Qualitative features of the minimizers of energies and implications on modeling

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

Collaborators: S. Durand, R. Chan, M. Ng, F. Malgouyres, T. Chan, S. Esedoglu

Plenary - SIAM - IS 2008 (July) - San Diego

Outline

1. Energy minimization methods
2. Regularity of minimizers
3. Non-smooth regularization
4. Non-smooth data-fidelity
5. Nonsmooth data-fidelity and regularization
6. Non-convex regularization
7. Models are distorted by Bayesian MAP
8. Open questions

1. Energy minimization methods

1.1 The energy method

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

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

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

$\searrow$ coherent with priors and desiderata

$$\hat{u} = \underset{u\in\Omega}{\operatorname{argmin}} \mathcal{F}_v(u)$$

Ω - constraints

$$\mathcal{F}_v(u) = \Psi(u,v) + \beta\Phi(u)$$

energy $\qquad$ data-fidelity $\qquad$ prior

Related formulation: minimize $\left\{\Phi(u) \text{ such that } u \in \Omega, \quad \Psi(u,v) \leq \tau\right\}$

$\beta = \beta(\tau) > 0$ - Lagrange parameter

Applications: Denoising, Segmentation, Deblurring, Tomography, Seismic imaging, Zooming, Superresolution, Compression, Learning, Motion estimation, etc.

1.2 Background to define energies

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

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

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

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

Ψ is a degradation model, good prior is needed to compensate for the loss of information

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

- Bayesian approach $\boldsymbol{\Phi(u) = \sum_i \varphi(\|G_i u\|)}$
- Variational approach $\boldsymbol{\Phi(u) = \int_\Omega \varphi(\|\nabla u\|)dx}$

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

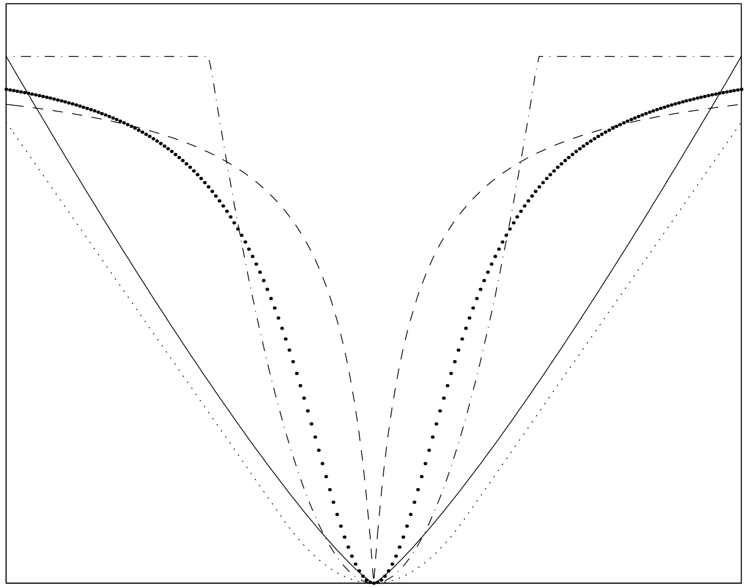
Both approaches lead to similar energies

Well-known difficulties to control the solution

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

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

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



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

φ is edge-preserving if $\lim\limits_{t\to\infty} \dfrac{\varphi'(t)}{t} = 0$ - a frequent requirement

$\varphi(t) = \min\{\alpha t^2, 1\}$ - discrete version of Mumford-Shah

ℓ_0-norm: $\|u\|_0 = \#\{i : u_i \neq 0\} = \sum_i \varphi(u_i)$ for $\varphi(0)=0, \ \varphi(t)=1$ if $t \neq 0$

1.3 Minimizer approach

(the core of this plenary)

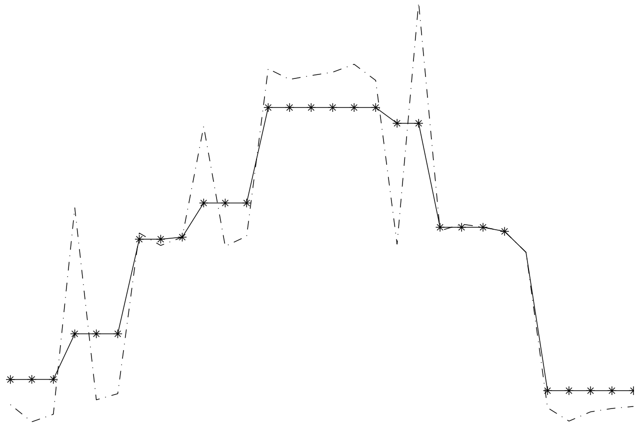
Analysis of the main features exhibited by the (local) minimizers $\hat{u}$ of $\mathcal{F}_v$ as a function of the shape of $\mathcal{F}_v$

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

$\Rightarrow$ A new rigorous approach for modeling:

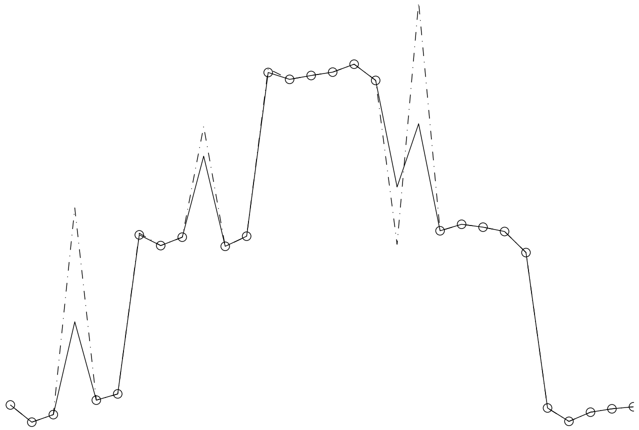
Choose $\mathