimizers $\hat{\boldsymbol{u}}$ of $\boldsymbol{\mathcal{F}_v}$ as a function of the shape of $\boldsymbol{\mathcal{F}_v}$

- Point of view able to yield strong results on the solutions
  (not yet explored in a systematic way)
- Focus on the singularities of $\mathcal{F}_v$ (non-smoothness, non-convexity)

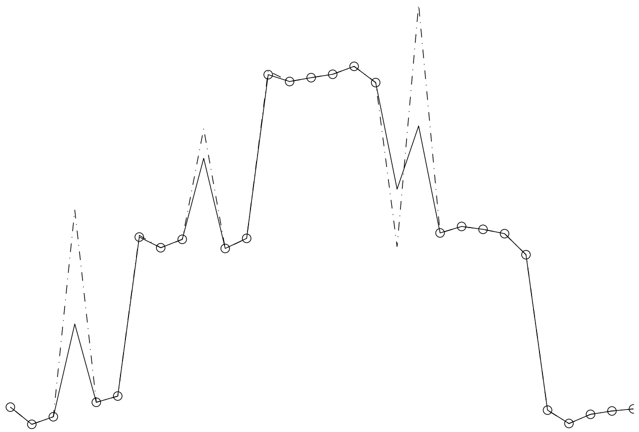
Illustration: the role of the smoothness of $\mathcal{F}_v$



STAIR-CASING

$$\mathcal{F}_v(u) = \sum_{i=1}^{p}(u_i - v_i)^2 + \beta \sum_{i=1}^{p-1}|u_i - u_{i+1}|$$

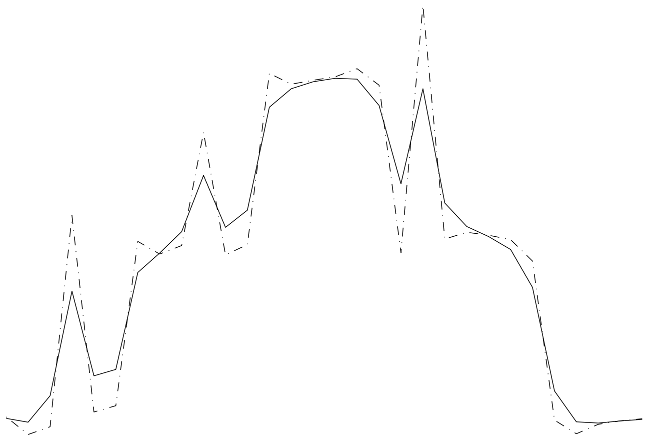
smooth **non-smooth**



EXACT DATA-FIT

$$\mathcal{F}_v(u) = \sum_{i=1}^{p}|u_i - v_i| + \beta \sum_{i=1}^{p-1}(u_i - u_{i+1})^2$$

non-smooth **smooth**



$$\mathcal{F}_v(u) = \sum_{i=1}^{p}(u_i - v_i)^2 + \beta \sum_{i=1}^{p-1}(u_i - u_{i+1})^2$$

smooth **smooth**

Data (−. − .−), Minimizer (—)

2 Regularity results

2.1 Preliminaries

[Hiriart-Urruty &Lemaréchal 96, Rockafellar 97, Ciarlet 00]

- $\mathcal{F}_v : \mathbb{R}^p \to \mathbb{R}$ is coercive if $\lim_{\|u\|\to\infty} \mathcal{F}_v(u) = \infty \quad \Rightarrow \quad \exists \hat{\boldsymbol{u}}$ **such that** $\boldsymbol{\mathcal{F}_v(\hat{u}) = \inf_{u\in\Omega} \mathcal{F}_v(u)}$

- $\Omega \subset \mathbb{R}^p$ is convex if $u, w \in \Omega$ and $\theta \in (0,1) \;\Rightarrow\; \theta u + (1-\theta)w \in \Omega$

  an affine manifold $\{u \in \mathbb{R}^p : Au = b\}$ is convex

- $\mathcal{F}_v : \mathbb{R}^p \to \mathbb{R}$ is convex if $\mathcal{F}_v\big(\theta u + (1-\theta)w\big) \le \theta\mathcal{F}_v(u) + (1-\theta)\mathcal{F}_v(w)$, $\forall u, w$ and $\theta \in (0,1)$

  $\mathcal{F}_v$ is strictly convex if the equality holds only for $u = w$

  $D^2\mathcal{F}_v \succ 0$ (positive definite) $\Rightarrow \mathcal{F}_v$ strictly convex

- $\mathcal{F}_v : \Omega \to \mathbb{R}$ convex, coercive, continuous, Ω convex $\;\Rightarrow\;$ the set of its minimizers

  $\Big\{\hat{u} \in \Omega : \mathcal{F}_v(\hat{u}) = \inf_{u\in\Omega} \mathcal{F}_v(u)\Big\}$ is closed and convex

- $\mathcal{F}_v : \mathbb{R}^p \to \mathbb{R}$ coercive, strictly convex $\Rightarrow \exists \hat{\boldsymbol{u}}$, unique : $\mathcal{F}_v(\hat{u}) = \inf_{u\in\mathbb{R}^p} \mathcal{F}_v(u)$

- Right-derivative of $\mathcal{F}_v$ at u along w: $\delta\mathcal{F}_v(u)(w) = \lim_{t\downarrow 0} \frac{\mathcal{F}_v(u+tw)-\mathcal{F}_v(u)}{t}$

  The relevant left-derivative is $\quad -\delta\mathcal{F}_v(u)(-w)$

  $\mathcal{F}_v$ differentiable at u along $w \Rightarrow -\delta\mathcal{F}_v(u)(-w) = \delta\mathcal{F}_v(u)(w) = \frac{d}{dt}\mathcal{F}_v(u+tw)\Big|_{t=0}$

- **Theorem 2.1**

$$\boxed{\begin{array}{c} \mathcal{F}_v \text{ has a (local) minimum at } \hat{u} \;\Rightarrow\; \delta\mathcal{F}_v(\hat{u})(w) \geq 0, \quad \forall w \in \mathbf{R}^p \\ \Leftrightarrow \; -\delta\mathcal{F}_v(\hat{u})(-w) \leq 0 \leq \delta\mathcal{F}_v(\hat{u})(w), \quad \forall w \in \mathbf{R}^p \\ \textbf{left derivative} \leq 0 \leq \textbf{right derivative} \end{array}}$$

  $\mathcal{F}_v$ differentiable at $\hat{u} \Rightarrow \langle \nabla\mathcal{F}_v(\hat{u}), w\rangle = 0, \; \forall w \in \mathbf{R}^p \quad \Leftrightarrow \; \nabla\mathcal{F}_v(\hat{u}) = 0$

- $\underline{\mathcal{F}_v : \mathbf{R}^p \to \mathbf{R} \textit{ convex}}$

  Subdifferential of $\mathcal{F}_v$ at u $\quad \partial\mathcal{F}_v(u) = \Big\{ g \in \mathbf{R}^p : \langle g, w\rangle \leq \delta\mathcal{F}_v(u)(w), \; \forall w \in \mathbf{R}^p \Big\}$

  $\partial\mathcal{F}_v$ is nonempty, compact and convex

  $\mathcal{F}_v$ differentiable at $u \;\Rightarrow\; \partial\mathcal{F}_v(u) = \{\nabla\mathcal{F}_v(u)\}$

  $g \in \partial\mathcal{F}_v(u)$ is a subgradient of $\mathcal{F}_v$ at u

  Theorem 2.2 $\boxed{\mathcal{F}_v \textit{ has a minimum at } \hat{u} \;\Leftrightarrow\; 0 \in \partial\mathcal{F}_v(\hat{u}) \;\Leftrightarrow\; \delta\mathcal{F}_v(\hat{u})(w) \geq 0, \forall w}$

We focus on

$$\boxed{\begin{aligned} \mathcal{F}_v(u) &= \|Au - v\|^2 + \beta\Phi(u) \\ \Phi(u) &= \sum_{i=1}^{r} \varphi(\|G_i u\|) \end{aligned}}$$

- $\{G_i\}$ linear operators $\mathrm{R}^p \to \mathrm{R}^s$, $s \geq 1$ (typically $s = 1$ or $s = 2$), $\quad G = [G_1^*, \ldots, G_r^*]^*$

  systematic assumption:

$$\ker A \cap \ker G = \{0\}$$

- $\mathcal{U} : \mathbf{R}^q \to \mathbf{R}^p$ *(strict) minimizer function*: $\forall v$, $\mathcal{F}_v$ has a (strict) minimum at $\mathcal{U}(v)$

- $\mathrm{rank} A = q$, $\varphi \sim \mathcal{C}^{m \geq 2}$ convex, $\varphi'(0) = 0 \quad \Rightarrow \quad \mathcal{U} \sim \mathcal{C}^{m-1}$

  (implicit functions theorem for $\nabla \mathcal{F}_v(\mathcal{U}(v)) = 0$)

- $\varphi'(0) > 0 \quad \Rightarrow \quad \mathcal{F}_v$ is nonsmooth on $\bigcup_i \left\{u : G_i u = 0\right\}$

- $\mathcal{F}_v$ **nonconvex** $\Rightarrow$ **there may be many local minima**

2.2 Stability of the solution (Φ possibly nonconvex)

[Durand &Nikolova 06]

Assumptions

- $\varphi : \mathrm{R}_+ \to \mathrm{R}$ is continuous and $\mathcal{C}^{m\geq 2}$ on $\mathrm{R}_+ \setminus \{\theta_1, \theta_2 ....\}$
- $\mathrm{rank}(A) = p$

A. LOCAL MINIMIZERS

(knowing local minimizers is important)

$$\mathbf{\Gamma}_{\mathrm{loc}} = \left\{ \boldsymbol{v} \in \mathrm{R}^q : \begin{array}{l} \text{if } \hat{u} \text{ is a strict local minimizer of } \mathcal{F}_v \text{ then there is a } \mathcal{C}^{m-1}\text{strict (local)} \\ \text{minimizer function } \boldsymbol{\mathcal{U}} : O \to \mathrm{R}^p \text{ such that } v \in O \subset \mathrm{R}^q \text{ and } \hat{\boldsymbol{u}} = \boldsymbol{\mathcal{U}}(\boldsymbol{v}) \end{array} \right\}$$

$=$ all data leading to (local) minimizers having good regularity properties

$\mathbf{\Gamma}^c_{\mathbf{loc}} = \mathrm{R}^q \setminus \Gamma_{\mathrm{loc}}$ contains all nonstrict minimizers, $\overline{\mathbf{\Gamma}^c_{\mathbf{loc}}}$ is its closure

Theorem 2.2

The Lebesgue measure of $\overline{\Gamma^c_{\mathrm{loc}}}$ *in* R^q *is* $\boldsymbol{\mathcal{L}^q}\left(\overline{\mathbf{\Gamma}^c_{\mathbf{loc}}}\right) = \mathbf{0}$

- no special assumptions if Φ smooth (result follows from Sard's theorem)

B. Important intermediate results

- <u>Φ *smooth*</u>

$$\boldsymbol{\Gamma_0^c} = \left\{ v \in \mathbf{R}^q : \exists \hat{u} \in \mathbf{R}^p \; : \; \nabla \mathcal{F}_v(\hat{u}) = \mathbf{0} \text{ and } \det \nabla^2 \mathcal{F}_v(\hat{u}) = \mathbf{0} \right\}$$

$$\Rightarrow \boldsymbol{\Gamma_0} \subset \boldsymbol{\Gamma_{\mathrm{loc}}}$$

We prove that $\boxed{\mathcal{L}^q\left(\overline{\boldsymbol{\Gamma_0^c}}\right) = \mathbf{0}}$

$\Rightarrow$ For almost every $v \in \mathrm{R}^q$ every local minimizer $\hat{u}$ of $\mathcal{F}_v$ is strict and $\nabla^2 \mathcal{F}_v(\hat{u})$ is positive definite

- Φ convex $\Rightarrow$ $\Gamma_0 = \Gamma_{\mathrm{loc}} = \mathrm{R}^p$ $\quad (\nabla^2 \mathcal{F}_v(u) \succ 0 \; \forall u)$

- <u>$\boldsymbol{\Phi}$ *piecewise smooth* $(\boldsymbol{\varphi'(0) > 0})$</u>

<u>*Notations*</u>: $\hat{u}$ MINIMIZER $\rightarrow \left\{ \begin{array}{lcl} \hat{h} & = & \left\{ i : G_i \hat{u} = 0 \right\} \\ K_{\hat{h}} & = & \left\{ w \in \mathbf{R}^p : G_i w = 0, \forall i \in \hat{h} \right\} \end{array} \right.$

($\varphi(\|G_i \hat{u}\|)$ is nonsmooth $\forall i \in \hat{h}$ but $\mathcal{F}_v$ is smooth on $K_{\hat{h}}$ near $\hat{u} \in K_{\hat{h}}$)

We prove that $\exists\, \boldsymbol{\Gamma_0} \subset \boldsymbol{\Gamma_{\text{loc}}}$, such that $\boxed{\boldsymbol{\mathcal{L}^q\left(\overline{\Gamma_0^c}\right) = 0}}$ **and :**

$$\forall v \in \boldsymbol{\Gamma_0} \quad \begin{cases} \boldsymbol{\nabla \mathcal{F}_v\big|_{K_{\hat{h}}}(\hat{u}) = 0,\ \nabla^2 \mathcal{F}_v|_{K_{\hat{h}}}(\hat{u}) \succ 0} & \text{(sufficient condition for strict minimum on } \boldsymbol{K_{\hat{h}}}) \\ \boldsymbol{\delta \mathcal{F}_v(\hat{u})(w) > 0,\quad \forall w \in K_{\hat{h}}^{\perp} \setminus \{0\}} & \text{(sufficient condition for strict minimum on } \boldsymbol{K_{\hat{h}}^{\perp}}) \end{cases}$$

where $\hat{u}$ is a minimizer of $\mathcal{F}_v$

- For almost every $v \in \mathrm{R}^q$, $\mathcal{F}_v$ has a finite number of local minimizers

$\boxed{\text{C. GLOBAL MINIMIZERS}}$ $\boldsymbol{\Gamma = \{v \in \mathbf{R}^q : \mathcal{F}_v}$ has a unique global minimizer$\boldsymbol{\}}$

Theorem 2.3

$\boldsymbol{\mathcal{L}^q(\Gamma^c) = 0}$ *and the interior of* Γ *is dense in* R^q. *The global minimizer function* $\boldsymbol{\hat{\mathcal{U}} : \Gamma \to \mathbf{R}^p}$ *is* $\boldsymbol{\mathcal{C}^{m-1}}$ *on an open subset of* Γ *which is dense in* R^q.

$\Rightarrow$ $\boxed{\textit{the optimal solution } \boldsymbol{\hat{u}} \textit{ is almost surely unique}}$

2.3 Nonasymptotic bounds on minimizers

[Nikolova 07]

- $\{G_i u\}_{i=1}^r$: discrete gradients or 1st-order differences between neighbors $\Rightarrow$ $\mathbf{ker}G = \mathbf{span}(\mathbb{1})$

  If u 1-D signal, e.g. $G_i u = u_{i+1} - u_i$ $\qquad \mathbb{1} = [1, \ldots, 1]$

  $u \sim (m \times n)$ image $G_k u = \begin{bmatrix} u_{i,j} - u_{i+1,j} \\ u_{i,j} - u_{i,j+1} \end{bmatrix}$ or $\begin{aligned} G_{2k-1} u &= u_{i,j} - u_{i+1,j} \\ G_{2k}\, u &= u_{i,j} - u_{i,j+1} \end{aligned}$

  $k = (j-1)m + i$ if image(:)

Classical bounds hold for $\beta \searrow 0$ or $\beta \nearrow \infty$

Assumptions

- φ is $\mathcal{C}^1$ on R_+, $\varphi'(t) \geq \mathbf{0}$ on $\mathbf{R}_+$
- **or** $\varphi(t) = \min\{\alpha t^2, 1\}$

$\varphi(t) = \min\{\alpha t^2, 1\}$ $\Rightarrow$ $\mathcal{F}_v$ is smooth at every local minimizer $\hat{u}$ and $\nabla \mathcal{F}_v(\hat{u}) = 0$

nonsmooth at $\frac{1}{\sqrt{\alpha}}$ [Nikolova 04]

A. Bounds on restored data $\boldsymbol{A\hat{u}}$

Theorem 2.4

- *assume* $\operatorname{rank} A = p$ **or** $\varphi'(t) > 0, \ \forall t > 0$

  If $\boldsymbol{\mathcal{F}_v}$ *has a (local) minimum at* $\boldsymbol{\hat{u}}$, *then* $\|\boldsymbol{A\hat{u}}\| \leq \|\boldsymbol{v}\|$ ("maximum principle")

- *assume* $\operatorname{rank} A = p \geq 2$, $\ker G = \operatorname{span}(\mathbb{1})$ *and* $\varphi'(t) > 0, \ \forall t > 0$

  $\Rightarrow \ \exists \boldsymbol{N} \subset \mathbf{R}^q$ *with* $\boldsymbol{\mathcal{L}^q(N) = 0}$, *such that* $\forall \boldsymbol{v} \in \mathbf{R}^q \setminus \boldsymbol{N}$, $\ \|\boldsymbol{A\hat{u}}\| < \|\boldsymbol{v}\|$

- A orthonormal (e.g. $A = \mathrm{I}$) $\Rightarrow \|\hat{\boldsymbol{u}}\| \leq \|v\|$
- Φ smooth $\Rightarrow \quad N = \left\{v \in \mathbf{R}^q : A^*v \propto A^*A\mathbb{1}\right\} \cup \ker(A^*), \quad \dim N < q$

B. The mean of restored data

- Usually **mean(noise)=0**
- If $\boldsymbol{A = I}$, then $\mathbf{mean}(\hat{\boldsymbol{u}}) = \mathbf{mean}(v)$ *[Aubert 06]*
- $\mathbf{mean}(\boldsymbol{A\hat{u}}) = \mathbf{mean}(v)$ if $\mathbb{1} \in \ker(\boldsymbol{G})$ and $\boldsymbol{A}\mathbb{1} \propto \mathbb{1}$
- In general $\mathbf{mean}(\boldsymbol{A\hat{u}}) \neq \mathbf{mean}(\boldsymbol{v})$ (try $\varphi(t) = t^2$ and any A invertible)

C. Residuals for edge-preserving regularization

Additional assumption: $\qquad \|\varphi'\|_\infty = \sup_{0\le t<\infty} \left|\varphi'(t)\right| < \infty \qquad$ (φ edge-preserving)

Theorem 2.5

Let $\mathrm{rank}(A) = q$. *For every* $v \in \mathbf{R}^q$, *if* $\mathcal{F}_v$ *has a (local) minimum at* $\hat{u}$ *then*

$$\|v - A\hat{u}\|_\infty \le \frac{\beta}{2}\, \|\varphi'\|_\infty\, \left|\!\left|\!\left|(AA^*)^{-1}A\right|\!\right|\!\right|_\infty |\!|\!|G|\!|\!|_1$$

Reminder: $|\!|\!|C|\!|\!|_1 = \max_j \sum_i |C_{i,j}|$ and $|\!|\!|C|\!|\!|_\infty = \max_i \sum_j |C_{i,j}|$

- Signal ($|\!|\!|G|\!|\!|_1 = 2$) and $A = I \quad \Rightarrow \quad \|v - \hat{u}\|_\infty \le \beta\|\varphi'\|_\infty$
- Image ($|\!|\!|G|\!|\!|_1 = 4$) and $A = I \quad \Rightarrow \quad \|v - \hat{u}\|_\infty \le 2\beta\|\varphi'\|_\infty$

Sketch of the proof for Φ smooth

$$
\begin{array}{rcll}
\nabla \mathcal{F}_v(\hat{u}) = 0 \quad \Rightarrow \quad 2A^*(v - A\hat{u}) & = & \beta \nabla \Phi(\hat{u}) & \times \frac{1}{2}(AA^*)^{-1}A \\
v - A\hat{u} & = & \frac{\beta}{2}(AA^*)^{-1}A \; \nabla \Phi(\hat{u}) & \\
\|v - A\hat{u}\|_\infty & \leq & \frac{\beta}{2} \left\vert\kern-0.25ex\left\vert\kern-0.25ex\left\vert (AA^*)^{-1}A \right\vert\kern-0.25ex\right\vert\kern-0.25ex\right\vert_\infty \|\nabla \Phi(\hat{u})\|_\infty &
\end{array}
$$

$$
\begin{array}{rcl}
\left|\frac{d}{du_n}\Phi(u)\right| & = & \left|\sum_i \left(G_i[\cdot, n]\right)^* \frac{\varphi'(\|G_i x\|)}{\|G_i x\|} G_i x\right| \leq \|\varphi'\|_\infty \sum_i \sum_j \left|G_i[j,n]\right| \\
\|\nabla \Phi(\hat{u})\|_\infty & = & \max_n \left|\frac{d}{du_n}\Phi(u)\right| \leq \|\varphi'\|_\infty \max_n \sum_i \sum_j \left|G_i[j,n]\right| = \|\varphi'\|_\infty |||G|||_1
\end{array}
$$

each $\boldsymbol{G_i}$ is a $\boldsymbol{s \times p}$ matrix

- **Remark:** $\|Au - v\|^2 \quad \sim$ white Gaussian noise (unbounded)

  However the noise estimates $\boldsymbol{(v - A\hat{u})_i}$ are tightly bounded

D. Bounds on the reconstructed differences $\boldsymbol{G\hat{u}}$

Proposition 2.6

$$\boxed{A \text{ orthonormal} \;\Rightarrow\; \|\boldsymbol{G\hat{u}}\| \le \|\!|\boldsymbol{G}|\!\|_{\mathbf{2}}\ \|\boldsymbol{v}\|}$$

Remind: $|\!|\!|C|\!|\!|_2 = \max\{\sqrt{\lambda} : \lambda =$ eigenvalue of $C^*C\} = \sup\limits_{\|u\|=1} \Big\|Cu\Big\|$

Smooth regularization $(\varphi'(0) = 0)$

We compare $\boldsymbol{G\hat{u}}$ with $\boldsymbol{G\hat{z}}$ where

$$\hat{\boldsymbol{z}} = \text{arg min}_{\boldsymbol{u}} \|\boldsymbol{Au} - \boldsymbol{v}\|^{\mathbf{2}} = (\boldsymbol{A}^*\boldsymbol{A})^{-\mathbf{1}}\boldsymbol{A}^*\boldsymbol{v} \qquad \text{(the least-squares solution)}$$

Theorem 2.7

Let $\operatorname{rank}(A) = p$. *For every* $v \in \mathrm{R}^q$, *if* $\boldsymbol{\mathcal{F}_v}$ *has a (local) minimum at* $\hat{\boldsymbol{u}}$, *then there is an* $r \times r$ *linear operator* $\boldsymbol{H_v}$ *such that*

$$\boldsymbol{G\hat{u}} = \boldsymbol{H_v}\ \boldsymbol{G}\ \hat{\boldsymbol{z}},$$

$$\text{Spectral Radius}(\boldsymbol{H_v}) \quad \le \quad \mathbf{1}$$

More precisely: $\boldsymbol{H_v} = \Big(\boldsymbol{I} + \frac{\boldsymbol{\beta}}{\mathbf{2}}\boldsymbol{G}(\boldsymbol{A}^*\boldsymbol{A})^{-\mathbf{1}}\boldsymbol{G}^*\mathbf{diag}(\boldsymbol{\theta})\Big)^{-\mathbf{1}}, \quad \boldsymbol{\theta_i} \ge \mathbf{0}, \forall \boldsymbol{i}$

Similar result for nonsmooth regularization $(\varphi'(0) > 0)$

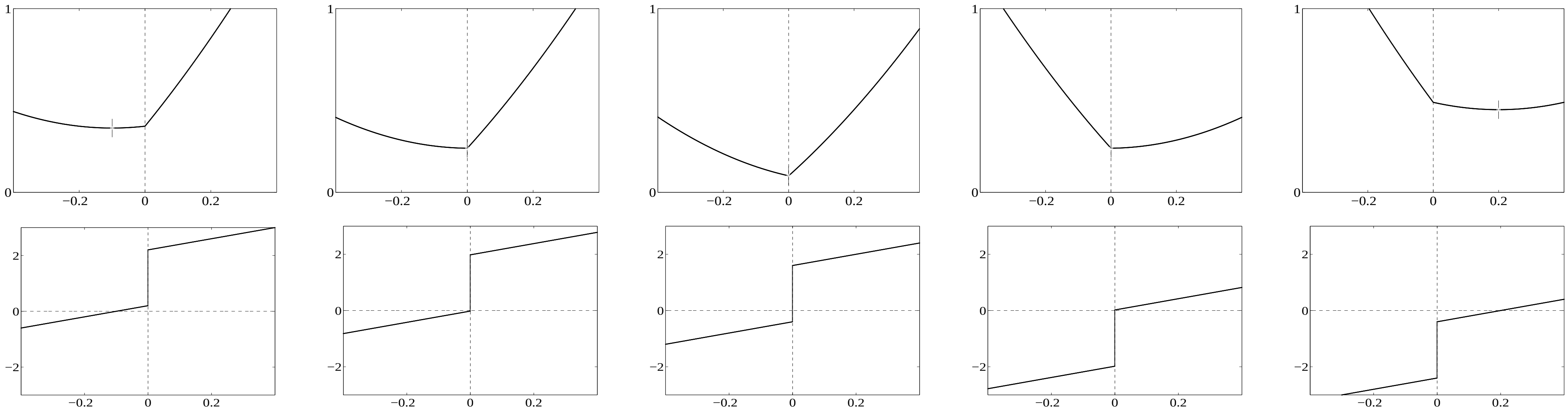
3 Minimizers under Non-Smooth Regularization

3.1 General case

[Nikolova 97,01,04]

$$\boldsymbol{\Psi} \in \mathcal{C}^{m\geq 2}, \;\; \varphi \in \mathcal{C}^m(\mathrm{R}_+) \text{ and } \varphi'(0) > 0 \qquad\qquad \mathcal{F}_v(u) = \Psi(u,v) + \beta \sum_{i=1}^{r} \varphi(\|G_i u\|)$$

Examples: $\varphi(t) = t^\alpha, \; 0 < \alpha \leq 1, \;\; \varphi(t) = \frac{\alpha t}{(1+\alpha t)}$ *[Besag89,Geman92,Rudin92,Black96...]*



$v = -0.6, \hat{u} = -0.1$ $\qquad v = -0.49, \hat{u} = 0$ $\qquad v = 0.3, \hat{u} = 0$ $\qquad v = 0.49, \hat{u} = 0$ $\qquad v = 0.7, \hat{u} = 0.2$

$\mathcal{F}_v(u) = (u-v)^2 + |u|$ (1^{st} row). Check $0 \in \partial\mathcal{F}_v(\hat{u})$ (2^{nd} row): $\hat{u} = 0$ if $|v| \leq \frac{1}{2}$, $\hat{u} = v - \frac{1}{2}\mathrm{sign}(v)$ else

Theorem 3.1 (Generalizes the above observation)

$\boldsymbol{\Psi(u,v) = \|Au - v\|^2}$, **rank**$\boldsymbol{A = p}$.
For almost every $\boldsymbol{v} \in \mathbf{R}^{\boldsymbol{q}}$ $(\forall \boldsymbol{v} \in \boldsymbol{\Gamma_0})$ *if* $\boldsymbol{\mathcal{F}_v}$ *has a (local) minimum at* $\boldsymbol{\hat{u}}$

$\Rightarrow \quad \exists\, \boldsymbol{O} \subset \mathrm{R}^q$ *open,* $\boldsymbol{v} \in \boldsymbol{O}$ *and* $\exists\, \boldsymbol{\mathcal{U}} \in \mathcal{C}^{m-1}$ (local minimizer function)

$$\boldsymbol{v'} \in \boldsymbol{O} \;\Rightarrow\; \boldsymbol{\mathcal{F}_{v'}} \text{ has a minimum at } \boldsymbol{\hat{u}' = \mathcal{U}(v')} \text{ and } \begin{cases} \boldsymbol{G_i \hat{u}' = 0} & \text{if} \quad \boldsymbol{i \in \hat{h}} \\ \boldsymbol{G_i \hat{u}' \neq 0} & \text{if} \quad \boldsymbol{i \in \hat{h}^c} \end{cases}$$

where $\boldsymbol{\hat{h} = \{i : G_i \hat{u} = 0\}}$

Theorem 3.2 (General Ψ)

$\boldsymbol{\mathcal{F}_v}$ *has a (local) minimum at* $\boldsymbol{\hat{u}}$

- $\delta\mathcal{F}_v(\hat{u})(w) > 0,\ \forall w \in K_{\hat{h}}^{\perp} \setminus \{0\}$
- $\mathcal{U}_{\hat{h}}$—*local minimizer function for* $\mathcal{F}_v|_{K_{\hat{h}}}$*, continuous at* $\boldsymbol{v}$

$\Rightarrow$ *same conclusion (Theorem 3.1)*

$\hat{h} \subset \{1, .., r\}$

$$\boxed{\mathcal{O}_{\hat{h}} \stackrel{\text{def}}{=} \{v \in \mathbf{R}^q : G_i \mathcal{U}(v) = 0, \ \forall i \in \hat{h}\}}$$

Theorems 3.1-3.2 $\Rightarrow$ $\mathcal{L}^q(\mathcal{O}_{\hat{h}}) > 0$ $\Rightarrow$ noisy data do come across $\mathcal{O}_{\hat{h}}$

Data v yield (local) minimizers $\hat{u}$ of $\mathcal{F}_v$ such that $G_i \hat{u} = 0$ for a set of indexes $\hat{h}$

$G_i \approx \nabla_{u_i}$ $\Rightarrow$ $\hat{h}$ =constant regions $\Rightarrow$ **stair-casing:** $\hat{u}_i = \hat{u}_j$ for many neighbors (i, j)

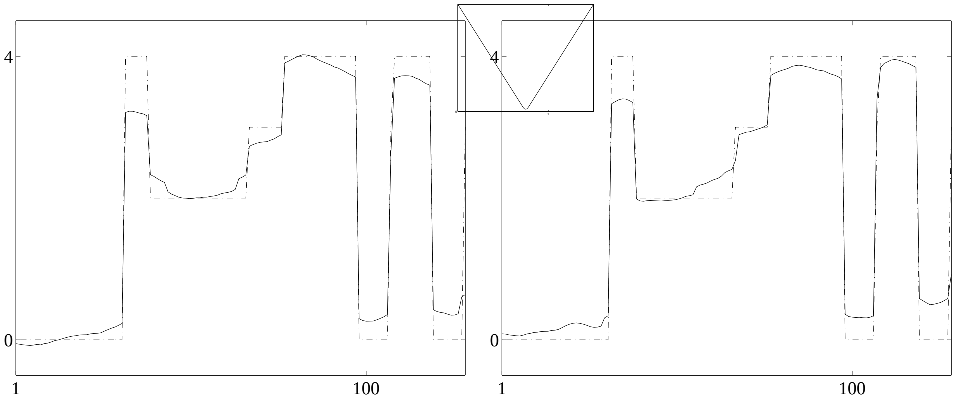
THE SHAPE OF $\mathcal{O}_{\hat{h}}$ FOR 1D SIGNALS

$$\mathcal{F}_v(u) = \|Au - v\|^2 + \beta \sum_{i=1}^{p-1} |u_{i+1} - u_i|, \quad A \in \mathrm{R}^{p \times p} \text{ invertible}$$

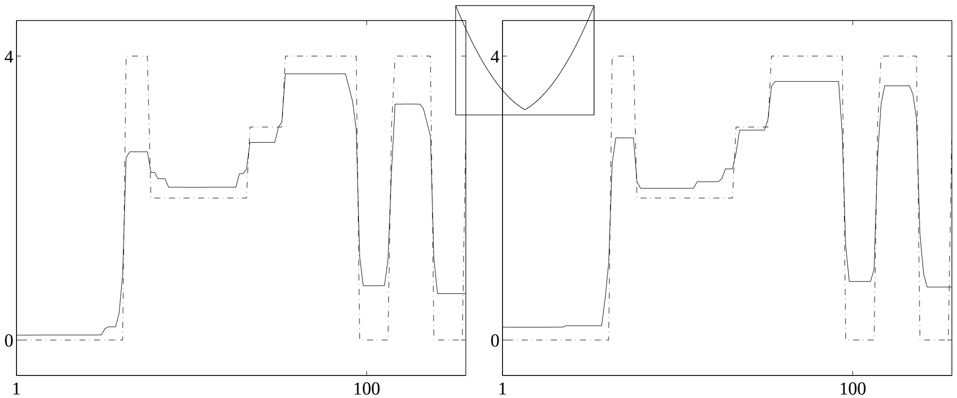
- $\forall \hat{u} \in \mathrm{R}^p$ $\exists W_{\hat{u}}$ polyhedron $\mathbf{dim}(W_{\hat{u}}) = \#\hat{h}$

$$v \in W_{\hat{u}} \quad \Rightarrow \quad \hat{u} = \arg\min_{u \in \mathbf{R}^p} \mathcal{F}_v(u)$$

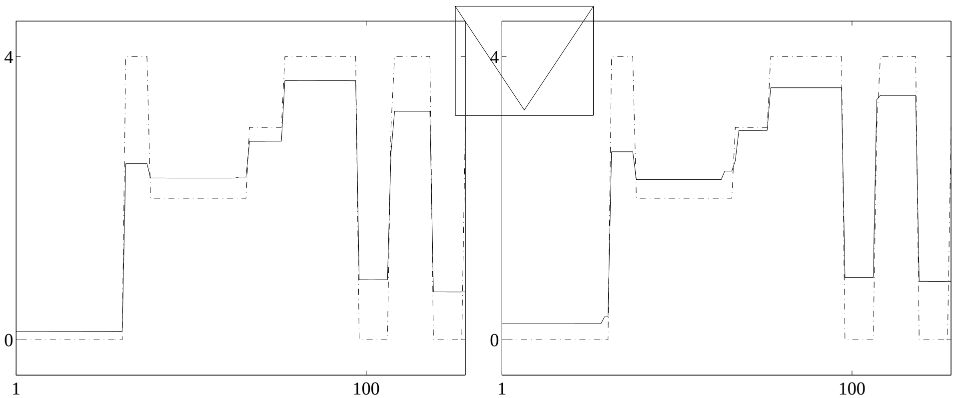
- $\forall \hat{h} \subset \{1, \ldots, p-1\}$ $\quad \mathcal{O}_{\hat{h}} = \bigcup (2^{p-\#\hat{h}-1} \textit{ polyhedra of } \mathrm{R}^q)$, $\quad \mathbf{R}^q = \text{closure}\left(\bigcup_{\hat{h}} \mathcal{O}_{\hat{h}}\right)$



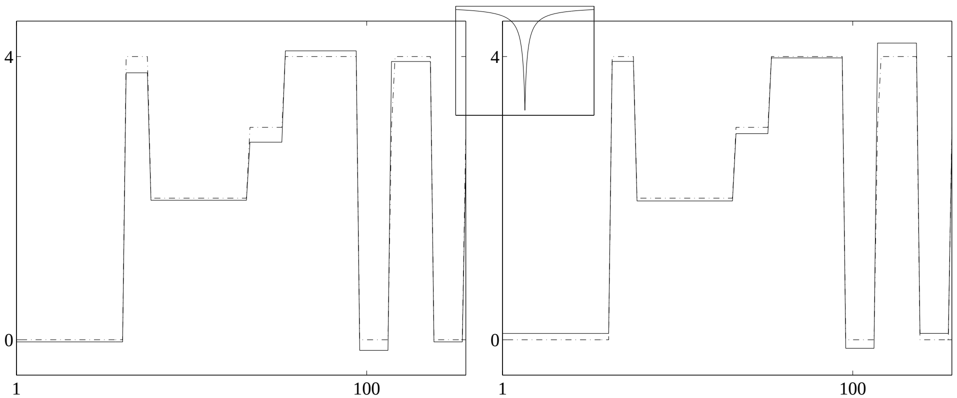
$\varphi(t) = t^2/2$ if $|t| \le \alpha,\ \varphi(t) = \alpha|t| - \alpha^2/2$ else



$\varphi(t) = (t + \alpha \operatorname{sign}(t))^2$



$\varphi(t) = |t|$



$\varphi(t) = \alpha|t|/(1 + \alpha|t|)$

3.2 Illustration: soft thresholding

[*Donoho94*]

$$\boldsymbol{\mathcal{F}_v(u) = \sum_{i=1}^{q}(u_i - v_i)^2 + \beta \sum_{i=1}^{q}|u_i|}$$

$$\min_u \mathcal{F}_v(u) = \sum_{i=1}^{q} \min_{u_i} \left\{(u_i - v_i)^2 + \beta|u_i|\right\}$$

$$\mathcal{U}_i(v) = \begin{cases} 0 & \text{if} \quad |v_i| \leq \frac{\beta}{2} \\ v_i - \frac{\beta}{2}\operatorname{sign}(v_i) & \text{if} \quad |v_i| > \frac{\beta}{2} \end{cases} \qquad \Rightarrow \qquad \hat{h} = \{i : \mathcal{U}_i(v) = 0\} = \left\{i : |v_i| \leq \frac{\beta}{2}\right\}$$

- $\delta\mathcal{F}_v(u)(e_i) = 2(u_i - v_i) + \beta\operatorname{sign}(u_i)$ $\quad$ $\{e_i\}$ is the canonical basis of R^q
- $\boldsymbol{\mathcal{O}_h = \left\{v \in \mathbf{R}^q : \ |v_i| \leq \frac{\beta}{2}, \forall i \in h \ \text{ and } \ |v_i| > \frac{\beta}{2}, \forall i \in h^c\right\}}$
  - $\boldsymbol{\mathcal{L}^q(\mathcal{O}_h) > 0, \ \forall h}$
  - $\boldsymbol{\{ \ \mathcal{O}_h : h \subset \{1,..,q\} \ \}}$ is a partition of R^q
- $\boldsymbol{\hat{u}}$ is sparse

3.3 An application: recovery of Quasi-Binary images

[Nikolova 98]

- The sought image is binary (e.g. document) $\boldsymbol{u \in \{0,1\}^p}$, Data $\boldsymbol{v = u + n}$

- Classical approach: Binary Markov models $\Rightarrow$ calculation troubles
  (direct calculation is infeasible, SA approximation is costly, ICM yields poor solutions)

  Surrogate methods (convex criteria, median filtering) unsatisfactory

- Instead, define *continuous-valued quasi-binary minimizers*

  of <u>*convex*</u> energies which $\begin{cases} \text{discourage nonbinary values} \\ \text{enforce stair-casing} \end{cases}$

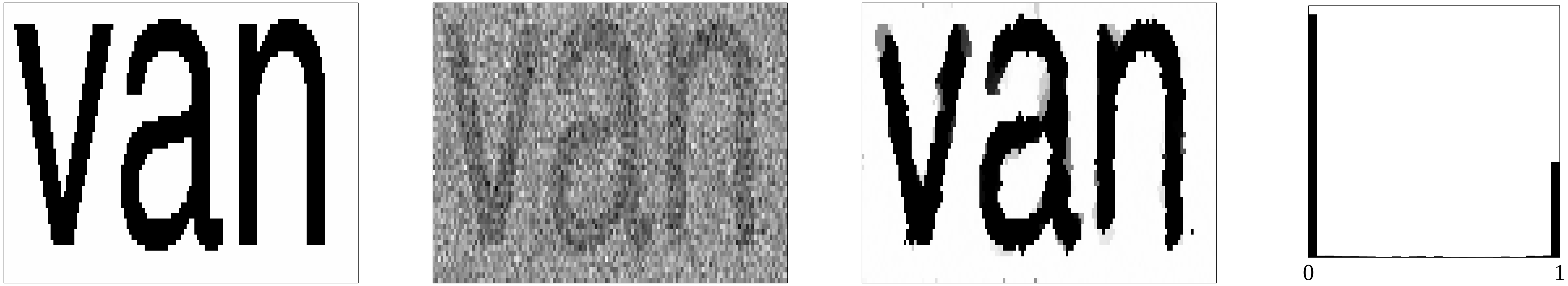
  minimize $\boldsymbol{\mathcal{F}_v}$ subject to $\boldsymbol{u \in [0,1]^p}$ (convex constraint)

$$\boldsymbol{\mathcal{F}_v(u) \quad = \quad \sum_i (u_i - v_i)^2 - \gamma \sum_i \left(u_i - \tfrac{1}{2}\right)^2 + \beta \sum_{i\sim j} \left|u_i - u_j\right|, \quad \gamma \lessapprox 1}$$

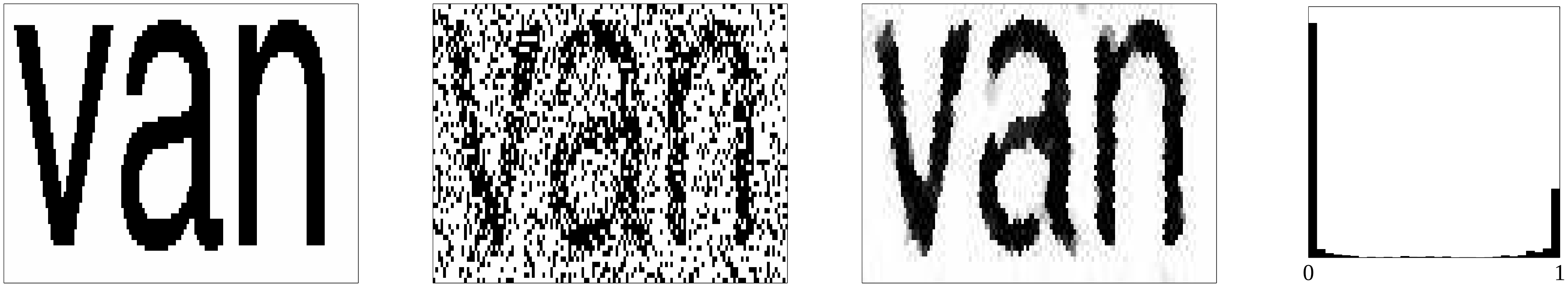
  where $\boldsymbol{i \sim j}$ means that $\boldsymbol{i}$ and $\boldsymbol{j}$ are neighbors

- Extends to $\boldsymbol{v = Au + n}$ if $\mathbf{rank}\boldsymbol{(A) = p}$

  Applications for blind channel estimation *[Alberge et al 02, 06]*



Original image u Data=u+Gaussian noise Proposed method Histogram(solution)



Original image u Data $=u$+salt &pepper Proposed method Histogram(solution)

3.4 Total Variation (TV) regularization

Image $u : \Omega \subset \mathrm{R}^2 \to \mathrm{R}$ is of *bounded variation* (BV) if $\mathbf{TV}(\boldsymbol{u}) < \infty$

(Total variation) $$\mathbf{TV}(\boldsymbol{u}) = \int_\Omega |Du| = \sup\left\{ \int_\Omega u \ \mathrm{div}\xi \ : \ \xi \in \mathcal{C}_c^1(\Omega), |\xi| \le 1 \right\}$$

Coarea formula $$= \int_{\mathbf{R}} \mathbf{Perimeter}\Big(\boldsymbol{E_u(t)}\Big)\boldsymbol{dt} \qquad \boldsymbol{E_u(t)} = \Big\{\boldsymbol{x} \in \boldsymbol{\Omega} : \boldsymbol{u(x)} \ge \boldsymbol{t}\Big\}$$

the $\boldsymbol{t}$-level set of $\boldsymbol{u}$

If $\boldsymbol{u}$ differentiable $\mathbf{TV}(\boldsymbol{u}) = \int_{\boldsymbol{\Omega}} \|\boldsymbol{\nabla u}\|$

BV = the space of functions of bounded variation

BV is often a reasonable functional model—recovers edges (discontinuities) and TV is convex

- *Natural images do not belong to BV* [Gousseau &Morel 2001]

Since [Rudin,Osher &Fatemi 92] TV is a very popular regularization : $\boldsymbol{\Phi(u)} = \mathbf{TV}(\boldsymbol{u})$

$$\boldsymbol{\mathcal{F}_v(u)} = \|\boldsymbol{u} - \boldsymbol{v}\|_{\mathbf{2}}^{\mathbf{2}} + \boldsymbol{\beta}\ \mathbf{TV}(\boldsymbol{u})$$

SEVERAL FACTS ABOUT TV REGULARIZATION

- Full description of $\hat{\boldsymbol{u}}$ when $\boldsymbol{v} = \mathbb{1}_{\boldsymbol{\Sigma}}$, $\boldsymbol{\Sigma} \subset \boldsymbol{\Omega}$ convex *[Alter, Caselles &Novaga05]*

$$\mathbb{1}_{\Sigma}(x) = \begin{cases} 1 & \text{if } x \in \Sigma \\ 0 & \text{else} \end{cases} \qquad \text{(the characteristic function of } \Sigma\text{)}$$

- New stability result *[Caselles, Chambolle &Novaga07]*

$$\Big\{\textit{edges of } \hat{\boldsymbol{u}}\Big\} \subseteq \Big\{\textit{edges of } \boldsymbol{v}\Big\}$$

- Discrete equivalents for an image $u \in \mathrm{R}^{m \times n}$

$$\mathrm{TV}_1(u) = \sum_{i,j} \Big(\big|u_{i,j} - u_{i+1,j}\big| + \big|u_{i,j} - u_{i,j+1}\big| \Big) = \sum_{i,j} \|\nabla_{i,j} u\|_1$$

$$\mathrm{TV}_2(u) = \sum_{i,j} \sqrt{(u_{i,j} - u_{i+1,j})^2 + (u_{i,j} - u_{i,j+1})^2} = \sum_{i,j} \|\nabla_{i,j} u\|_2$$

For a signal $u \in \mathrm{R}^p$, $\mathrm{TV}(u) = \sum_i \big|u_i - u_{i+1}\big|$

  - TV_1 is a classical Markovian enery *(median-pixel prior)* *[Besag 89]*
  - TV_2 is rotation invariant, TV_1 is easier for computation

- Theorems 3.1-3.2 apply with $\hat{\boldsymbol{h}} =$ the constant zones in $\hat{\boldsymbol{u}}$ $\Rightarrow$ sparsity

- Quantitative characterization of sparsity *[Malgouyres 06, 07]*
in connection with the data distribution

- Since *[Meyer 01]* texture restoration *[Vese &Osher 03, Aujol &Chambolle 05,...]*

$$\text{minimize } \|\boldsymbol{w}\|_G \quad + \quad \boldsymbol{\beta}\mathbf{TV}(\boldsymbol{u}) \quad \text{subject to} \quad \boldsymbol{v} = \boldsymbol{u} + \boldsymbol{w} \;=\; \textbf{cartoon} + \textbf{texture}$$

$$\|\boldsymbol{w}\|_G \stackrel{\text{def}}{=} \mathbf{inf}\left\{\|\boldsymbol{g}\|_\infty : \boldsymbol{w} = \mathbf{div}(\boldsymbol{g})\right\} \quad \text{oscillating patterns (like texture)}$$

- How to avoid stair-casing (when undesirable)
  - regularize TV, e.g. $\sqrt{\alpha + (\mathrm{TV}(u))^2}, \quad \alpha \stackrel{>}{\approx} 0$
  - Minimize $\Psi(u,v)$ subject to $\mathrm{TV}(u) \leq \tau$ *[Combettes &Pesquet04]*
  - Iterative regularization using Bregman distance (d_B) *[Osher at al.05]*

$$u^{(1)} = \arg\min_u \left\{\|u-v\|^2 + \beta\mathrm{TV}(u)\right\}$$

for $k = 2, \ldots, K$: $$u^{(k)} = \arg\min_u \left\{\|u-v\|^2 + \beta d_B(u, u^{(k-1)})\right\}$$

$$d_B(u,v) = \mathrm{TV}(u) - \mathrm{TV}(v) - \langle g, u-v\rangle, \;\; g \in \partial\mathrm{TV}(v)$$

- Equivalence results for signals $\boldsymbol{u} \in \mathbf{R}^{\boldsymbol{p}}$ *[Steidl, Weickert, Mrázek & Welk 04]*

Space discrete TV diffusion = TV regularization, $\boldsymbol{\beta}$ = time step

Soft Haar wavelet shrinkage and TV diffusion on two-pixel signals are equivalent

These equivalences are false in general

4 Minimizers relevant to non-smooth data-fidelity

4.1 Main result

[Nikolova 01,02]

$\psi \in \mathcal{C}^m$ and $\psi'(0) > 0$, $\Phi \in \mathcal{C}^m$ $\qquad \mathcal{F}_v(u) = \sum_{i=1}^{q} \psi(|\langle a_i, u\rangle - v_i|) + \beta\Phi(u)$

Example: $\psi(t) = t$ *[Alliney 94,97]*

Theorem 4.1

If $\mathcal{F}_v$ *has a strict (local) minimum at* $\hat{u}$, *denote* $\hat{h} = \{i : \langle a_i, \hat{u}\rangle = v_i\}$

$$\mathcal{K}_{\hat{h}}(v) = \left\{u \in \mathbf{R}^p : \langle a_i, u\rangle = v_i \ \forall i \in \hat{h}\right\}$$

$$K_{\hat{h}} = \left\{u \in \mathbf{R}^p : \langle a_i, u\rangle = 0 \ \forall i \in \hat{h}\right\}$$

(the tangent of $\mathcal{K}_{\hat{h}}(v)$)

- $\nabla \mathcal{F}_v|_{\mathcal{K}_{\hat{h}}(v)}(\hat{u}) = 0$ *and* $\nabla^2 \mathcal{F}_v|_{\mathcal{K}_{\hat{h}}(v)}(\hat{u}) > 0$
- $\delta \mathcal{F}_v(\hat{u})(w) > 0, \quad \forall w \in K_{\hat{h}}^{\perp} \setminus \{0\}$

$\Rightarrow$ $\exists\, O \subset \mathrm{R}^q$ *open*, $v \in O$ *and* $\exists\, \mathcal{U} \sim \mathcal{C}^{m-1}$ *such that*

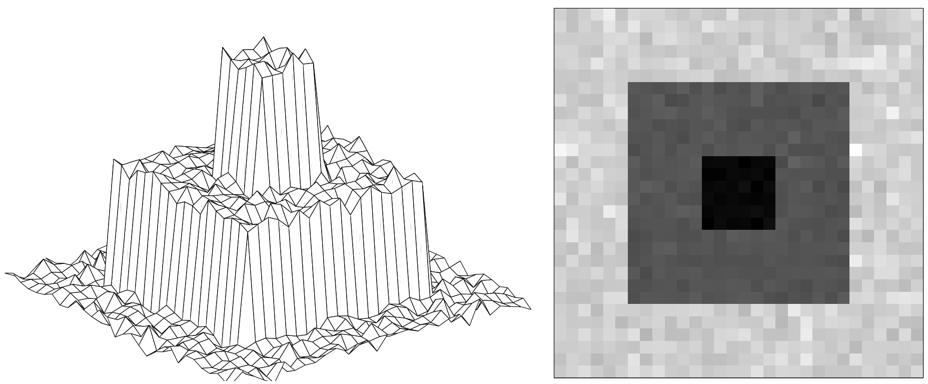
$v' \in O \ \Rightarrow\ \mathcal{F}_{v'}$ *has a minimum at* $\hat{u}' = \mathcal{U}(v')$ *and* $\begin{cases} \langle a_i, \hat{u}'\rangle = v_i & \text{if } i \in \hat{h} \\ \langle a_i, \hat{u}'\rangle \neq v_i & \text{if } i \in \hat{h}^c \end{cases}$

- Weak assumptions—details in *[Nikolova 02]*

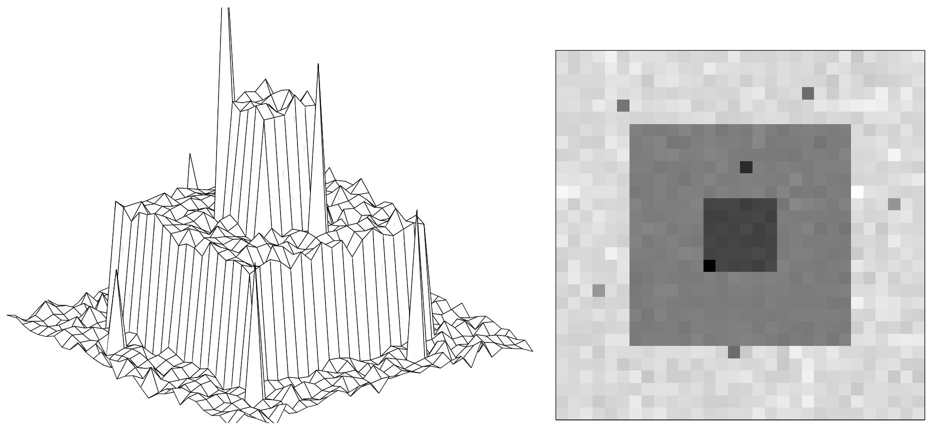
- For $\hat{h} \subset \{1, .., q\}$ and $\mathcal{U}$, define $\boxed{\mathcal{O}_{\hat{h}} = \left\{v \in \mathbf{R}^q : \langle a_i, \mathcal{U}(v)\rangle = v_i, \forall i \in \hat{h}\right\}}$

Theorem 4.1 $\Rightarrow$ $\mathcal{L}^q(\mathcal{O}_{\hat{h}}) > 0$ $\Rightarrow$ noisy data do come across $\mathcal{O}_{\hat{h}}$

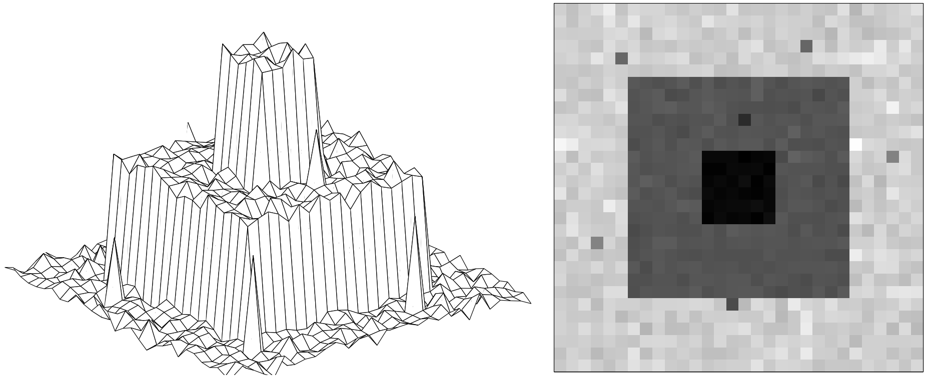
$\Rightarrow$ Noisy data v yield (local) minimizers $\hat{u}$ of $\mathcal{F}_v$ which achieve an **exact fit to data** $\langle a_i, \hat{u}\rangle = v_i$ for a certain number of indexes i



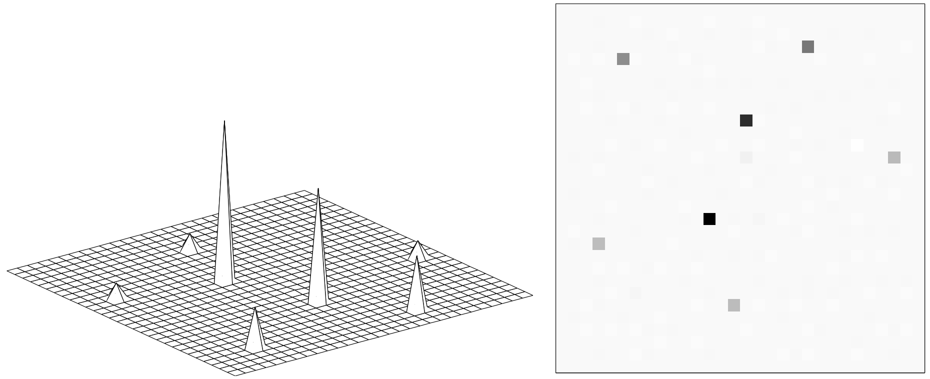
Original u_o



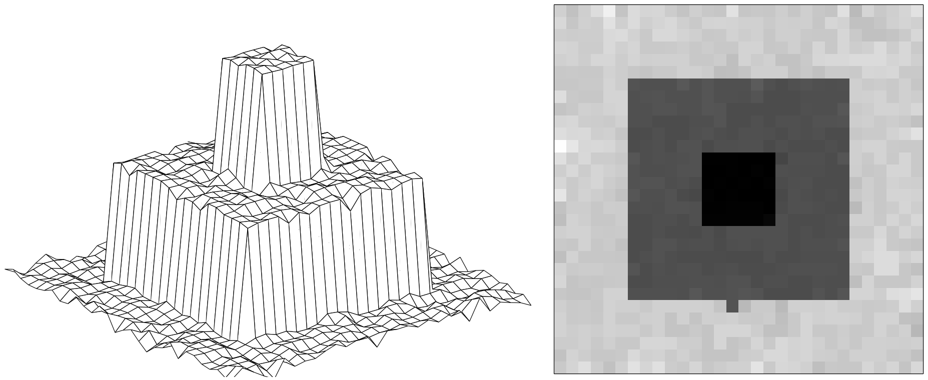
Data $v = u_o$+outliers



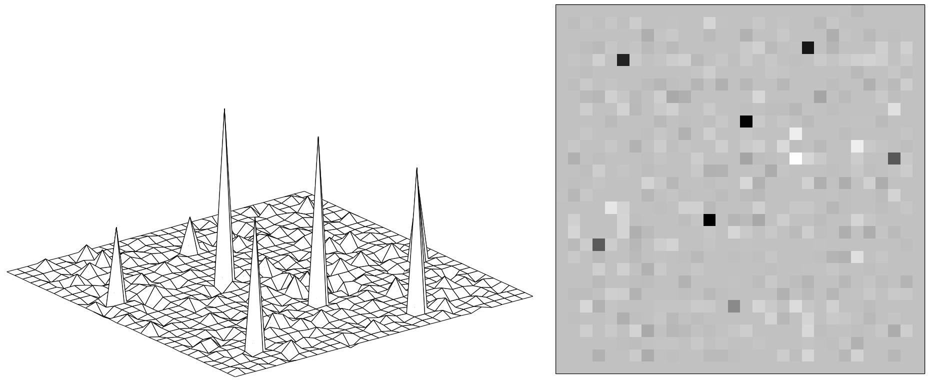
Restoration $\hat{u}$ for $\boldsymbol{\beta = 0.14}$



Residuals $v - \hat{u}$

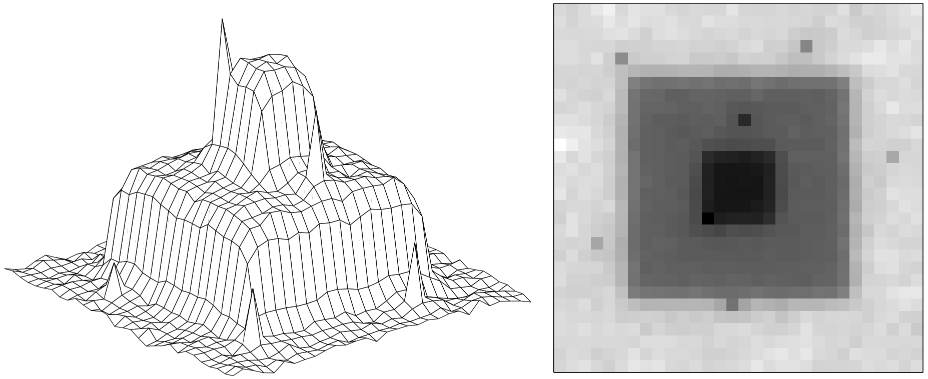


Restoration $\hat{u}$ for $\boldsymbol{\beta = 0.25}$

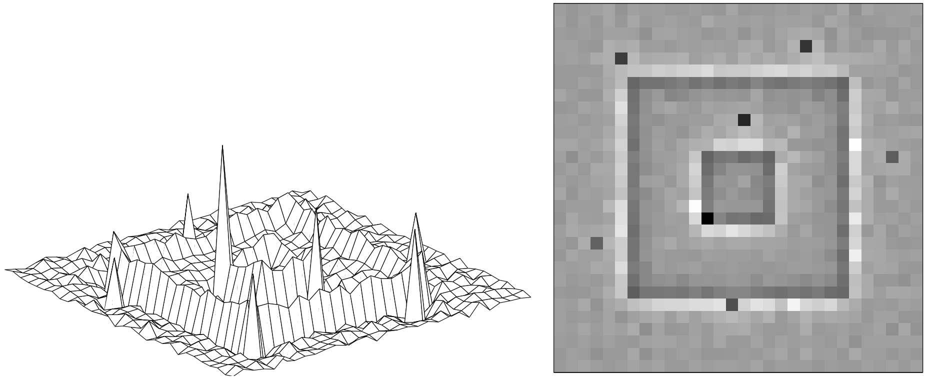


Residuals $v - \hat{u}$

$$\boldsymbol{\mathcal{F}_v(u) = \sum_i \Big|u_i - v_i\Big| + \beta \sum_{i \sim j} |u_i - u_j|^{1.1}}$$

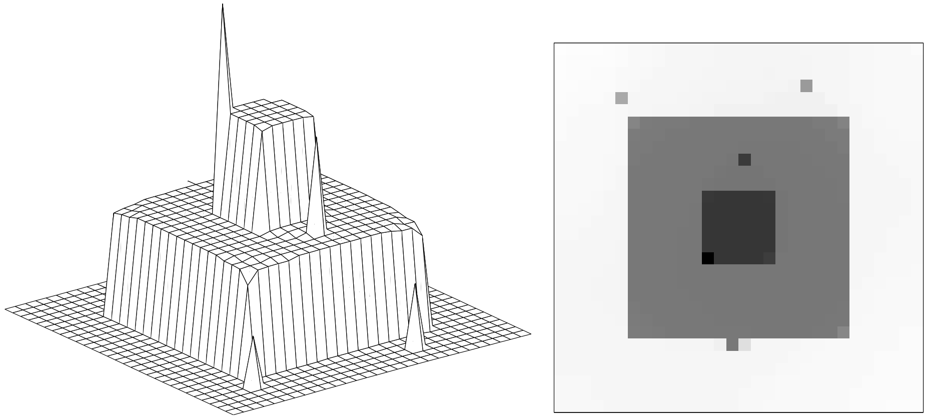


Restoration $\hat{u}$ for $\beta = 0.2$

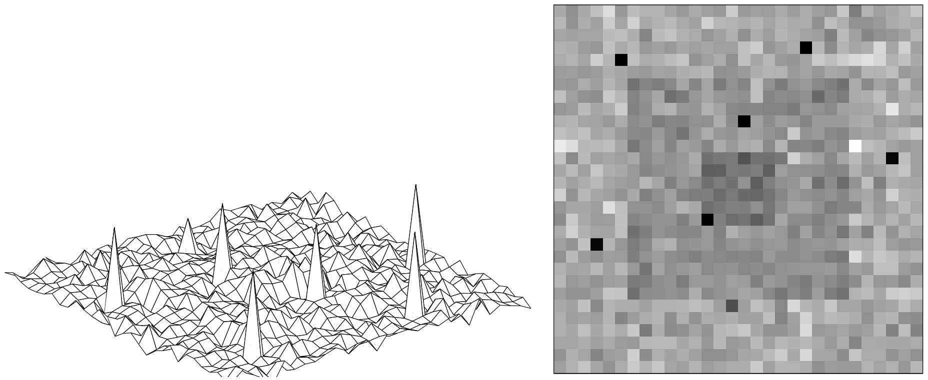


Residuals $v - \hat{u}$

Smooth energy: $\boldsymbol{\mathcal{F}_v(u) = \sum_i \big(u_i - v_i\big)^2 + \beta \sum_{i \sim j} (u_i - u_j)^2}$



Restoration $\hat{u}$ for $\beta = 0.2$



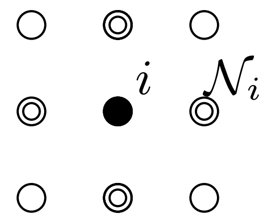
Residuals $v - \hat{u}$

Non-smooth regularization: $\boldsymbol{\mathcal{F}_v(u) = \sum_i (u_i - v_i)^2 + \beta \sum_{i \sim j} |u_i - u_j|}$

4.2 Detection and cleaning of outliers using ℓ_1 data-fidelity

[Nikolova 04]

$$\boxed{\mathcal{F}_v(u) = \sum_{i=1}^{p} |u_i - v_i| + \frac{\beta}{2} \sum_{i=1}^{p} \sum_{j \in \mathcal{N}_i} \varphi(|u_i - u_j|)}$$



φ: smooth, convex, *edge-preserving*

Assumptions:
$$\begin{cases} \text{data } v \text{ contain uncorrupted samples } v_i \\ v_i \text{ is outlier if } \quad |v_i - v_j| \gg 0, \quad \forall j \in \mathcal{N}_i \end{cases}$$

$$v \in \mathbb{R}^p \quad \Rightarrow \quad \hat{u} = \arg\min_u \mathcal{F}_v(u), \qquad \hat{h} = \{i : \hat{u}_i = v_i\}$$

$$\begin{cases} v_i \text{ is regular} & \text{if } i \in \hat{h} \\ v_i \text{ is outlier} & \text{if } i \in \hat{h}^c \end{cases}$$

Outlier detector:	$v \to \hat{h}^c(v)$	$=$	$\{i \,:\, \hat{u}_i \neq v_i\}$
Smoothing:	$\hat{u}_i$ *for* $i \in \hat{h}^c$	$=$	*estimate of the original* $u_{o\,i}$

Justification based on the properties of $\hat{\boldsymbol{u}}$

- Regular data samples are fitted exactly $\quad (\hat{u}_i = v_i$ for $i \in \hat{h})$
- $\boldsymbol{v_i}$ is an outlier if it is too dissimilar with respect to its (recovered) neighbors

  $\boldsymbol{\omega}$ patch of outliers—$\hat{\boldsymbol{u}}_{\boldsymbol{\omega}}$ depends *only* on the regular samples surrounding $\boldsymbol{\omega}$
- The same set of outliers $\hat{\boldsymbol{h}}$ if
  - small perturbation on the regular samples
  - arbitrarily large perturbation on outliers

  $\Rightarrow$ stability of the outlier detection
- The solution $\hat{u}$ remains unchanged under arbitrarily large perturbation on outliers
- Remark: $\boldsymbol{\beta} \leq \big(\|\boldsymbol{\nabla \Phi(v)}\|_{\boldsymbol{\infty}}\big)^{\boldsymbol{-1}} \;\Rightarrow\; \hat{\boldsymbol{u}} = \boldsymbol{v}$
- More formal results in *[Nikolova 04]*

Original image u_o

10% random-valued noise

Median ($\|\hat{u}-u_o\|_2=4155$)

Recursive CWM ($\|\hat{u}-u_o\|_2=3566$)

PWM ($\|\hat{u}-u_o\|_2=3984$)

Proposed ($\|\hat{u}-u_o\|_2 = 2934$)

Original image u_o

45% salt-and-pepper noise

Recursive median ($\|\hat{u}-u_o\|_2$=7825)

Recursive CWM ($\|\hat{u}-u_o\|_2=7497$)

PWM ($\|\hat{u}-u_o\|_2=6265$)

Proposed ($\|\hat{u}-u_o\|_2 = 6126$)

4.3 Restoration of frame coefficients using ℓ_1 data-fidelity

[Durand &Nikolova07]

- Noisy data $\quad v = u_o +$ noise
- Noisy frame coefficients $\quad y = Wv = Wu_o +$ noise
- Hard thresholding keeps the relevant information $\quad y_{T\,i} = \begin{cases} 0 & \text{if} \quad |y_i| \le T \\ y_i & \text{if} \quad |y_i| > T \end{cases}$

  $\tilde{u} = W^* y_T$ – Gibbs oscillations and wavelet-shaped artifacts $\qquad W^*$ left inverse

- Hybrid methods—combine y_T with prior $\Phi(u)$

  Different energies *[Bobichon &Bijaoui 97, Coifman &Sowa 00, Durand &Froment 03...]*

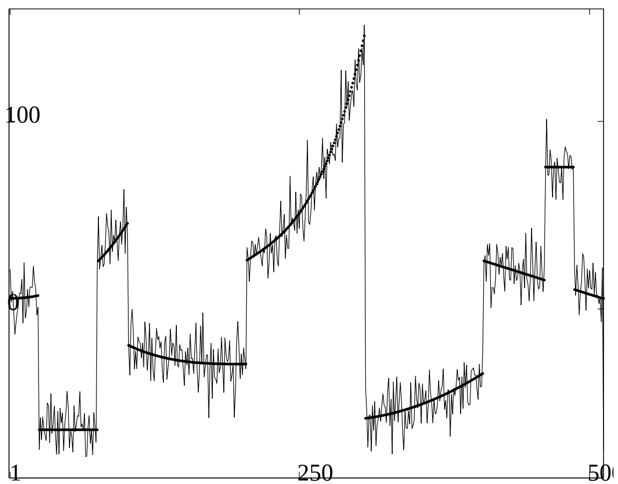
Our choice:

$$\text{minimize} \quad \mathcal{F}_y(x) = \sum_i \lambda_i \big|(x - y_T)_i\big| + \int_\Omega \varphi(|\nabla W^* x|) \qquad (\varphi \approx \mathrm{TV})$$

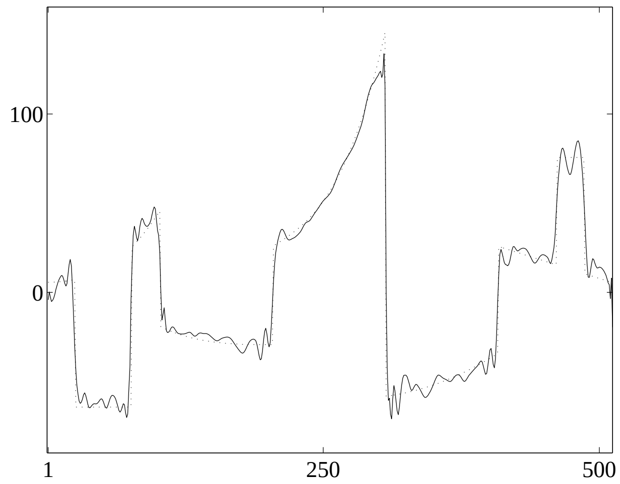
$$\hat{u} = W^* \hat{x}$$

Rationale

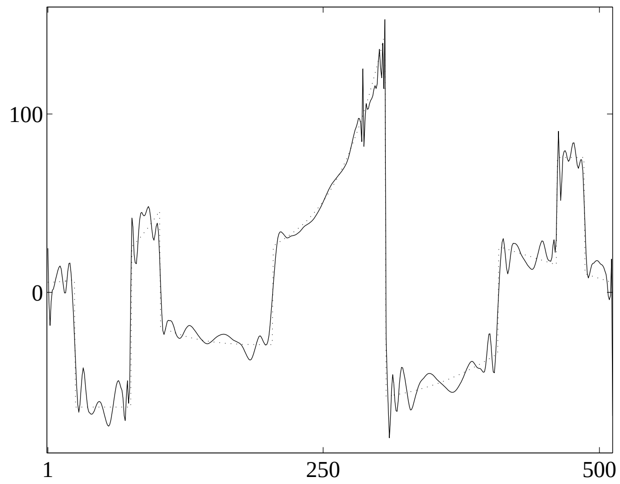
Keep $\hat{x}_i = y_{T\,i}$	Restore $\hat{x}_i \neq y_{T\,i}$	
significant coefs $y_i \approx (Wu_o)_i$	outliers $\quad |y_i| \gg |(Wu_o)_i|$	*(frame-shaped artifacts)*
thresholded coefs if $(Wu_o)_i \approx 0$	edge coefs $|(Wu_o)_i| > |y_{T\,i}| = 0$	*(Gibbs oscillations)*



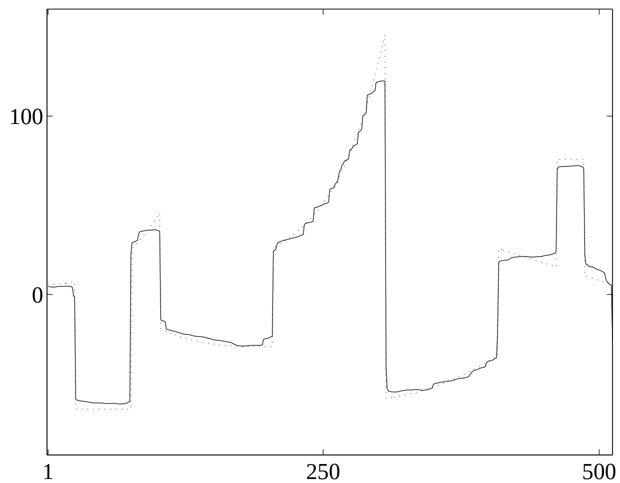
Original and data



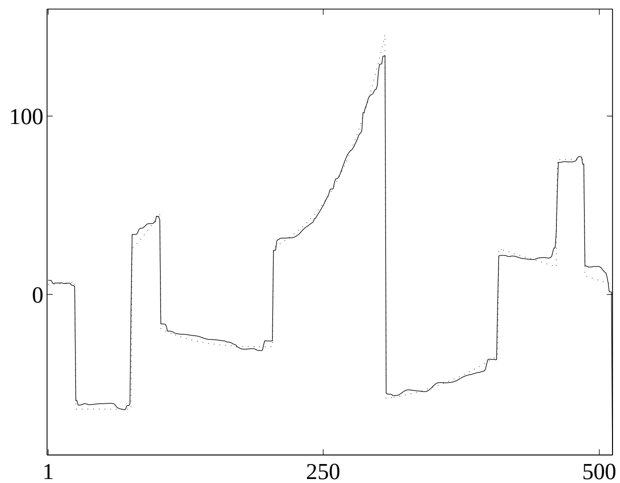
Sure-shrink method



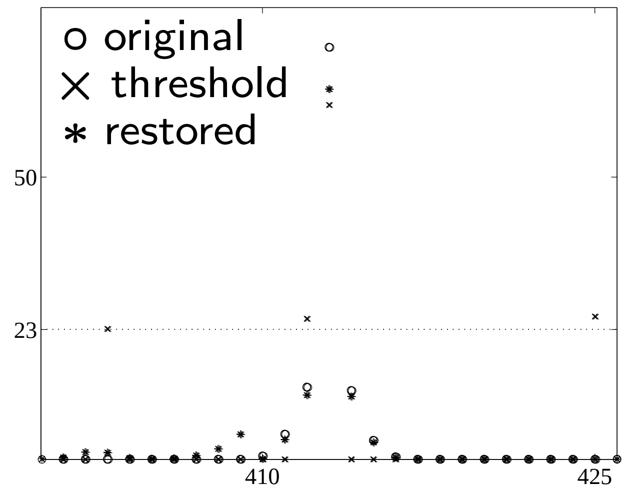
Hard thresholding



Total variation



The proposed method



Magnitude of coefficients

Restored signal (—), original signal (- -).

Original image

Noisy image

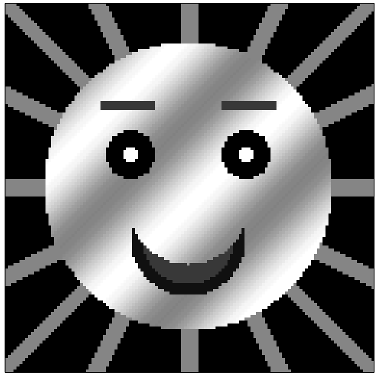
TV regularization

Optimal threshold $T = 100$

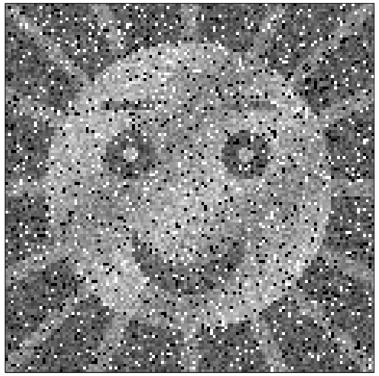
$(\star)$ Our data $T = 50$

Proposed method

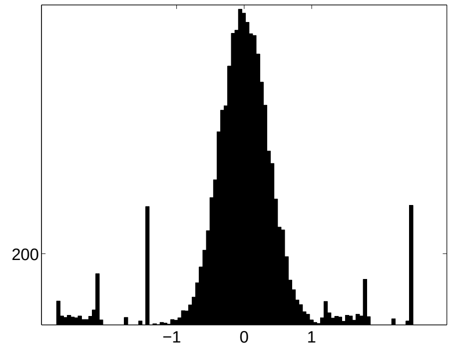
4.4 Fast Cleaning of Noisy Data From Impulse Noise



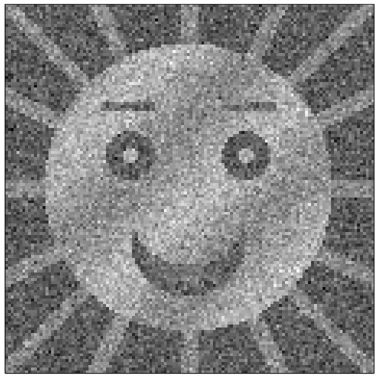
Original image u^*



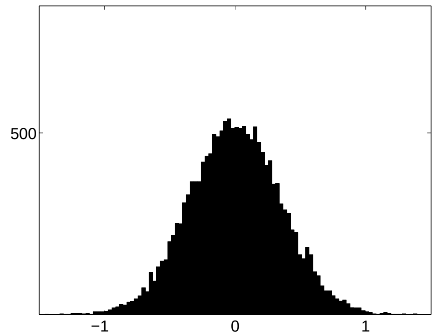
Data $\boldsymbol{v} = \boldsymbol{u_o} + \boldsymbol{\omega}$ (impulse noise)



Histogram of $n + \omega$ (all the noise)



Original data $\boldsymbol{u_o} = \boldsymbol{u}^* + \boldsymbol{n}$



Histogram of $\boldsymbol{n}$

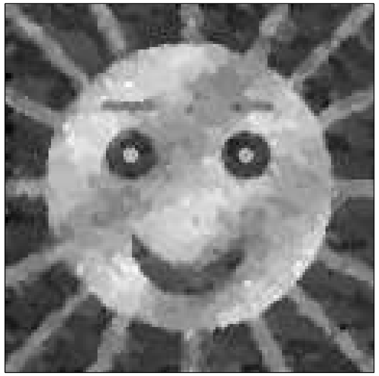
Data u_o &Histogram n

to recover based on v

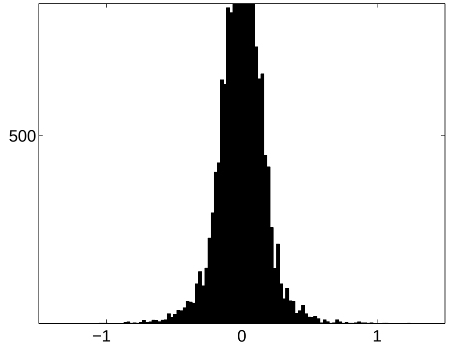
$\Longleftarrow$

$$\text{minimize} \quad \boldsymbol{\mathcal{F}_v(u) = \|u - v\|_1 + \beta\|Gu\|^2}$$

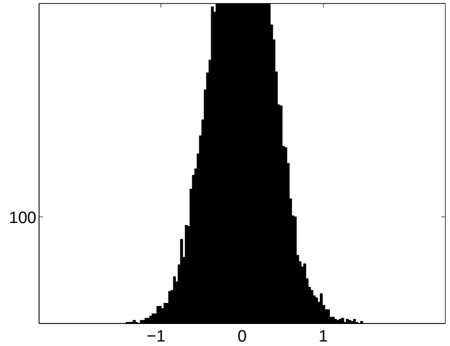
Fast minimization—the scheme in §9.2-A is explicit



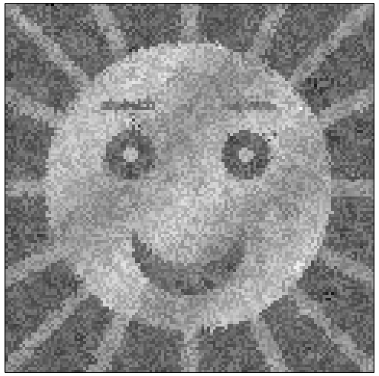
median Filter



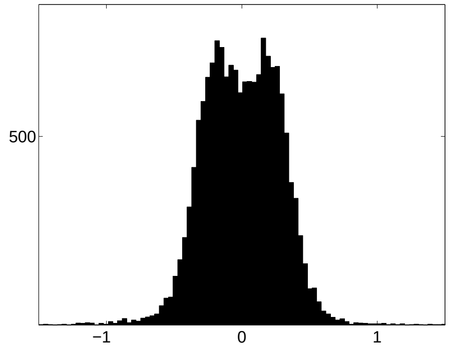
noise estimate $\hat{n} = \hat{u} - u_o$



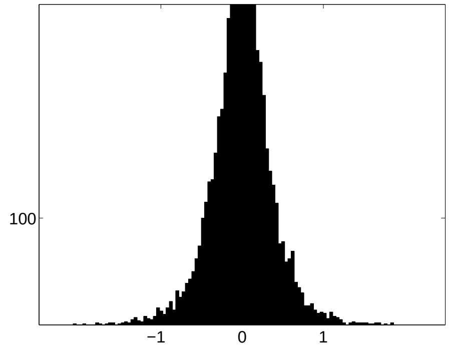
$\|\hat{u} - u_o\|_1 = 43,\ \|.\|_2 = 46$



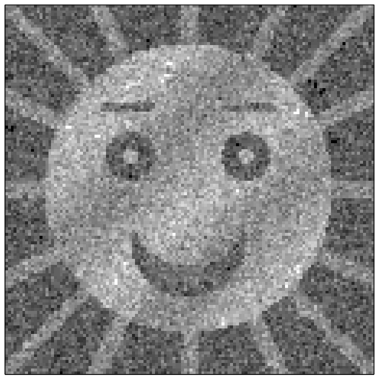
center-weighted median



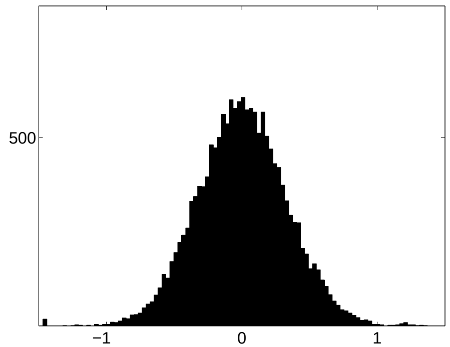
noise estimate $\hat{n} = \hat{u} - u_o$



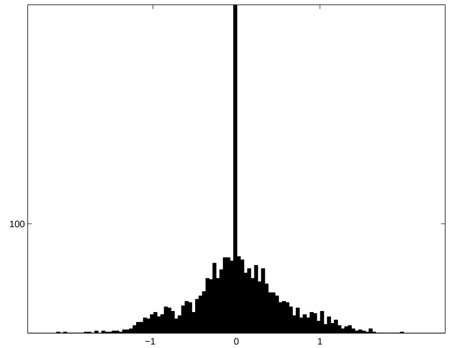
$\|\hat{u} - u_o\|_1 = 22,\ \|.\|_2 = 35$



Proposed method



noise estimate $\hat{n} = \hat{u} - u_o$



$\|\hat{u} - u_o\|_1 = 14,\ \|.\|_2 = 26$

4.5 Practical 2-stage methods to clean impulse noise

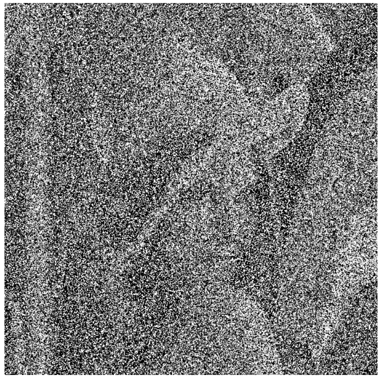
[Chan et al. 04, 05]

Approximate the outlier-detection stage by rank-order filter

$\Rightarrow$ increase speed and accuracy

Methods for denoising (+ deblurring under mixed noise)

- Corrupted pixels $\hat{h}^c = \{i : \hat{v}_i \neq v_i\}$ where $\hat{v}$=Rank-Order Filter (v)
  - salt &pepper (SP) noise by adaptive median
  - Random-valued noise by center-weighted median
- Restore $\left\{\hat{u}_i : i \in \hat{h}^c\right\}$ by an edge-preserving variational method subject to $\hat{u}_i = v_i$ for all $i \in \hat{h}$
  - Fast optimization, pertinent initialization



70%SPnoise(6.7dB)

MED (23.2 dB)

PSM (19.5 dB)

MSM (19.0 dB)

DDBSM (17.5 dB)

NASM (21.8 dB)

ISM filter (23.4 dB)

Adapt.med.(25.8dB)

Variational (24.6)

Our method(29.3dB)

Original Lena

30% random noise

ACWMF with $s = 0.6$

DPVM with $\beta = 0.19$

Our method

50% random noise

ACWMF with $s = 0.3$

DPVM with $\beta = 0.19$

Our method

5. Nonsmooth data-fidelity and regularization

A consequence of §3 and §4: if Φ and Ψ are non-smooth (as specified)

$$\begin{array}{lll} \boldsymbol{G_i \hat{u} = 0} & \text{for} & \boldsymbol{i \in \hat{h}_\varphi \neq \emptyset} \\ \boldsymbol{\langle a_i, \hat{u} \rangle = v_i} & \text{for} & \boldsymbol{i \in \hat{h}_\psi \neq \emptyset} \end{array}$$

5.1 L1 data-fidelity and TV regularization

$$\boldsymbol{\mathcal{F}_v(u) = \|u - v\|_1 + \beta\, TV(u)}$$

A. Restoration of characteristic functions

[Chan &Esedoglu 06]

- The regularization imposed on the solution is more geometric than with $\|u-v\|^2$
- Data $\boldsymbol{v = 1\!\!1_\Omega}$ for $\boldsymbol{\Omega} \subset \mathrm{R}^2$ bounded
- For almost every $\boldsymbol{\beta > 0}$, $\exists \boldsymbol{\Sigma} \subset \boldsymbol{\Omega}$ such that $\boldsymbol{1\!\!1_\Sigma}$ is the **unique** minimizer of $\boldsymbol{\mathcal{F}_v}$
- $\boldsymbol{\beta \longrightarrow \|v - \hat{u}_\beta\|_1}$ is discontinuous (critical values for β)

$\boxed{\text{B. RESTORATION OF BINARY IMAGES}}$

[Chan, Esedoglu &Nikolova 06]

Classical approach to find a binary image $\hat{u} = \mathbb{1}_{\hat{\Sigma}}$ from binary data $\mathbb{1}_{\Omega}$, $\quad \Omega \subset \mathrm{R}^2$

$$\hat{\Sigma} = \arg\min_{\Sigma} \Big\{ \big\| \mathbb{1}_{\Sigma} - \mathbb{1}_{\Omega} \|_2^2 + \beta \mathrm{TV}(\mathbb{1}_{\Sigma}) \Big\}$$

nonconvex problem

$$= \arg\min_{\Sigma} \{ \mathsf{Surface}(\Sigma \,\Delta\, \Omega) + \beta \mathsf{Per}(\Sigma) \}$$

(symmetric difference)

usual techniques (curve evolution, level-sets) fail

Instead—*the convex problem*

$$\boxed{\hat{\boldsymbol{u}} = \arg\min_{\boldsymbol{u}} \Big\{ \|\boldsymbol{u} - \mathbb{1}_{\boldsymbol{\Omega}}\|_1 + \boldsymbol{\beta} \mathbf{TV}(\boldsymbol{u}) \Big\}}$$

is solved for $\hat{\boldsymbol{u}} = \mathbb{1}_{\boldsymbol{\Sigma}}$

$\Rightarrow$ Algorithm for finding the global minimum

$\boxed{\text{C. DISCRETE FORMULATIONS}}$

- $\mathcal{F}_v^{\rho}(u) = \sum_{i,j} \Big(|u_{i,j} - v_{i,j}| + \beta \|\nabla_{i,j} u\|_{\rho} \Big), \quad \rho = 1$ or $\rho = 2$ (see §3.4)
- minimize $\boldsymbol{\mathcal{F}_v^1(u)}$ *yields a morphological filter* *[Darbon-Sigelle06]*

6 Comparison with Smooth Energies

[Nikolova 04]

Theorem

$\boldsymbol{\mathcal{F}_v(u) = \Psi(u,v) + \beta\Phi(u)}$, $\boldsymbol{\mathcal{F}} \in \mathcal{C}^{m\geq 2}$ *+ assumptions.* *If* $\boldsymbol{h} \neq \emptyset \;\Rightarrow$

$\{\boldsymbol{v} \in \mathrm{R}^q : \mathcal{F}_v\text{—}\textit{minimum at } \hat{\boldsymbol{u}},\ \boldsymbol{G_i \hat{u} = 0},\ \forall \boldsymbol{i} \in \boldsymbol{h}\}$ **closed and**

$\{\boldsymbol{v} \in \mathrm{R}^q : \mathcal{F}_v\text{—}\textit{minimum at } \hat{\boldsymbol{u}},\ \langle \boldsymbol{a_i}, \hat{\boldsymbol{u}}\rangle = \boldsymbol{v_i},\ \forall \boldsymbol{i} \in \boldsymbol{h}\}$ **negligible in** R^q

- The chance that noisy data come across $\mathcal{O}_h$, for any $h \neq \emptyset$, is null

The chance that noisy data $\boldsymbol{v}$ yield a minimizer $\hat{\boldsymbol{u}}$ of $\boldsymbol{\mathcal{F}_v}$ which satisfies <u>*exactly*</u> $\boldsymbol{G_i \hat{u} = 0}$, or $\langle \boldsymbol{a_i}, \hat{\boldsymbol{u}}\rangle = \boldsymbol{v_i}$, for some $\boldsymbol{i}$, is <u>*null*</u>

- Almost all $v \in \mathrm{R}^q$ lead to $\hat{u} = \mathcal{U}(v)$ satisfying $\boldsymbol{G_i \hat{u} \neq 0}$, $\forall \boldsymbol{i}$ and $\langle \boldsymbol{a_i}, \hat{\boldsymbol{u}}\rangle \neq \boldsymbol{v_i}$, $\forall \boldsymbol{i}$

`Salient features of the minimizers of smooth energies are tricky to obtain`

7 Non-convex regularization

[Nikolova 04]

7.1 Edge enhancement

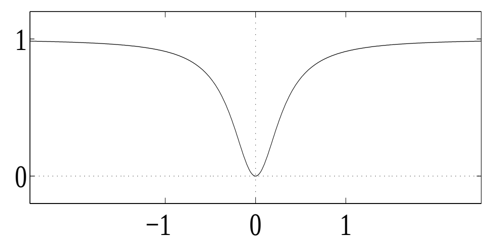
$$\mathcal{F}_v(u) = \|Au - v\|^2 + \beta \sum_{i=1}^{r} \varphi(\|G_i u\|)$$

<u>*Assumptions on φ:*</u> $\mathcal{C}^2$ on $\mathbb{R}_+$ and $\lim_{t\to\infty} \varphi''(t) = 0$

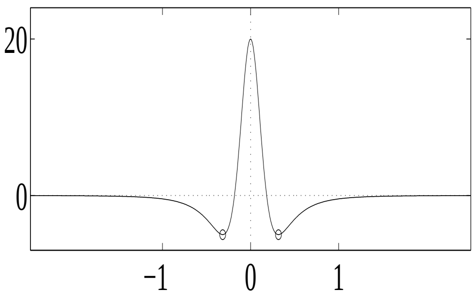
<u>$\varphi'(0) = 0$ (Φ is smooth)</u>

$$0 < \tau < \mathcal{T} \;:\; \varphi'' \begin{cases} \geq 0 \text{ on} & [0, \tau] \\ < 0 \text{ decrease} & (\tau, \mathcal{T}] \\ \leq 0 \text{ increase} & [\mathcal{T}, \infty) \end{cases}$$

$\varphi(t) = \frac{\alpha t^2}{1+\alpha t^2}$



$\varphi''(t)$

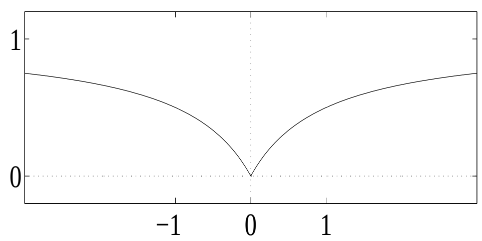


<u>$\varphi'(0) > 0$ (Φ is nonsmooth)</u>

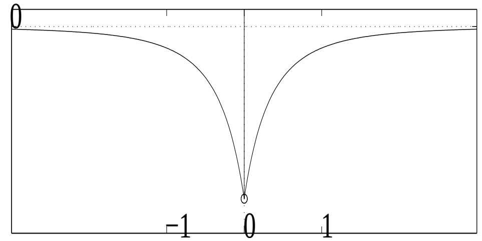
$\varphi''(0^+) < 0$

$\varphi'' \leq 0$ increase on $(0, \infty)$

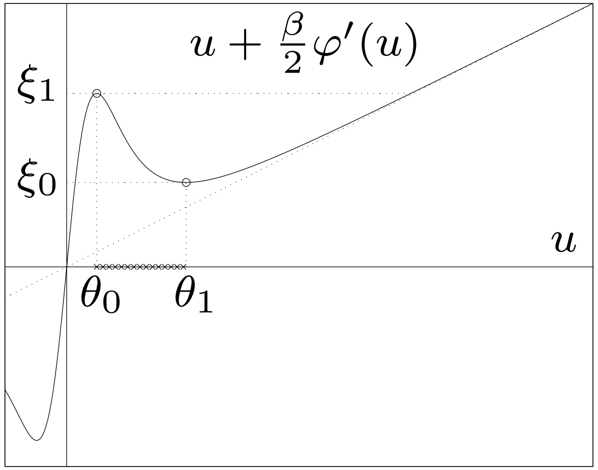
$\varphi(t) = \frac{\alpha|t|}{1+\alpha|t|}$



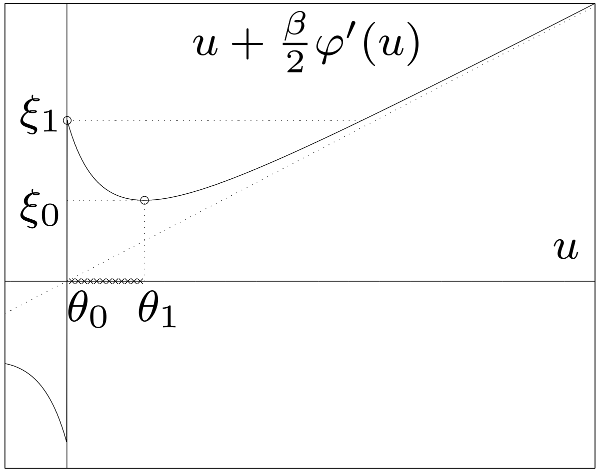
$\varphi''(t)$



$\boxed{\text{A. ILLUSTRATION ON R}}$ $\qquad$ $\boldsymbol{\mathcal{F}_v(u) = (u-v)^2 + \beta\varphi(|u|), \quad u, v \in \mathbf{R}}$



$\varphi'(0) = 0$



$\varphi'(0) > 0$

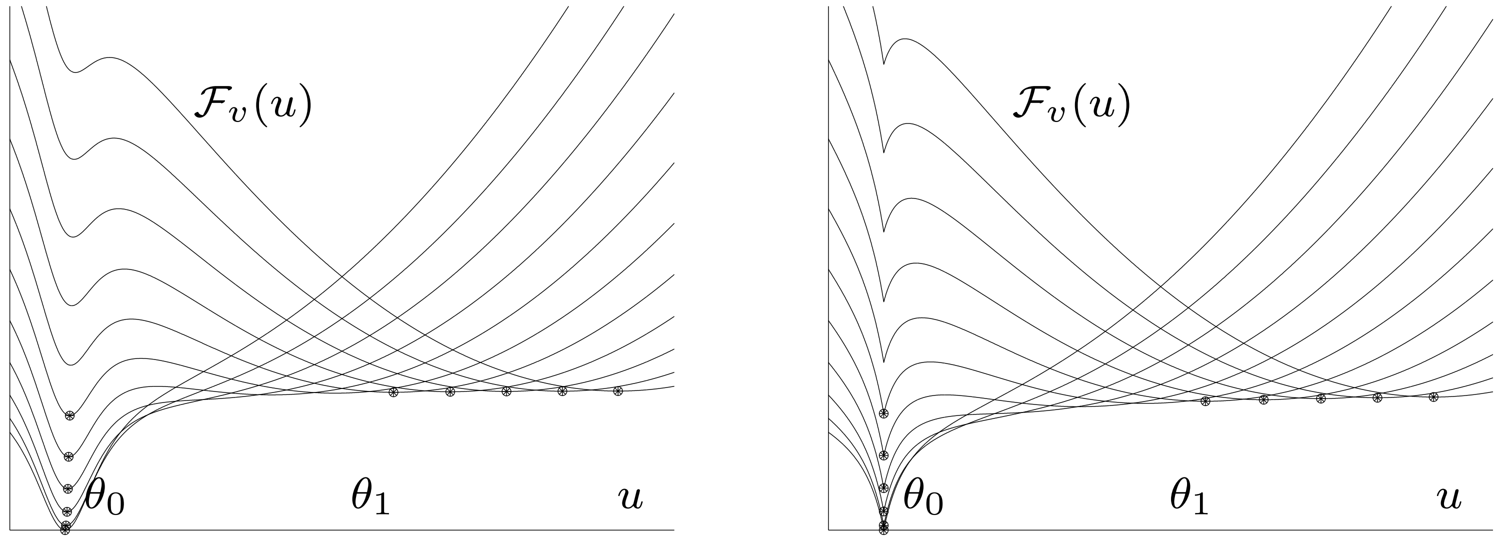
no local minimizer lies in $\boldsymbol{(\theta_0, \theta_1)}$ $= \left\{t > 0 : \varphi''(t) \le -2/\beta\right\}$ $\quad (F_v''(u) < 0)$

$$\exists\ \boldsymbol{\xi_0 > 0}, \quad \exists\ \boldsymbol{\xi_1 > \xi_0}$$

- $|\boldsymbol{v}| \le \boldsymbol{\xi_1}$ $\Rightarrow$ $\exists$ local minimizer $|\hat{\boldsymbol{u}}_{\mathbf{0}}| \le \boldsymbol{\theta_0}$ *(strong smoothing)*
- $|\boldsymbol{v}| \ge \boldsymbol{\xi_0}$ $\Rightarrow$ $\exists$ local minimizer $|\hat{\boldsymbol{u}}_{\mathbf{1}}| \ge \boldsymbol{\theta_1}$ *(loose smoothing)*

- $\exists\, \boldsymbol{\xi} \in (\boldsymbol{\xi_0}, \boldsymbol{\xi_1})$ $\quad |\boldsymbol{v}| \leq \boldsymbol{\xi} \;\Rightarrow\;$ global minimizer $= \hat{\boldsymbol{u}}_{\mathbf{0}}$ *(strong smoothing)*

  $|\boldsymbol{v}| \geq \boldsymbol{\xi} \;\Rightarrow\;$ global minimizer $= \hat{\boldsymbol{u}}_{\mathbf{1}}$ *(loose smoothing)*



$\mathcal{F}_v(u) = (u-v)^2 + \beta \frac{\alpha u^2}{(1+\alpha u^2)}$ $\qquad$ $\mathcal{F}_v(u) = (u-v)^2 + \beta \frac{\alpha |u|}{(1+\alpha |u|)}$

Each curve $\boldsymbol{u} \to \boldsymbol{\mathcal{F}_v(u)}$ for a different $\boldsymbol{v}$

For $\boldsymbol{v} = \boldsymbol{\xi}$ *the global minimizer jumps from* $\hat{\boldsymbol{u}}_{\mathbf{0}}$ *to* $\hat{\boldsymbol{u}}_{\mathbf{1}}$

$\equiv$ *decision on the presence of an "edge"*

Since [Geman²1984] various nonconvex Φ to produce minimizers with smooth regions and sharp edges

B. Either shrinkage or enhancement of differences

Theorem 7.1 [Φ smooth]

φ *nonconvex and* $\varphi'(0)=0$, GG^* *invertible*, $\beta > K(A,G,\varphi)$

$\Rightarrow\ \exists\ \boldsymbol{\theta_0} \in (\boldsymbol{\tau}, \boldsymbol{\mathcal{T}})$ *and* $\boldsymbol{\theta_1} > \boldsymbol{\mathcal{T}}$: *if* $\boldsymbol{\mathcal{F}_v}$ *has a (local) minimum at* $\hat{\boldsymbol{u}}$ *then*

$$\textbf{either}\quad \|\boldsymbol{G_i}\hat{\boldsymbol{u}}\| \le \boldsymbol{\theta_0} \quad \textbf{or} \quad \|\boldsymbol{G_i}\hat{\boldsymbol{u}}\| \ge \boldsymbol{\theta_1}, \quad \forall \boldsymbol{i}$$

$\widehat{\boldsymbol{h}}_0 = \{\ \boldsymbol{i}\ :\|\boldsymbol{G_i}\hat{\boldsymbol{u}}\| \le \boldsymbol{\theta_0}\}$ — **homogeneous regions**

$\widehat{\boldsymbol{h}}_1 = \{\ \boldsymbol{i}\ :\ \|\boldsymbol{G_i}\hat{\boldsymbol{u}}\| \ge \boldsymbol{\theta_1}\}$ — **edges**

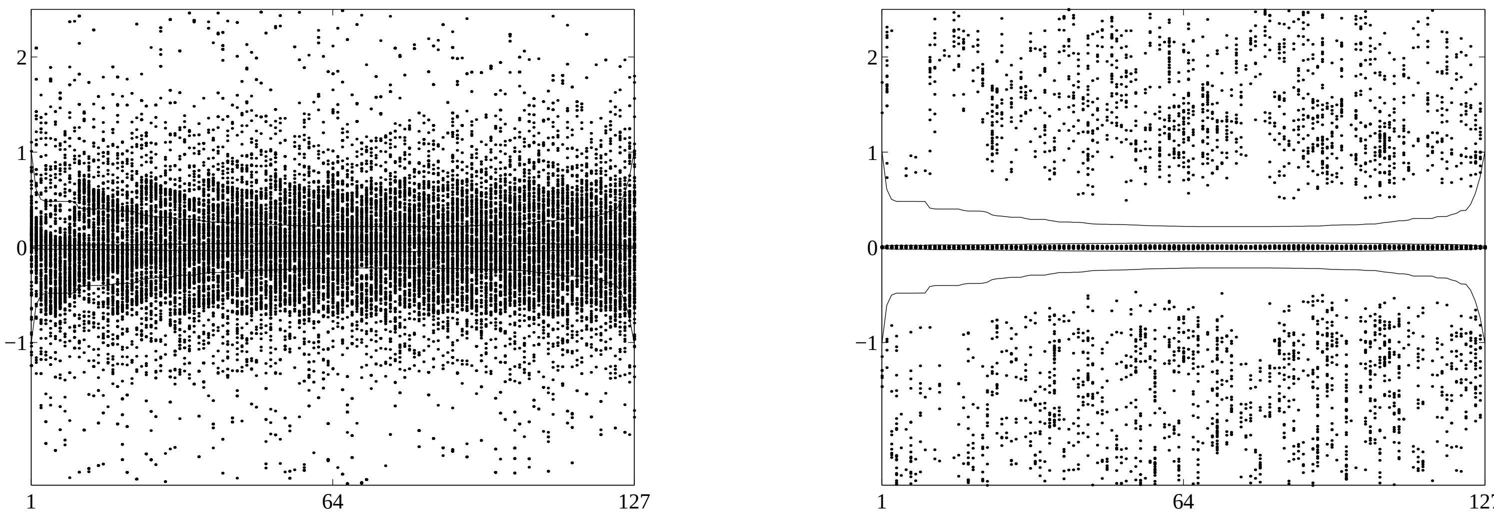
Truncated Quadratic PF $\boldsymbol{\varphi(t) = \min\{\alpha t^2,\ 1\}}$ (the discrete Mumford-Shah)

Proposition 7.2 *[Nikolova 00]*

$\boldsymbol{\mathcal{F}_v}$ *has a global minimum at* $\hat{\boldsymbol{u}}$ $\Rightarrow$ $\forall \boldsymbol{i}$

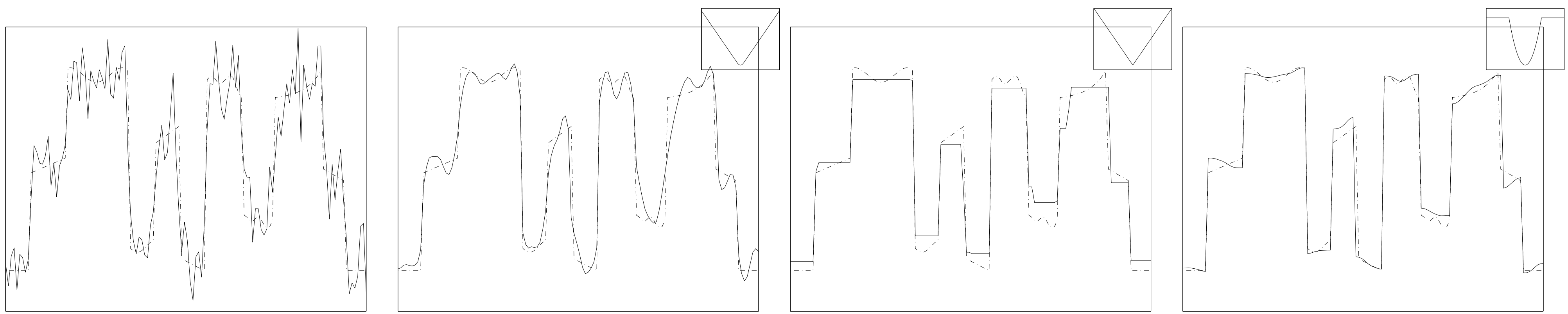
$$\textit{either}\quad \left|\hat{\boldsymbol{u}}_{\boldsymbol{i+1}} - \hat{\boldsymbol{u}}_{\boldsymbol{i}}\right| \le \frac{\boldsymbol{\Gamma}_i}{\sqrt{\boldsymbol{\alpha}}} \quad \textit{or} \quad \left|\hat{\boldsymbol{u}}_{\boldsymbol{i+1}} - \hat{\boldsymbol{u}}_{\boldsymbol{i}}\right| \ge \frac{\mathbf{1}}{\sqrt{\boldsymbol{\alpha}}\boldsymbol{\Gamma}_i}$$

$\boldsymbol{\Gamma_i < 1}$ *(explicit form)*



$u_{oi} - u_{oi+1}$ vs i for 100 original signals $\hat{u}_i - \hat{u}_{i+1}$ vs i for the global minimizers

X-axis: positions of differences $i = 1, \ldots, 127$. Y-axis: a dot at position i is the value of the ith difference of a signal. Thresholds $\pm\Gamma_i/\sqrt{\alpha}$, $\pm 1/\sqrt{\alpha}\Gamma_i$ for $i = 1, \ldots, 127$ (—).



Noisy data $v = u_o * a + n$, $a_k = \exp^{-\frac{0.4k^2}{1.4}}$, $|k| \leq 5$ and n white Gaussian noise, 10 dB SNR

Theorem 7.3 [Φ nonsmooth] *[Nikolova 04]*

$\boldsymbol{\varphi'(0)>0}$ *nonconvex,* $\beta > K(A,G,\varphi)$ $\Rightarrow$ $\exists\ \boldsymbol{\theta_1>0}$ *: if* $\boldsymbol{\mathcal{F}_v}$ *has a (local) minimum at* $\hat{\boldsymbol{u}}$ *then*

$$\textit{either} \quad \|\boldsymbol{G_i\hat{u}}\| = \boldsymbol{0} \quad \textit{or} \quad \|\boldsymbol{G_i\hat{u}}\| \geq \boldsymbol{\theta_1}, \quad \forall \boldsymbol{i}$$

Strong result with no special assumptions

$$\widehat{\boldsymbol{h}}_{\boldsymbol{0}} = \Big\{\ \boldsymbol{i}\ :\|\boldsymbol{G_i\hat{u}}\| = \boldsymbol{0}\Big\} \qquad \widehat{\boldsymbol{h}}_{\boldsymbol{1}} = \Big\{\ \boldsymbol{i}\ :\ \|\boldsymbol{G_i\hat{u}}\| \geq \boldsymbol{\theta_1}\Big\}$$

strongly homogeneous regions | **neat edges**

$\Rightarrow$ **Enhanced stair-casing, high sparsity**

"0 − 1" *PF* $\boldsymbol{\varphi(0) = 0}$ *and* $\boldsymbol{\varphi(t) = 1}$ *if* $\boldsymbol{t \neq 0}$ (Potts)

Proposition 7.4

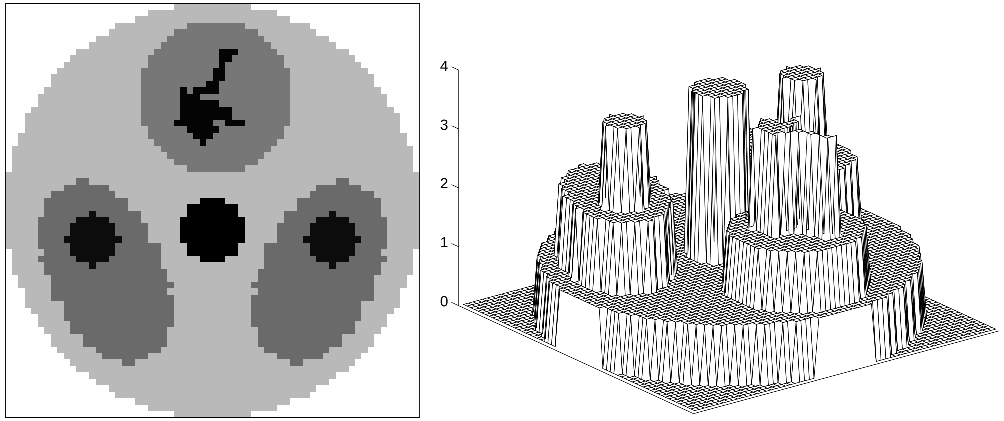
$\boldsymbol{\mathcal{F}_v}$ *has a global minimum at* $\hat{\boldsymbol{u}}$ $\Rightarrow$ $\forall \boldsymbol{i}$

$$\textit{either} \quad \hat{\boldsymbol{u}}_{\boldsymbol{i+1}} = \hat{\boldsymbol{u}}_{\boldsymbol{i}} \quad \textit{or} \quad |\hat{\boldsymbol{u}}_{\boldsymbol{i+1}} - \hat{\boldsymbol{u}}_{\boldsymbol{i}}| \geq \frac{\sqrt{\boldsymbol{\beta}}}{\boldsymbol{\Gamma_i}}$$

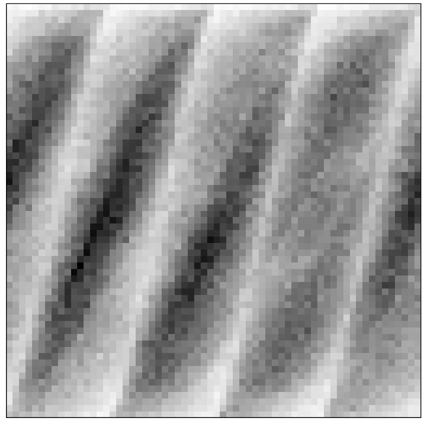
Explicit formula for Γ_i. *Strict inequality if unique global minimizer.*

- Necessary condition for $\hat{u}$ to be global minimizer

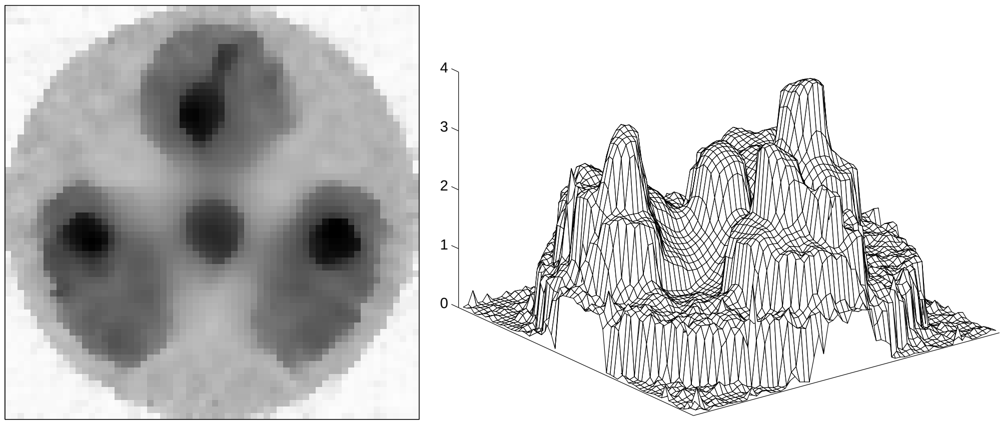
Image Reconstruction in Emission Tomography



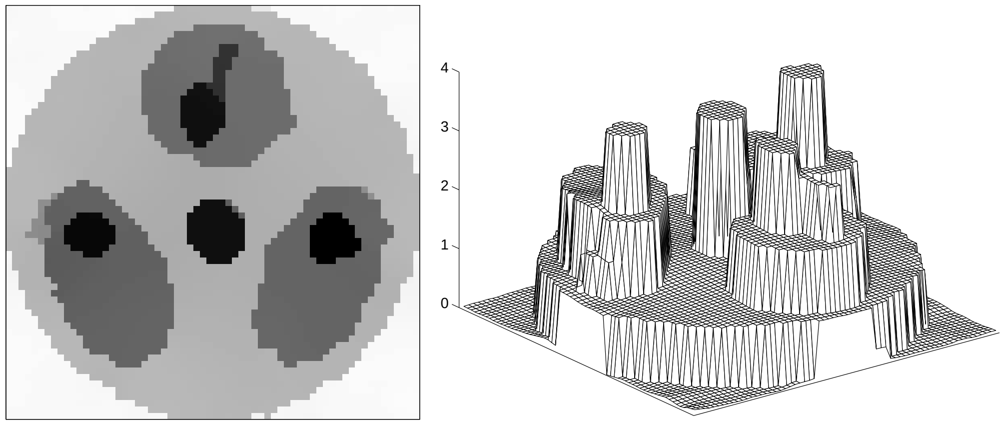
Original phantom



Emission tomography simulated data



φ is smooth (Huber function)



$\varphi(t) = t/(\alpha + t)$ (non-smooth, non-convex)

Reconstructions by minimizing $\mathcal{F}_v(u) = \Psi(u, v) + \beta \sum_{i \sim j} \varphi(|u_i - u_j|)$, $\quad \Psi =$ smooth, convex

7.2 Selection for the global minimizer

<u>*Additional assumptions:*</u> $\|\varphi\|_\infty = 1$, $G_i : \mathbb{R}^p \to \mathbb{R}$—first-order differences, A^*A invertible

$$\mathbb{1}_{\Sigma i} = \begin{cases} 1 & \text{if } i \in \Sigma \\ 0 & \text{else} \end{cases}$$ (the characteristic function of Σ)

$h_1 = \{i : \|G_i \mathbb{1}_\Sigma\| \neq 0\}$ (edges)

$h_0 = \{i : \|G_i \mathbb{1}_\Sigma\| = 0\}$ (constant zones)

Original: $u_o = \xi \mathbb{1}_\Sigma, \ \xi > 0$

Data: $v = \xi \ A \ \mathbb{1}_\Sigma \ = Au_o$

$$\hat{u} = \text{global minimizer of } \mathcal{F}_v$$

Theorem 7.5 [Φ as in Thm 7.1.] $\exists \ 0 < \xi_0 < \xi_1 \ \Rightarrow \begin{cases} \xi < \xi_0 & \hat{u} \text{ is smooth} \\ \xi > \xi_1 & \hat{u} \text{ has correct edges} \end{cases}$

Proposition 7.6 [$\varphi(t) = \min\{\alpha t^2, \ 1\}$]

$$\hat{z} = (A^*A + \beta\alpha G^*G)^{-1} A^*A\mathbb{1}_\Sigma$$ (the regularized least-squares for $\xi = 1$)

$$\exists \ \ 0 < \xi_0 < \xi_1 : \begin{cases} \xi \in (0, \xi_0) & \Rightarrow & \hat{u} = \xi \ \hat{z} & \text{(no edges)} \\ \xi > \xi_1 & \Rightarrow & \hat{u} = \xi \ \mathbb{1}_\Sigma & \text{(the original)} \end{cases}$$

Theorem 7.7 [Φ **non-smooth**]

The context of Theorem 7.3. Then $\exists\ \boldsymbol{\xi_0 > 0}, \quad \exists\ \boldsymbol{\xi_1 > \xi_0}$

- $\boldsymbol{\xi \in (0, \xi_0) \quad \Rightarrow \quad \hat{u} = c\xi 1\!\!1} \qquad c = \langle A1\!\!1, A1\!\!1_\Sigma \rangle \|A1\!\!1\|^{-2}$ *(constant solution)*

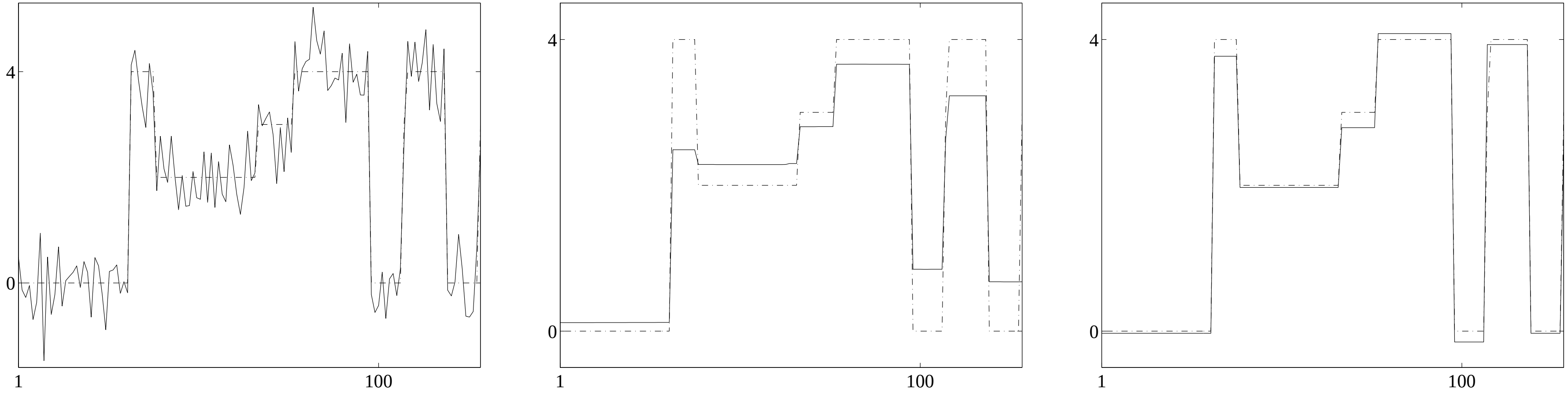
- $\boldsymbol{\xi > \xi_1} \quad \Rightarrow \quad \begin{cases} \boldsymbol{\|G_i \hat{u}\| = 0} & \boldsymbol{\forall i \in h_0} \\ \boldsymbol{\|G_i \hat{u}\| \geq \theta_1} & \boldsymbol{\forall i \in h_1} \end{cases}$ *(perfect segmentation)*

  If $\boldsymbol{\Sigma}$ *connected, then* $\boldsymbol{\hat{u} \to \xi 1\!\!1_\Sigma}$ *as* $\boldsymbol{\xi \to \infty}$

Proposition 7.8 [*"0-1" PF* $\boldsymbol{\varphi(0) = 0}$ *and* $\boldsymbol{\varphi(t) = 1}$ *if* $\boldsymbol{t \neq 0}$]

$$\exists\ \boldsymbol{\xi_0 > 0}, \quad \exists\ \boldsymbol{\xi_1 > h_0} \begin{cases} \boldsymbol{\xi \in (0, \xi_0)} & \Rightarrow \quad \boldsymbol{\hat{u} = c\ \xi\ 1\!\!1} & \text{(constant solution)} \\ \boldsymbol{\xi > \xi_1} & \Rightarrow \quad \boldsymbol{\hat{u} = \xi\ 1\!\!1_\Sigma} & \text{(the original)} \end{cases}$$

7.3 Comparison with Convex Edge-Preserving Regularization

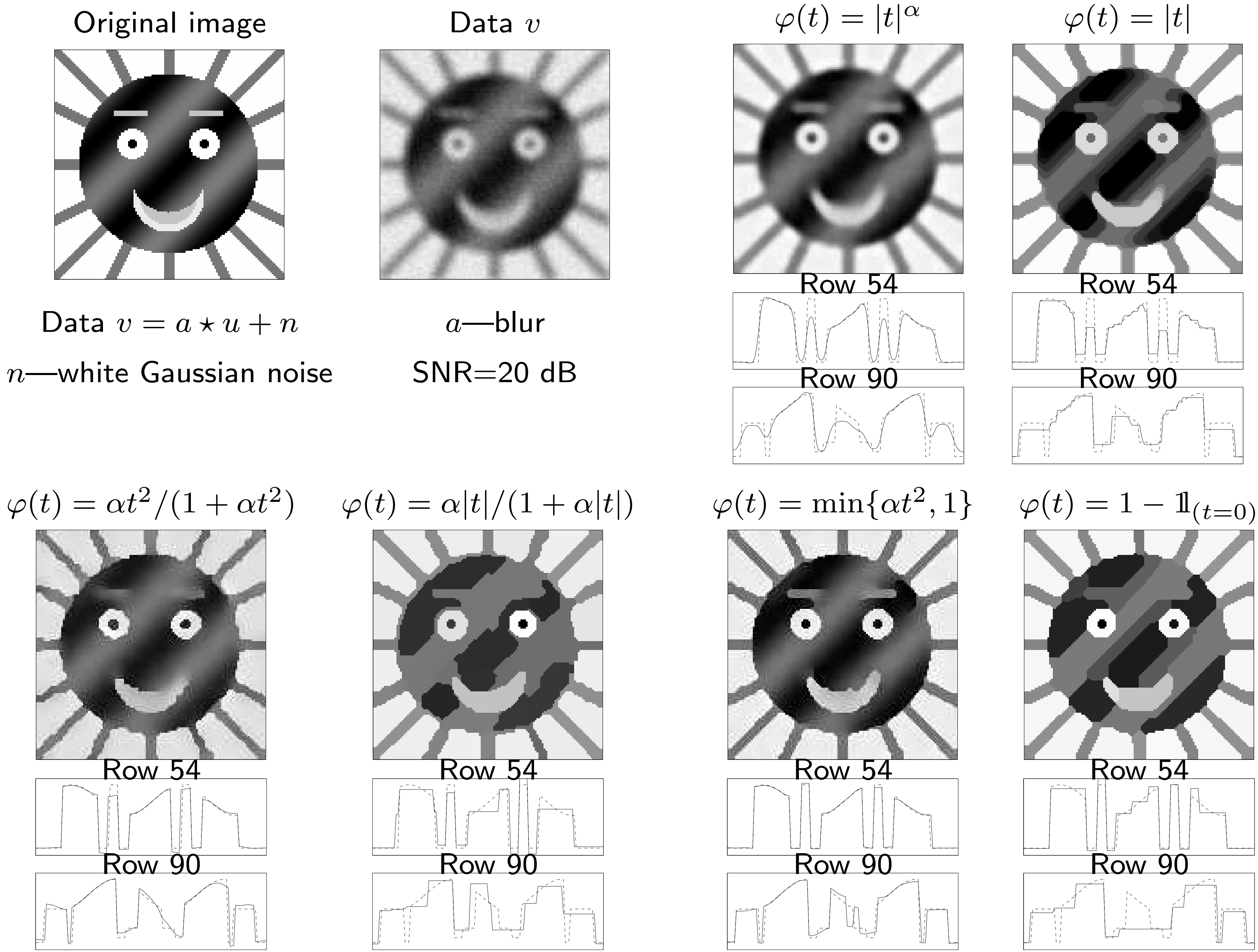


Data $v = u_o + n$ $\qquad$ $\varphi(t) = |t|$ $\qquad$ $\varphi(t) = \alpha|t|/(1 + \alpha|t|)$

- If $\mathcal{F}_v$ is convex, then $\|G_i \hat{u}\|$ can take any value on R
- TV (convex, edge-preserving) creates constant zones, a fortiori these are separated by edges whose amplitude is underestimated
- Edge-detection using φ non-convex is fundamentally different: it relies on the concurrence between different local minima corresponding to different edge configurations

  The discontinuity of $v \to \mathcal{U}(v)$ at some points plays a key role for the detection of edges

Illustration of all properties


Original image
Data v
$\varphi(t) = |t|^\alpha$
$\varphi(t) = |t|$
Data $v = a \star u + n$
a—blur
n—white Gaussian noise
SNR=20 dB
Row 54
Row 90
$\varphi(t) = \alpha t^2/(1+\alpha t^2)$
$\varphi(t) = \alpha|t|/(1+\alpha|t|)$
$\varphi(t) = \min\{\alpha t^2, 1\}$
$\varphi(t) = 1 - 1\!\!1_{(t=0)}$

8 Model distortions under Bayesian MAP

[Nikolova 07]

8.1. MAP estimators to combine noisy data and prior

Likelihood $f_{V|U}(v|u)$ $\quad + \quad$ Prior $f_U(u) \propto \exp\{-\lambda\Phi(u)\}$

<u>*MAP $\hat{\boldsymbol{u}}$ = the most likely solution given the recorded data $\boldsymbol{V} = \boldsymbol{v}$:*</u>

$$\hat{u} = \arg\max_u f_{U|V}(u|v) \;=\; \arg\min_u \big(-\ln f_{V|U}(v|u) - \ln f_U(u) \big)$$
$$= \arg\min_u \big(\Psi(u, v) \; + \; \beta\Phi(u) \big)$$

Realist models for data-acquisition $\boldsymbol{f_{V|U}}$ and prior $\boldsymbol{f_U}$

$\Rightarrow$ $\hat{\boldsymbol{u}}$ must be coherent with $\boldsymbol{f_{V|U}}$ and $\boldsymbol{f_U}$

$$\boxed{\left\{ \begin{array}{r} \boldsymbol{U} \sim \boldsymbol{f_U} \\ \boldsymbol{AU} - \boldsymbol{V} \sim \boldsymbol{f_N} \end{array} \right. \Rightarrow \text{we need} \left\{ \begin{array}{l} \boldsymbol{f_{\hat{U}}} \approx \boldsymbol{f_U} \\ \boldsymbol{f_{\hat{N}}} \approx \boldsymbol{f_N}, \quad \boldsymbol{\hat{N}} = \boldsymbol{A\hat{U}} - \boldsymbol{V} \end{array} \right.}$$

In general $f_{\hat{U}}$ and $f_{\hat{N}}$ cannot be calculated

Example: MAP shrinkage estimators

[Simoncelli99, Belge-Kilmer00, Moulin-Liu00, Antoniadis02]

- Noisy wavelet coefficients

$$y = Wv = Wu_o + n$$

where $n \sim \mathcal{N}(0, \sigma^2)$

Original coefficients $x_o = Wu_o$

- Prior: x_i are i.i.d., Generalized Gaussian (GG)

$$\begin{aligned} x_i \sim f_X(t) &= \frac{1}{Z} e^{-\lambda |t|^\alpha} \\ f_X(x) &= \prod_i f_X(x_i) \end{aligned}$$

Experiments have shown that usually $\alpha \in (0,1)$ for real-world images

- MAP restoration $\hat{x} = \arg\max_x f(y|x) f(x) = \arg\min_x \sum_i \Big((x_i - y_i)^2 + \lambda_i |x_i|^\alpha \Big)$

$$\Leftrightarrow \qquad \boldsymbol{\hat{x}_i = \arg\min_{t \in \mathbf{R}} \left((t - y_i)^2 + \lambda_i |t|^\alpha \right), \quad \forall i}$$

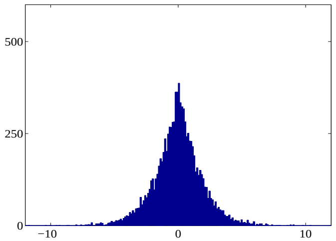
- Restored image or signal $\boldsymbol{\hat{u} = W^* \hat{x}}$

(α, λ) and σ fixed—

10 000 independent trials:

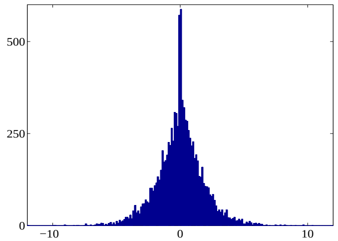
- sample $x \in \mathrm{R}$ from f_X
- $y = x + n, \quad n \sim \mathcal{N}(0, \sigma^2)$
- compute the true MAP $\hat{x}$

$f_{X|Y}(., y)$—one mode if $\alpha \geq 1$

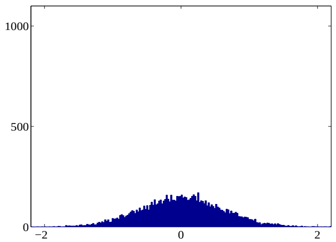


GG prior

$\alpha = 1.2, \lambda = 0.5$

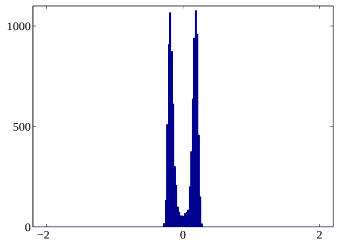


The true MAP $\hat{x}$



Noise $\mathcal{N}(0, \sigma^2)$

$\sigma = 0.6$

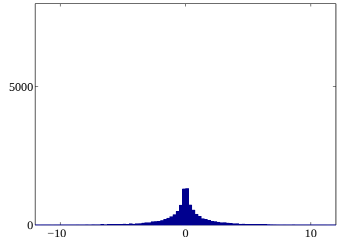


Noise estimate

$\hat{n} = y - \hat{x}$

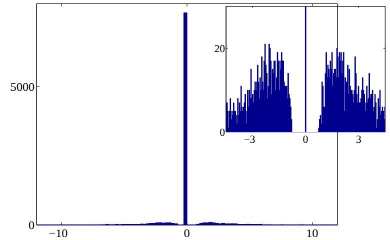
$0 < \alpha < 1$—modes $\hat{u}_1 = 0$ and $|\hat{u}_2| > 0.47$

$\Rightarrow$ $\mathrm{P}_{\hat{X}}(0) = \text{Dirac}$ and $\mathrm{P}_{\hat{X}} = 0$ on $(-\theta, 0) \cup (0, \theta)$



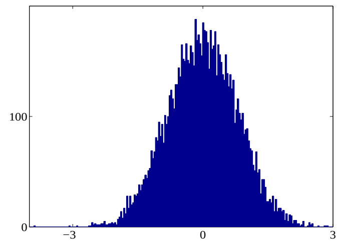
Prior f_X for

$\alpha = 0.5, \lambda = 2$



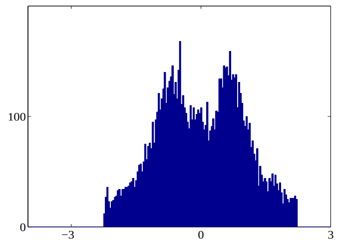
True MAP $\hat{x}$

$+$ zoom



Noise $\mathcal{N}(0, \sigma^2)$

for $\sigma = 0.8$



Noise estimate

$\hat{n} = y - \hat{x}$

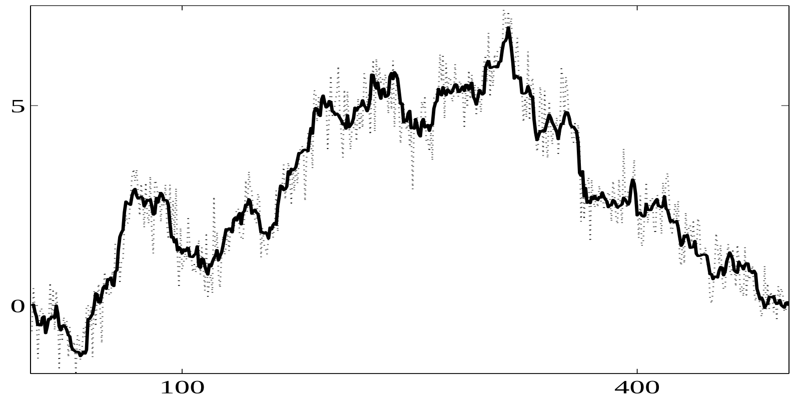
Hist$(\hat{x}_i) \neq f_X$ and Hist$(y_i - \hat{x}_i) \neq f_N$ $\Rightarrow$ $\boxed{\text{Gap between model and estimate}}$

8.2. Non-smooth at zero priors

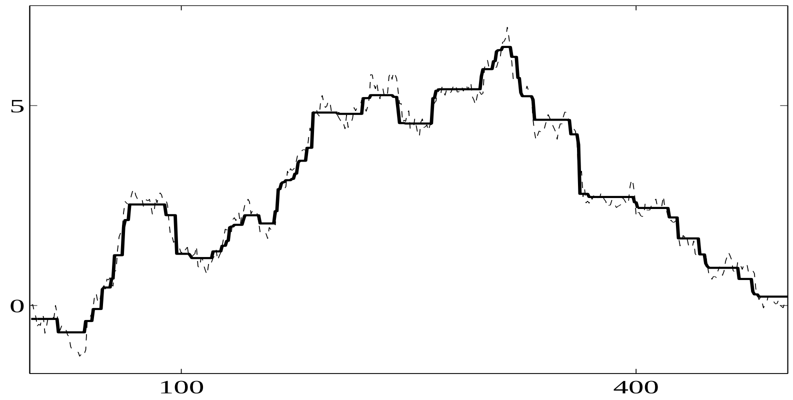
A. Experiment: Laplacian Markov chain corrupted with Gaussian noise

- Markov chain : $f_U \propto \exp\Big(-\lambda\Phi(u)\Big)$ with $\Phi(u) = \lambda\sum_{i=1}^{p-1}|u_i - u_{i+1}|, \quad \lambda > 0$

  $\boldsymbol{U_i - U_{i+1}}$ —i.i.d. Laplacian—$\boldsymbol{f_{\Delta U}(t) = \frac{\lambda}{2}\exp\Big(-\lambda|t|\Big)}$
- Data $V = U + N, \quad N \sim \mathcal{N}(0, \sigma^2 I)$
- MAP energy $\boldsymbol{\mathcal{F}_v(u) \; = \; \|u - v\|^2 + \beta\sum_i \Big|u_i - u_{i+1}\Big|, \quad \beta = 2\sigma^2\lambda}$



Original u (—), $u_i - u_{i+1}$ sampled from $f_{\Delta U}$ for $\lambda = 8$ and data $v = u + n$ $(\cdots)$ for $\sigma = 0.5$.



The true MAP $\hat{u}$ (—) versus the original u (- - -). $\hat{u}$ involves 92% null differences

The same experiment 40 times: no zero-valued difference $u_{oi} - u_{oi+1}$ was sampled whereas 87% of all restored differences $\hat{u}_i - \hat{u}_{i+1}$ are null $\Rightarrow$ `the MAP solution does not fit the prior`

B. Explanation

Statistical consequence of Theorem 3.1:

$$\boldsymbol{v} \in \mathcal{O}_{\hat{\boldsymbol{h}}} \quad \text{and} \quad \hat{\boldsymbol{u}} = \arg\max_{\boldsymbol{u}\in\mathbf{R}^p} f_{U|V}(\boldsymbol{u}|\boldsymbol{v}) \quad \Rightarrow \quad \left[\boldsymbol{G}_i\hat{\boldsymbol{u}} = \boldsymbol{0} \;\; \forall \boldsymbol{i} \in \hat{\boldsymbol{h}} \quad \Leftrightarrow \quad \hat{\boldsymbol{u}} \in \boldsymbol{K}_{\hat{\boldsymbol{h}}}\right]$$

$$\Rightarrow \quad \mathbf{Pr}(\hat{\boldsymbol{U}} \in \boldsymbol{K}_{\hat{\boldsymbol{h}}}) \geq \mathbf{Pr}(\boldsymbol{V} \in \mathcal{O}_{\hat{\boldsymbol{h}}}) = \int_{\mathcal{O}_{\hat{\boldsymbol{h}}}} f_{\boldsymbol{V}}(\boldsymbol{v})d\boldsymbol{v} > \boldsymbol{0}$$

since $f_V(\boldsymbol{v}) = \frac{1}{Z}\int \exp\big(-\mathcal{F}_v(\boldsymbol{u})\big)d\boldsymbol{u} > \boldsymbol{0}$ and $\mathcal{L}^q(\mathcal{O}_{\hat{\boldsymbol{h}}}) > \boldsymbol{0}$

The "prior" model on the unknown U effectively *realized by the MAP estimator $\hat{U}$ corresponds to images and signals such that $G_i\hat{U} = 0$ for a certain number of indexes i.*

If $\{G_i\}$=first-order, then effective prior model for locally constant images and signals.

According to the prior distribution, for any nonempty $h \subset \{1, \ldots, r\}$

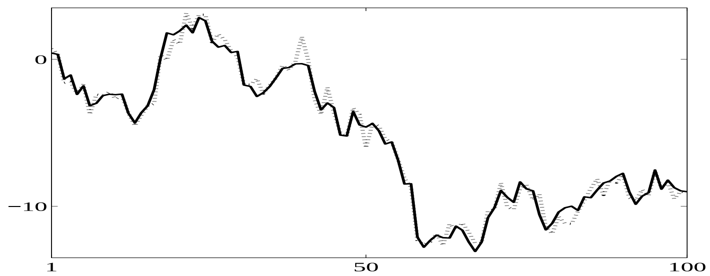
$$\mathbf{Pr}(\boldsymbol{U} \in \boldsymbol{K}_{\boldsymbol{h}}) = \int_{\boldsymbol{K}_{\boldsymbol{h}}} f_{\boldsymbol{U}}(\boldsymbol{u})d\boldsymbol{u} = \boldsymbol{0}$$

since f_U is continuous and $\mathbf{dim}\, \boldsymbol{K}_{\boldsymbol{h}} < \boldsymbol{p}$

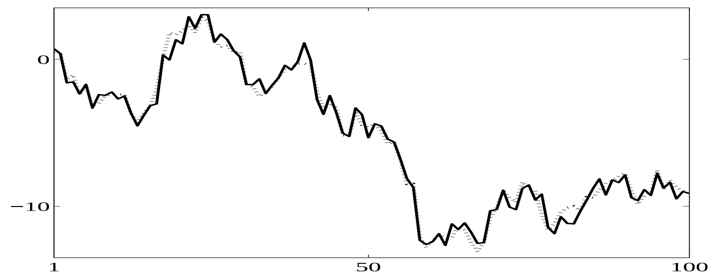
8.3 Non-smooth at zero noise models

A. Experiment: Generalized Gaussian Markov chain under Laplace noise

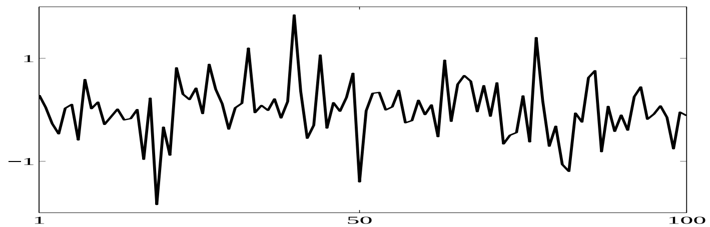
- U — Markov chain, $U_i - U_{i-1} \sim f_{\Delta U}$ are i.i.d., $f_{\Delta U}(t) = \frac{1}{Z} e^{-\lambda |t|^\alpha}$
- Data $V = U + N$ where N_i are i.i.d. with $f_N(t) = \frac{\sigma}{2} e^{-\sigma |t|}$
- MAP energy: $\boldsymbol{\mathcal{F}_v(u) = \sum_{i=1}^p \left| u_i - v_i \right| + \beta \sum_i |u_i - u_{i+1}|^\alpha}$ where $\boldsymbol{\beta = \frac{\lambda}{\sigma}}$



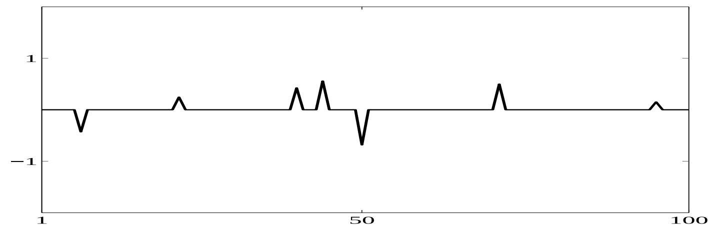
u_o (—) for $\alpha = 1.2, \lambda = 1$, data v $(\cdots)$



True MAP $\hat{u}$ (—) vs u_o $(\cdots)$



Laplacian noise n for $\sigma = 2.5$



The noise estimate $\hat{n} = y - \hat{u}$.

$\boldsymbol{u_{oi} \neq v_i}$ for all i whereas the MAP $\hat{u}$ contains 93% samples satisfying $\boldsymbol{\hat{u}_i = v_i}$ ($f_N(0) =$Dirac)

The same experiment 1000 times $\Rightarrow$ `the true MAP cannot efficiently clean Laplacian noise`

B. Explanation

$\boldsymbol{V} = \boldsymbol{A}\boldsymbol{U} + \boldsymbol{N}$, $N_i \sim f_N$ are i.i.d. $\boldsymbol{f_N(t) = \frac{1}{Z} exp\big(-\sigma\psi(t)\big)}$ continuous, $\boldsymbol{\psi'(0^+) > 0}$

$$\boldsymbol{f_{V|U}(v|u)} = \prod_i f_N(\langle a_i, u\rangle - v_i) \quad \propto \quad \mathbf{exp}\,\big(-\boldsymbol{\sigma\Psi(u,v)}\big)$$

$$\boldsymbol{\Psi(u,v) \quad = \quad \textstyle\sum_{i=1}^{q} \psi(\langle a_i, u\rangle - v_i)}$$

$$\Rightarrow \quad \mathbf{Pr}\Big(\langle \boldsymbol{a_i}, \boldsymbol{U}\rangle - \boldsymbol{V_i} = \mathbf{0}\Big) = \mathbf{Pr}\,\big(\boldsymbol{N_i} = \mathbf{0}\big) = \mathbf{0}, \quad \forall \boldsymbol{i}$$

For $X \sim$ Gibbsian where $\Phi \sim \mathcal{C}^m$, the true MAP $\hat{\boldsymbol{u}}$ minimizes

$$\boldsymbol{\mathcal{F}_v(u) = \Psi(u,v) + \beta\Phi(u), \quad \beta = \frac{\lambda}{\sigma}}$$

By Theorem 4.1

$$\boldsymbol{v \in \mathcal{O}_{\hat{h}}} \quad \text{and} \quad \hat{\boldsymbol{u}} \quad = \quad \mathbf{arg\,max}_{\boldsymbol{u}\in\mathbf{R}^p}\, \boldsymbol{f_{U|V}(u|v)} \quad \Rightarrow \quad \langle \boldsymbol{a_i}, \hat{\boldsymbol{u}}\rangle = \boldsymbol{v_i} \quad \forall \boldsymbol{i} \in \hat{\boldsymbol{h}}$$

$$= \mathbf{arg\,min}_{\boldsymbol{u}}\, \boldsymbol{\mathcal{F}_v(u)}$$

$$\Rightarrow \quad \mathbf{Pr}\Big(\langle \boldsymbol{a_i}, \hat{\boldsymbol{U}}\rangle - \boldsymbol{V_i} = \mathbf{0}, \forall \boldsymbol{i} \in \hat{\boldsymbol{h}}\Big) \geq \mathbf{Pr}(\boldsymbol{V} \in \boldsymbol{\mathcal{O}_{\hat{h}}}) = \int_{\boldsymbol{\mathcal{O}_{\hat{h}}}} \boldsymbol{f_V(v)dv} > \mathbf{0}$$

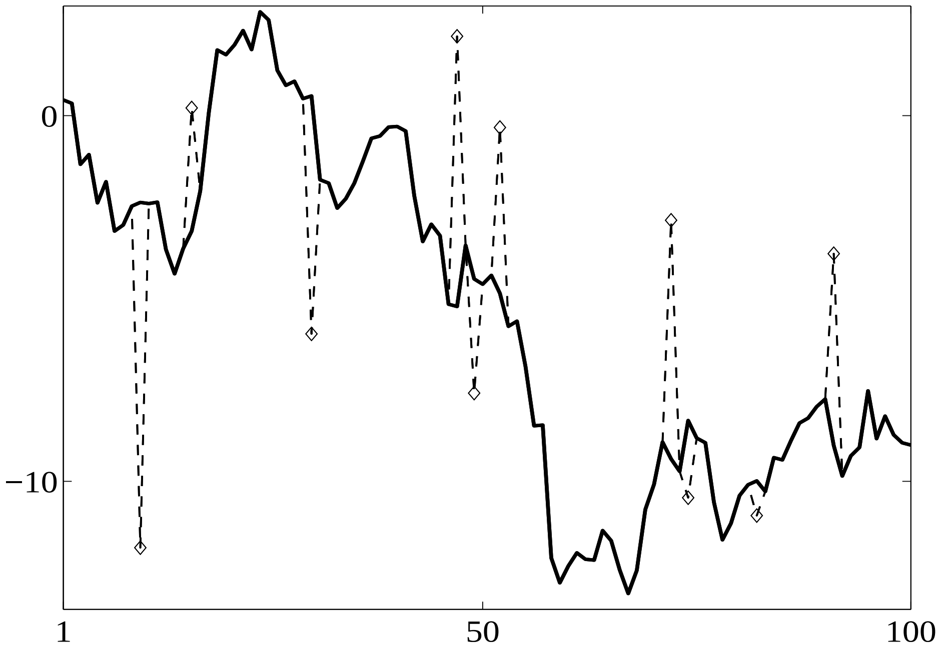
since $\boldsymbol{f_V(v)} > \mathbf{0}$ and $\boldsymbol{\mathcal{L}^q(\mathcal{O}_{\hat{h}})} > \mathbf{0}$

For all $i \in \hat{h}$, the prior has no influence on the solution and the noise remains intact

C. Yet another experiment : Laplace noise model to remove impulse noise

- $\boldsymbol{A = I}$
- Data samples $\boldsymbol{v_i, \ i \in \hat{h}}$ are fitted exactly, hence they must be free of noise.

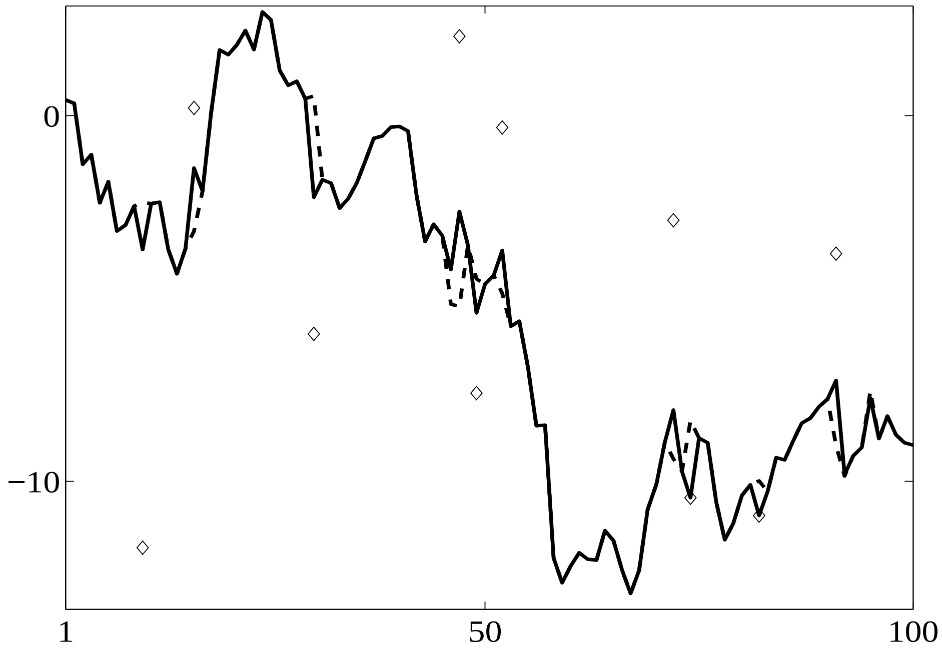
  If $\boldsymbol{i \in \hat{h}^c}$ then $\boldsymbol{v_i}$ is replaced by $\boldsymbol{\hat{u}_i = \mathcal{U}_i(\{v_j : j \in \hat{h}\})}$ which is *independent* of $\boldsymbol{v_i}$
- *The MAP estimator defined by $\mathcal{F}_v$ corresponds to an impulse noise model on the data*



Original u_o (—), data v (- - -)

with 10% random valued impulse noise.

(`More examples in` §4)



The minimizer $\hat{u}$ of $\mathcal{F}_v$ for $\beta = 0.4$ (—),

the original u_o (- - -), and $v_i \neq \hat{u}_i$ ($\diamond$)

$\hat{u}_i = v_i$ for 99% of the noise-free samples.

8.4 Priors with non-convex energies

$\boldsymbol{V = AU + N}$ with $\boldsymbol{N \sim \mathcal{N}(0, \sigma^2)}$ and $\boldsymbol{U \sim f_U(u) \propto e^{-\lambda \Phi(u)}}$ with $\boldsymbol{\Phi(u) = \sum_{i=1}^r \varphi(\|G_i u\|)}$, where φ and G_i as in section 7.1. The MAP $\hat{u}$ yields the (global) minimum of

$$\mathcal{F}_v(u) = \|Au - v\|^2 + \beta \sum_{i=1}^r \varphi(\|G_i u\|), \quad \beta = 2\sigma^2\lambda$$

A. Piecewise Gaussian Markov chain in Gaussian noise

[Nikolova 2000]

Piecewise GM chain = discrete 1D Mumford-Shah = weak-string model [Blake-Zisserman87]

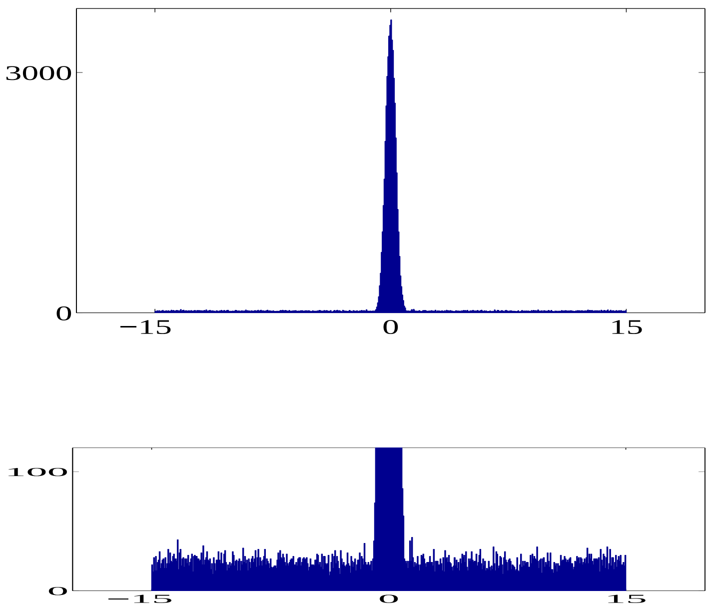
U such that $U_{i+1} - U_i$ i.i.d. $\sim f_{\Delta U}(t) \propto e^{-\lambda\varphi(t)}$

$$\varphi(t) = \begin{cases} \alpha t^2 & \text{if } |t| < \sqrt{\frac{1}{\alpha}} \\ 1 & \text{else} \end{cases} = \min\{\alpha t^2, 1\}, \qquad \Phi(u) = \sum_i \varphi(u_i - u_{i+1})$$

`For an illustration—see §7.1-B`

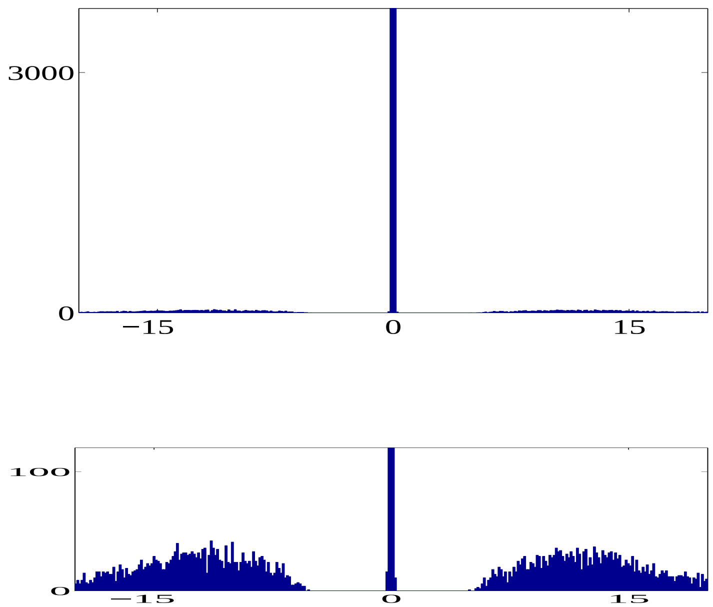
Repeat 200 times the following experiment:

- generate $U = u \in \mathrm{R}^{300}$ where $u_i - u_{i+1}$—sampled from $f_{\Delta U}$ for $\alpha = 1$, $\lambda = 5$ and $\gamma = 15$
- $v = u + n$ where $n \sim \mathcal{N}(0, \sigma^2 I)$, $\sigma = 4$
- compute $\hat{u} = \arg\min \mathcal{F}_v$ for the true parameter $\beta = 2\sigma^2\lambda = 160$.



Histogram of all original differences

$u_i - u_{i+1}$ (up) and zoom (bottom)



Histogram of the differences for all the

true MAP solutions $\hat{u}$ (up) and zoom (bottom)

B. Explanation

Theorem 7.1 and Prop. 7.2 imply $\quad \mathbf{Pr}\left(\boldsymbol{\theta_0} < \|\boldsymbol{G_i}\hat{\boldsymbol{U}}\| < \boldsymbol{\theta_1}\right) = \mathbf{0}, \quad \forall \boldsymbol{i}$

The prior model **effectively** *realized by the MAP estimator corresponds to images and signals whose differences are either smaller than θ_0 or larger than θ_1.*

In contrast, for the prior distribution $\quad \mathbf{Pr}\left(\boldsymbol{\theta_0} < \|\boldsymbol{G_i U}\| < \boldsymbol{\theta_1}\right) > \mathbf{0}$, $\forall \boldsymbol{i}$

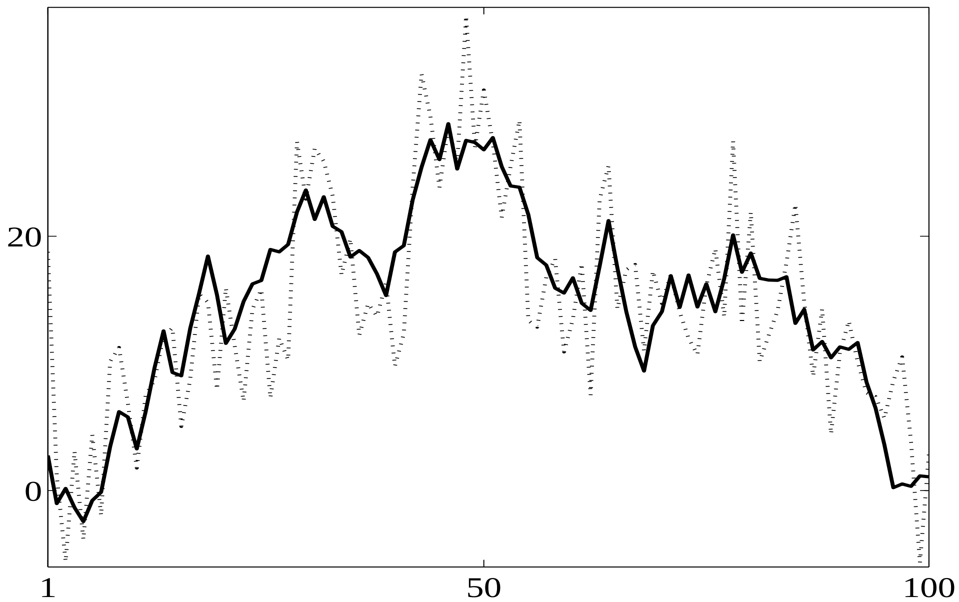
C. MAP for Φ non-smooth $(\varphi'(0) > 0)$

$$\text{Theorem 7.3} \quad \Rightarrow \quad \mathbf{Pr}\left(\|\boldsymbol{G_i}\hat{\boldsymbol{U}}\| = \mathbf{0}\right) > \mathbf{0} \qquad \mathbf{Pr}\left(\mathbf{0} < \|\boldsymbol{G_i}\hat{\boldsymbol{U}}\| < \boldsymbol{\theta_1}\right) = \mathbf{0}$$

If $\{G_i\}$—first-order differences between neighbors, every minimizer $\hat{u}$ of $\mathcal{F}_v$ is composed out of constant patches separated by edges higher than $\theta_1 \equiv$ the **effective** *model realized by the MAP*

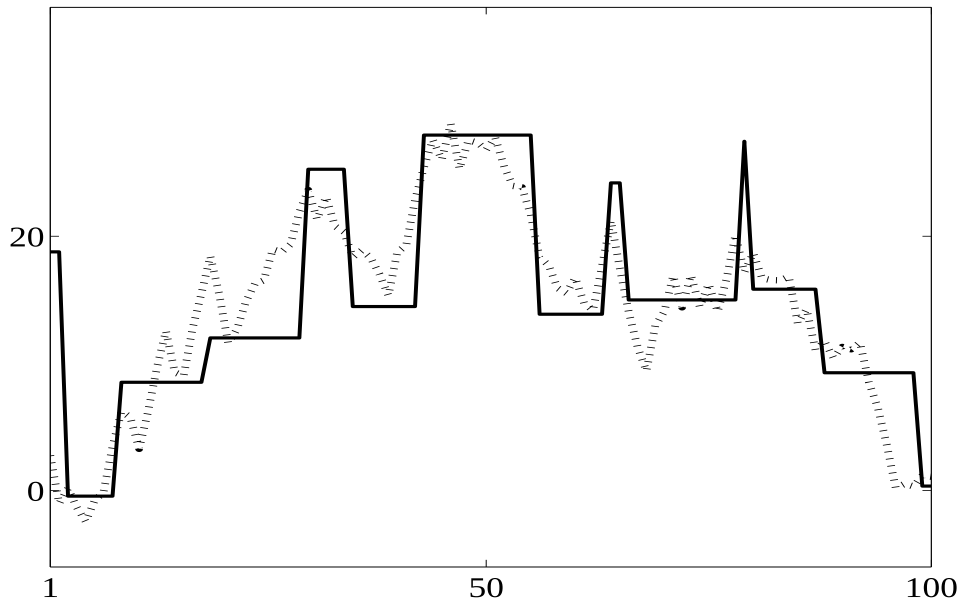
For the prior distribution $\mathbf{Pr}\left(\|\boldsymbol{G_i U}\| = \mathbf{0}\right) = \mathbf{0}$ and $\mathbf{Pr}\left(\mathbf{0} < \|\boldsymbol{G_i U}\| < \boldsymbol{\theta_1}\right) > \mathbf{0}$

Illustration: Original differences $U_i - U_{i+1}$ i.i.d.$\sim f_{\Delta U}(t) \propto e^{-\lambda\varphi(t)}$ on $[-\gamma, \gamma]$, $\boldsymbol{\varphi(t)} = \frac{\boldsymbol{\alpha|t|}}{\mathbf{1}+\boldsymbol{\alpha|t|}}$



Original u_o (—) by $f_{\Delta U}$ for $\alpha = 10$, $\lambda = 1$, $\gamma = 4$

data $v = u_o + n$ $(\cdots)$, $N \sim \mathcal{N}(0, \sigma^2 I)$, $\sigma = 5$.



The true MAP $\hat{u}$ (—), $\beta = 2\sigma^2\lambda$

versus the original u_o $(\cdots)$.

8.5 Comments on Bayesian MAP

- MAP estimators do not match the underlying models for the production of the data and for the prior
- Knowing the true distributions, with the true parameters, is not sufficient
- The models *effectively realized by the MAP solutions* can be characterized using different tools
- *Conjecture: similar problems generally arise with other Bayesian estimators too*
- Combining models is an open problem

9. Computational issues

9.1 Half-quadratic minimization

[Geman &Reynolds 1992],[Geman &Yang 1995]

A. MULTIPLICATIVE FORM $(\star)$

φ convex, $\varphi(\sqrt{.})$ concave *[Geman92,Charbonnier97,Idier01...]*

Augmented energy $\boldsymbol{F}$

$$\boldsymbol{F(u,b)} = \|\boldsymbol{Au - v}\|^2 + \boldsymbol{\beta} \sum_{i=1}^{r} \left(\frac{\boldsymbol{b_i}}{\mathbf{2}} \|\boldsymbol{G_i u}\|^2 + \boldsymbol{\psi(b_i)} \right) \qquad \text{(non-convex)}$$

$$\boldsymbol{\psi} \quad \text{such that} \quad \inf_{\mathrm{b} \in \mathbb{R}} \left\{ \frac{\mathbf{b}}{\mathbf{2}} \|\boldsymbol{G_i u}\|^2 + \boldsymbol{\psi}(\mathbf{b}) \right\} = \boldsymbol{\varphi}(\|\boldsymbol{G_i u}\|)$$

Alternate minimization

$$\begin{aligned} \boldsymbol{b_i^{(k)}} &= \frac{\boldsymbol{\varphi'(\|G_i u^{(k-1)}\|)}}{\boldsymbol{\|G_i u^{(k-1)}\|}}, \ \forall \boldsymbol{i} & \qquad b = [b_1^*, \ldots, b_r^*]^* \\ \boldsymbol{H(b^{(k)})} &= \boldsymbol{2A^*A + \beta G^* \mathrm{diag}(b^{(k)}) G} & \qquad \text{(update each iteration)} \\ \boldsymbol{u^{(k)}} &= \boldsymbol{\left(H(b^{(k)})\right)^{-1} 2A^* v} & \end{aligned}$$

Amounts to quasi-Newton: $\boldsymbol{u^{(k)} = u^{(k-1)} - \left(\mathcal{H}(u^{(k-1)})\right)^{-1} \nabla \mathcal{F}_v (u^{(k-1)})}$

B. Additive form (+)

φ convex, $\left(\frac{1}{2}t^2 - \varphi(t)\right)$ convex (in practice—$\varphi''(0) = 1$) [Geman95,Cohen96,Aubert97]

$$\boldsymbol{F(u,b) = \|Au - v\|^2 + \beta \sum_{i=1}^{r} \left(\frac{1}{2}\|G_i u - b_i\|^2 + \psi(b_i)\right)}$$

$$\boldsymbol{\psi} \quad \text{such that} \quad \inf_{\mathbf{b}} \left\{\frac{1}{2}\|\boldsymbol{G_i u} - \mathbf{b}\|^2 + \boldsymbol{\psi}(\mathbf{b})\right\} = \boldsymbol{\varphi(\|G_i u\|)}$$

For $\boldsymbol{H = 2A^*A + \beta G^*G}$ (fixed, easy to precondition) calculate

$$\begin{aligned} \boldsymbol{b_i^{(k)}} &= \boldsymbol{G_i u^{(k-1)} \left(I - \frac{\varphi'(\|G_i u^{(k-1)}\|)}{\|G_i u^{(k-1)}\|}\right)}, \quad \forall \boldsymbol{i} \\ \boldsymbol{u^{(k)}} &= \boldsymbol{H^{-1}\left(2A^*v + \beta G^* b^{(k)}\right)} \end{aligned}$$

It is quasi-Newton: $\boldsymbol{u^{(k)} = u^{(k-1)} - H^{-1}\nabla\mathcal{F}_v(u^{(k-1)})}$ $\qquad \nabla^2\mathcal{F}_v(u) \approx 2A^*A + \beta\varphi''(0)G^*G$

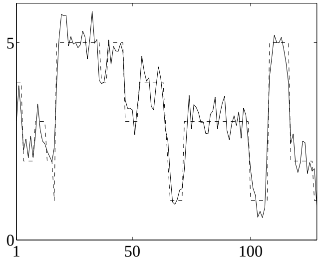
Comparison Linear convergence to $\boldsymbol{u^{(k)}} \to \boldsymbol{\hat{u}}$ for both forms [Nikolova &Ng 05, Allain 06]

Root-convergence factor $\quad \mathcal{R}(\star) < \mathcal{R}(+)$ (less iterations)

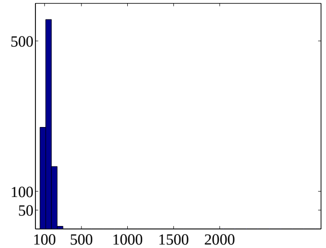
Cost-per-iteration $\quad C(\star) > C(+)$

`Both forms are faster than Steepest descent, Conjugated gradients, BFGS, DFP`

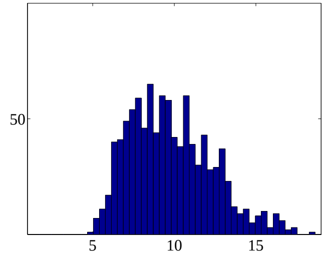
Illustration with a small-scale problem



Orig.(- -) Data(—)



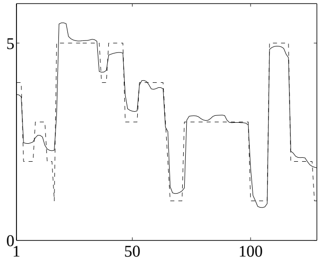
Iterations $\star$



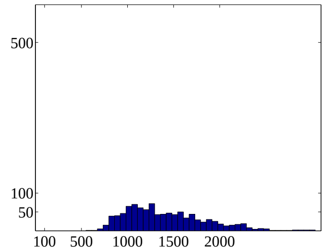
CPU time $\star$

Experiment with 1000 random signals

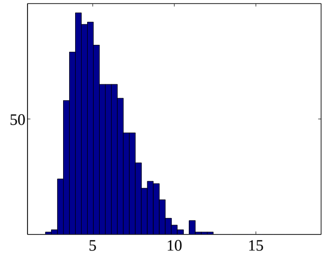
$$\varphi(t) = \sqrt{\alpha + t^2}$$



Restoration(—)



Iterations $+$



CPU time $+$

For large-scale problems, H^{-1} in "$+$" form may be difficult to store

$H(b)$ in "$\star$"-form may be ill-conditioned, it is inverted at each iteration

C. Truncated forms

[Labat-Idier 07]

- Truncated preconditioned conjugated gradient to find approximate (rough) inverses
- Convergence proven, CPU time considerably improved
- Similar approach to other quasi-Newton methods

9.2 Minimization of nonsmooth convex energies

A. Minimization scheme for ℓ_1 data-fidelity

[Nikolova 04]

Extension of *[Glowinski-Trémolières 76]* to $\boxed{\mathcal{F}_v(u) = \|u - v\|_1 + \frac{\beta}{2} \sum_{i,j \in \mathcal{N}_i} \varphi(|u_i - u_j|)}$

$$\hat{u} = \hat{z} + v \text{ where } \hat{z} \text{ minimizes } F_v$$

$$F_v(z) = \|z\|_1 + \tfrac{\beta}{2} \textstyle\sum_{i,j \in \mathcal{N}_i} \varphi(z_i + v_i - z_j - v_j)$$

Initialize with $z^{(0)} = 0$
Iteration k, $\forall i = 1, \ldots, p$

$$\xi_i^{(k)} \quad = \quad \beta \sum_{j \in \mathcal{N}_i} \varphi'(v_i - z_j - v_j), \qquad z = \left(z_1^{(k)}, \ldots, z_{i-1}^{(k)},\ 0,\ z_{i+1}^{(k-1)}, \ldots, z_p^{(k-1)}\right);$$

if $\quad |\xi_i^{(k)}| \leq 1 \quad \rightarrow \quad z_i^{(k)} = 0$

if $\quad |\xi_i^{(k)}| > 1 \quad \rightarrow \quad$ find $z_i^{(k)}$ by solving $\beta \sum_{j \in \mathcal{N}_i} \varphi'(z_i^{(k)} + y_i - z_j - y_j) = \mathrm{sign}(\xi_i^{(k)})$

knowing that $\quad \mathrm{sign}(z_i^{(k)}) = -\mathrm{sign}(\xi_i^{(k)})$

Under weak conditions $\hat{z}^{(k)} \rightarrow \hat{z}$

The calculation of each $z_i^{(k)}$ involves only the entries whose indexes are in $\mathcal{N}_i$.

B. Fast minimization for TV regularized signals

(Side-product of the method in A)

$$\boldsymbol{\mathcal{F}_v(u) = \|Au - v\|^2 + \beta \sum_{i=1}^{p-1} |u_i - u_{i+1}|}$$

$\boldsymbol{z = Tu}$: $z_i = u_i - u_{i+1}$, $1 \leq i \leq p-1$ and $z_p = \frac{1}{p}\sum_{i=1}^{p} u_i$, $\quad \boldsymbol{B = AT^{-1} = \left[\mathrm{b}_1, \ldots, \mathrm{b}_p\right]}$

Iteration k:

- $1 \leq i \leq p-1$ calculate $\xi_i^{(k)} = 2\mathrm{b}_i^* B \left(z_1^{(k)}, z_2^{(k)}, \ldots, z_{i-1}^{(k)},\ 0,\ z_{i+1}^{(k-1)}, \ldots, z_p^{(k-1)}\right) - 2\mathrm{b}_i^* v$

$$\begin{array}{llll} \text{if} & \left|\xi_i^{(k)}\right| \leq \beta, & \text{set} & z_i^{(k)} = 0 \\ \text{if} & \xi_i^{(k)} < -\beta & \text{set} & z_i^{(k)} = -\dfrac{\xi_i^{(k)} + \beta}{2\|\mathrm{b}_i\|^2} > 0 \\ \text{if} & \xi_i^{(k)} > \beta & \text{set} & z_i^{(k)} = -\dfrac{\xi_i^{(k)} - \beta}{2\|\mathrm{b}_i\|^2} < 0 \end{array}$$

- for $i = p$ $\qquad z_p^{(k)} = -\dfrac{\xi_p^{(k)}}{2\|\mathrm{b}_p\|^2}$

$\Rightarrow \quad \boldsymbol{T^{-1} z^{(k)} \rightarrow \hat{u}}$

This method cannot be extended to images

C. Fast minimization for $\ell_1 + \mathrm{TV}_1$ and $\ell_2 + \mathrm{TV}_1$

[Haoying, Ng, Nikolova &Barlow06]

$$\begin{aligned} \boldsymbol{x} &= \boldsymbol{Au} - \boldsymbol{v} &&= \boldsymbol{x}^+ - \boldsymbol{x}^- \\ \boldsymbol{y} &= \boldsymbol{\beta G u} &&= \boldsymbol{y}^+ - \boldsymbol{y}^- \end{aligned}$$

- minimize $\mathcal{F}(\boldsymbol{u}) = \|\boldsymbol{Au} - \boldsymbol{v}\|_1^1 + \boldsymbol{\beta}\|\boldsymbol{Gu}\|_1$, for $\boldsymbol{u} \geq \boldsymbol{0}$

$$\Longleftrightarrow \qquad \begin{array}{rl} \text{minimize} & \langle \mathbb{1}, \boldsymbol{x}^+\rangle + \langle \mathbb{1}, \boldsymbol{x}^-\rangle + \langle \mathbb{1}, \boldsymbol{y}^+\rangle + \langle \mathbb{1}, \boldsymbol{y}^+\rangle \\ \text{subject to} & \boldsymbol{Au} - \boldsymbol{v} = \boldsymbol{x}^+ - \boldsymbol{x}^- \\ & \boldsymbol{\beta G u} = \boldsymbol{y}^+ - \boldsymbol{y}^- \\ & \boldsymbol{x}^+ \geq \boldsymbol{0},\ \boldsymbol{x}^- \geq \boldsymbol{0},\ \boldsymbol{y}^+ \geq \boldsymbol{0},\ \boldsymbol{y}^- \geq \boldsymbol{0},\ \boldsymbol{u} \geq \boldsymbol{0} \end{array}$$

Linear Program (LP): $\boxed{\text{minimize } \langle \boldsymbol{c}, \boldsymbol{z}\rangle \text{ subject to } \boldsymbol{Hz} = \boldsymbol{b} \text{ and } \boldsymbol{z} \geq \boldsymbol{0}}$

- minimize $\mathcal{F}(\boldsymbol{u}) = \|\boldsymbol{Au} - \boldsymbol{v}\|_2^2 + \boldsymbol{\beta}\|\boldsymbol{Gu}\|_1$, for $\boldsymbol{u} \geq \boldsymbol{0}$

$$\Longleftrightarrow \qquad \begin{array}{rl} \text{minimize} & \|\boldsymbol{Au} - \boldsymbol{v}\|_2^2 + \langle \mathbb{1}, \boldsymbol{y}^+\rangle + \langle \mathbb{1}, \boldsymbol{y}^+\rangle \\ \text{subject to} & \boldsymbol{\beta G u} = \boldsymbol{y}^+ - \boldsymbol{y}^- \\ & \boldsymbol{y}^+ \geq \boldsymbol{0},\ \boldsymbol{y}^- \geq \boldsymbol{0},\ \boldsymbol{u} \geq \boldsymbol{0} \end{array}$$

Quadratic Program (QP): $\boxed{\text{minimize } \tfrac{1}{2}\boldsymbol{z}^*\boldsymbol{Qz} + \langle \boldsymbol{c}, \boldsymbol{z}\rangle,\ \boldsymbol{Q} \succ \boldsymbol{0}\ :\ \boldsymbol{Hz} = \boldsymbol{b} \text{ and } \boldsymbol{z} \geq \boldsymbol{0}}$

Solve LP and QP problems by interior point method + preconditioning

D. Minimization of TV regularized energy using duality

(Chambolle's method 04 to minimize $\mathcal{F}_v(u) = \|u - v\|^2 + \beta \mathbf{TV}(u)$)

- Image $u \in \mathbf{R}^{n\times n}$
- $\mathbf{div} : \mathrm{R}^{n\times n} \times \mathrm{R}^{n\times n} \to \mathrm{R}^{n\times n}$ (matrix operator)

  $(\mathrm{div}\ \xi)_{i,j} = \left(\xi^1_{i,j} - \xi^1_{i-1,j}\right) + \left(\xi^2_{i,j} - \xi^2_{i,j-1}\right)$ (+boundary conditions)
- $K = \left\{ \mathbf{div}\ \xi : \xi = (\xi^1, \xi^2) \in \mathrm{R}^{n\times n} \times \mathrm{R}^{n\times n}, \sqrt{\left(\xi^1_{i,j}\right)^2 + \left(\xi^2_{i,j}\right)^2} \leq 1, \forall (i,j) \right\}$ (closed convex)
- Discrete version of TV (§3.3)

$$\Phi(u) = \mathbf{TV}(u) = \sup \left\{ \langle u, z \rangle : z \in K \right\} \approx \|\nabla u\|_2$$

- $\Phi^*(w) \stackrel{def}{=} \left\{ \langle u, w \rangle - \sup_{z\in K} \langle u, z \rangle \right\} = \left\{ \begin{array}{ll} 0 & \text{if } w \in K \\ \infty & \text{if } w \notin K \end{array} \right\} = I_K(w)$

  (convex conjugate) (indicator function)
- $\hat{u} = \arg\min_u \mathcal{F}_v(u) \iff 0 \in \partial \mathcal{F}_v(u) \iff 0 \in \hat{u} - v + \frac{\beta}{2} \partial \Phi(\hat{u})$

$$\iff \hat{w} \stackrel{def}{=} \frac{2(v - \hat{u})}{\beta} \in \partial \Phi(\hat{u})$$

Property of subdifferentials: $\boxed{g \in \partial\Phi(u) \iff u \in \partial\Phi^*(g)}$ (Φ^* is the convex conjugate of Φ)

$$\Leftrightarrow \hat{\boldsymbol{u}} \in \boldsymbol{\partial\Phi^*}\left(\hat{\boldsymbol{w}}\right) \;\Leftrightarrow\; \mathbf{0} \in \hat{\boldsymbol{w}} - \tfrac{\mathbf{2}}{\boldsymbol{\beta}}\boldsymbol{v} + \tfrac{\mathbf{2}}{\boldsymbol{\beta}}\boldsymbol{\partial\Phi^*}(\hat{\boldsymbol{w}})$$

$\equiv$ necessary and sufficient condition for

$$\begin{aligned} \hat{\boldsymbol{w}} &= \mathbf{arg\,min}_{\boldsymbol{w}} \tfrac{\mathbf{1}}{\mathbf{2}}\|\boldsymbol{w} - \tfrac{\mathbf{2}\boldsymbol{v}}{\boldsymbol{\beta}}\|^{\mathbf{2}} + \tfrac{\mathbf{2}}{\boldsymbol{\beta}}\boldsymbol{I_K}(\boldsymbol{w}) \\ &= \mathbf{arg}\min_{\boldsymbol{w}\in\boldsymbol{K}} \|\boldsymbol{w} - \frac{\mathbf{2}\boldsymbol{v}}{\boldsymbol{\beta}}\|^{\mathbf{2}} \end{aligned}$$

$$\Leftrightarrow \hat{\boldsymbol{w}} = \boldsymbol{\Pi_K}\left(\tfrac{\mathbf{2}\boldsymbol{v}}{\boldsymbol{\beta}}\right) \;\Leftrightarrow\; \hat{\boldsymbol{u}} = \boldsymbol{v} - \tfrac{\boldsymbol{\beta}}{\mathbf{2}}\boldsymbol{\Pi_K}\left(\tfrac{\mathbf{2}\boldsymbol{v}}{\boldsymbol{\beta}}\right)$$

- Algorithm $\equiv$ compute the (nonlinear) projection $\boldsymbol{\Pi_K}\left(\tfrac{\mathbf{2}\boldsymbol{v}}{\boldsymbol{\beta}}\right)$
- Constrained minimization—formulation using Karush-Kuhn-Tucker (KKT) conditions
- Detailed algorithm and convergence conditions—*[Chambolle 04]*
- **Extension to** $\|\boldsymbol{Au} - \boldsymbol{v}\|^{\mathbf{2}} + \boldsymbol{\beta}\mathbf{TV}(\boldsymbol{u})$**—see** *[Bect et al 04]*

E. SOCP for TV regularization

[Goldfarb &Yin 05]

$$\boldsymbol{\min TV(u)} \text{ subject to } \boldsymbol{u+n=v,} \quad \boldsymbol{\|n\|^2 \le \sigma^2}$$

Second order-cone $\left\{\boldsymbol{(u,t) : \|u\| \le t}\right\}$

Standard second-order cone propgramming (SOCP) form

$$\begin{array}{lll}
\min & \displaystyle\sum_{i,j} t_{i,j} & \\
\text{s.t.} & u+n=v & \\
& \|\nabla_{i,j} u\| \le t_{i,j} & \text{(second-order cone)} \\
& \|n\| \le \sigma & \text{(second-order cone)}
\end{array}$$

Different dual formulations can be derived

SOCP is solved in polynomial time by interior point methods

F. Graph-cut methods for discrete TV regularization *[Darbon &Sigelle 06]*

$\boldsymbol{u}$—discrete-value image, $\boldsymbol{u_i \in \{0, \ldots, L-1\}, \ \forall i}$

$$\boldsymbol{\mathcal{F}_v(u) = \|u - v\|_\rho^\rho + \beta \sum_{i,j \in \mathcal{N}_i} |u_i - u_j|, \quad \rho \in \{1, 2\}}$$

- $\boldsymbol{u_i^\ell = 1\!\!1_{u_i \leq \ell}} = \begin{cases} 1 & \text{if } \ u_i \leq \ell \\ 0 & \text{else} \end{cases}$ $\qquad \boldsymbol{u^\ell}$ is a binary image, $\forall \ell$
- reformulate $\mathcal{F}_v$ into a set of binary images

$$\boldsymbol{\mathcal{F}_v(u) = \sum_\ell \mathcal{F}_v^\ell(u^\ell)} + \text{const}$$

- Separate minimization for each $\boldsymbol{\ell}$

$$\boldsymbol{\hat{u}^\ell = \arg\min_{u \in B} \mathcal{F}_v^\ell(u), \quad B = \{1, 0\}^p} \quad \text{(the set of binary images)}$$

The first clue: fast computation using **graph-cut** methods

- Reconstruction $\boldsymbol{\hat{u}_i = \min \left\{ \ell : \hat{u}_i^\ell = 1 \right\}, \ \forall i}$
- The second clue: proof that $\boldsymbol{\hat{u}}$ solves the problem

10. Open questions

- Analyzing the properties of the minimizers in connection with the shape of the energy provides strong results
- The results provide an alternative way for rigorous modeling
- Conception of specialized energies
- Minimization methods accounting for the features of the solution
- Open field for research...
- Extension to images and signals in functional spaces is necessary to capture the geometry
- What "features" and what "properties" ?
- Properties of solutions as a function of the randomness of the data ?
- Ultimate goal : conceive solutions that match pertinent models

11. References (partial list)

General references

G. Aubert and P. Kornprobst, *Mathematical problems in images processing*, Springer, 2nd ed., 2006

A. Blake and A. Zisserman, *Visual Reconstruction*, MIT Press, Cambridge, MA, 1987

P. G. Ciarlet, *Introduction à l'analyse numérique matricielle et à l'optimisation*, 5th ed., Dunod, 2000

J.-B. Hiriart-Urruty and C. Lemarchal, *Convex analysis and Minimization Algorithms, vol. I and II,* 1996, Springer-Verlag (Berlin)

R. T. Rockafellar, *Convex Analysis*, Princeton University Press, 2nd ed., 1970

R. T. Rockafellar and J. B. Wets, *Variational analysis*, 1997, Springer-Verlag (New York)

J. Weickert, *Anisotropic Diffusion in Image Processing*, 1998, Ed. B.G. Teubner, Stuttgart

Research papers

F. Alberge, P. Duhamel and M. Nikolova, *Adaptive Solution for Blind Identification/Equalization Using Deterministic Maximum Likelihood*, IEEE Trans. on Signal proc., 50(4), 2002, p.. 923-936

F. Alberge, M. Nikolova and P. Duhamel, *Blind Identification/Equalization Using Deterministic Maximum Likelihood and a Partial Prior on the Input*, IEEE Trans. on Signal proc., 54(2), 2006, p.. 724-737

M. Allain, J. Idier and Y. Goussard, *On Global and Local Convergence of Half-Quadratic Algorithms*, IEEE Trans. on Image Proc., 15(5), 2006, p. 1130-1142

S. Alliney,*Digital Filters as Absolute Norm Regularizers*, Trans. on IEEE Signal Processing, 42(6), 1992, pp 1548-1562

A. Alliney,*Recursive median filters of increasing order: a variational approach*, Trans. on IEEE Signal Processing, 44(6), 1996, p. 1346-1354

A. Alliney, *A property of the minimum vectors of a regularizing functional defined by means of absolute norm*, Trans. on IEEE Signal Processing, 45(4), 1997, p. 913-917

F. Alter, V. Caselles and A. Chambolle, *Evolution of characteristic functions of convex sets in the plane by minimizing total variation flow*, Interfaces Free Bound., 7(1), 2005, p. 2953

F. Alter, V. Caselles and A. Chambolle, *A characterization of convex calibrable sets in* $\mathcal{R}^N$, Math. Ann., 332(2), 2005, p. 329-366

J.-F. Aujol and A. Chambolle, *Dual norms and image decomposition models*, Int. J. of Computer Vision, 63(1), 2005, p. 85-104

J. Bect, L. Blanc-Fŕaud, G. Aubert et A. Chambolle, *A* ℓ^1 *unified variational framework for image restoration*, Lecture Notes in Computer Science, Computer Vision - ECCV 2004, Springer, p. 1-13

J. Besag, *Digital Image Processing : Towards Bayesian Image Analysis*, Journal of Applied Statistics, 16(3), 1989, p. 395-407

M. Black and A. Rangarajan, *On the Unification of Line Processes, Outlier Rejection, and Robust Statistics with Applications to Early Vision*, Int. J. of Computer Vision, 19(1), 1996, p. 57-91

Y. Bobichon and A. Bijaoui, *Regularized multiresolution methods for astronomical image enhancement*, Experiment. Astronom., 7, 1997, p. 239-255

V. Caselles, A. Chambolle and M. Novaga, *The discontinuity set of solutions of the TV denoising problem and some extensions*, to appear in SIAM J. on Multiscale Modeling and Simulation

A. Chambolle, *An Algorithm for Total Variation Minimization and Applications*, J. of Math. Imaging and Vision vol. 20, no. 1-2, 2004, p. 89-97

R. Chan, C. Hu and M. Nikolova, *An Iterative Procedure for Removing Random-Valued Impulse Noise*, IEEE Signal Proc. Letters, 11, 2004, p. 921-924

R. Chan, H. Chung-Wa and M. Nikolova, *Impulse Noise Removal by Median-type Noise Detector and Edge-preserving Regularization*, IEEE Trans. on Image Proc.,, 14(10), 2005, p. 1479-1485

T. Chan and S. Esedoglu, *Aspects of Total Variation Regularized L^1 Function Approximation*, SIAM J. Appl. Math., 65 (2005), p. 1817-1837

T. Chan, S. Esedoglu and M. Nikolova, *Algorithms for Finding Global Minimizers of Image Segmentation and Denoising Models*, SIAM J. Applied Maths, vol. 66, n. 5, 2006, p.1632-1648

P. L. Combettes and J.-C. Pesquet, *Image Restoration Subject to a Total Variation Constraint*, IEEE Trans. on Image Proc., 13(9), 2004, p. 1213-1222

R. R. Coifman and A. Sowa, *Combining the calculus of variations and wavelets for image enhancement*, Appl. Comput. Harmon. Anal., 9 (2000), p. 1-18

J. Darbon and M. Sigelle, *Image Restoration with Discrete Constrained Total Variation Part I: Fast and Exact Optimization*, Journal of Mathematical Imaging and Vision, 26(3), 2006, p. 261-276

S. Durand and M. Nikolova, *Stability of the Minimizers of Least Squares with a Non-Convex Regularization. Part I: Local Behavior*, J. of Applied Math. and Optimization, 53(2), 2006, p. 185-208

S. Durand and M. Nikolova, *Stability of the Minimizers of Least Squares with a Non-Convex Regularization. Part II: Global Behavior*, J. of Applied Math. and Optimization, 53(3), 2006, p. 259-277

S. Durand and M. Nikolova, *Denoising of frame coefficients using L1 data-fidelity term and edge-preserving regularization*, SIAM J. of Multiscale Modeling and Simulation, 6(2), 2007, p. 547-576.

S. Durand and J. Froment, *Reconstruction of wavelet coefficients using total variation minimization*, SIAM J. Sci. Comput., 24 (2003), p. 17541767

S. Geman and D. Geman, *Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images*, IEEE Trans. on Pattern An. and Machine Intelligence, PAMI-6(6), 1984, p. 721-741

D. Geman and G. Reynolds, *Constrained Restoration and Recovery of Discontinuities*, IEEE Trans. on Pattern An. and Machine Intelligence, PAMI-14(3), 1992, p. 367-383

D. Geman and C. Yang, *Nonlinear Image Recovery with Half-Quadratic Regularization*, IEEE Trans. on Image Proc., IP-4(7), 1995, p. 932-946

D. Goldfarb and W. Yin, *Second-order Cone Programming Methods for Total Variation-Based Image Restoration*, SIAM Journal on Scientific Computing, 27(2), 2005, p. 622-645

Y. Gousseau and J.-M. Morel, *Are natural images of bounded variation ?*, SIAM J. of Mathematical Analysis, 33, 3, 2001, pp 634-648

F. H. Haoying, M. Ng, M. Nikolova and J. Barlow, *Efficient minimization methods of mixed L1-L1 and L2-L1 norms for image restoration*, SIAM Journal on Scientific computing, 27(6), 2006, p. 1881-1992

C. Labat et J. Idier, *Convergence of truncated half-quadratic and Newton algorithms, with application to image restoration*, Rapp. tech., IRCCyN, June 2007

F. Malgouyres, *Rank related properties for Basis Pursuit and total variation regularization*, Signal Processing, 87(11), 2007, p. 2695-2707

F. Malgouyres, *Projecting onto a polytope simplifies data distributions*, University Paris 13 preprint, num. 2006-01, January 2006

F. Malgouyres, *Image compression through a projection onto a polyhedral set*, J. of Mathematical Imaging and Vision, 27(2), 2007, p. 193-200

Y. Meyer, *Oscillating patterns in image processing and in some nonlinear evolution equations*, 2001, The Fifteenth Dean Jacquelines B. Lewis Memorial Lectures

M. Nikolova, *Thresholding implied by truncated quadratic regularization*, IEEE Trans. on Signal Processing, 48, 2000, p. 3437-3450

M. Nikolova, *Estimation of binary images using convex criteria*, Proc. of IEEE Int. Conf. on Image Processing, Oct. 1998

M. Nikolova, *Local strong homogeneity of a regularized estimator*, SIAM Journal on Applied Mathematics, 61(2), 2000, p. 633-658

M. Nikolova, *Image restoration by minimizing objective functions with non-smooth data-fidelity terms*, IEEE Int. Conf. on Computer Vision / Workshop on Variational and Level-Set Methods, Jul. 2001

M. Nikolova, *Minimizers of cost-functions involving non-smooth data-fidelity terms. Application to the processing of outliers*, SIAM Journal on Numerical Analysis, 40(3), 2002, p. 965-994

Nikolova M., *Weakly constrained minimization. Application to the estimation of images and signals involving constant regions*, J. of Math Imaging and Vision, 21(2), 2004, p. 155-175

M. Nikolova, *A variational approach to remove outliers and impulse noise*, Journal of Mathematical Imaging and Vision, 20(1-2), 2004, p. 99-120

M. Nikolova, *Model distortions in Bayesian MAP reconstruction*, AIMS Journal on Inverse Problems and Imaging, 1(2), 2007, p. 399-422

M. Nikolova, *Analytical bounds on the minimizers of (nonconvex) regularized least-squares*, AIMS Journal on Inverse Problems and Imaging, 2007 (to appear)

S. Osher, M. Burger, D. Goldfarb, J. Xu, and W. Yin, *An iterative regularization method for total variation based image restoration*, SIAM J. on Multiscale Model Simul., 4(2), 2005, p. 460489

L. Rudin, S. Osher, and C. Fatemi, *Nonlinear total variation based noise removal algorithm*, Physica vol. 60 D, 1992, p. 259-268

O. Scherzer and J. Weickert, *Relations Between Regularization and Diffusion Filtering*, J. of Mathematical Imaging and Vision, 12(1), 2000, p. 43-63

G. Steidl, J. Weickert, P. Mrázek and M. Welk, *On the equivalence of soft wavelet shrinkage, total variation diffusion, total variation regularization, and SIDEs*, SIAM Journal on Numerical Analysis, 42(2), 2004, p. 686-713

L. Tenorio, *Statistical regularization of inverse problems*, SIAM Review, 43, 2001, p. 347366.

L. Vese and S. Osher, *Modeling textures with Total Variation minimization and oscillating patterns in image processing*, Journal of Scientific Computing, 19(1-3), 2003 , p. 553-572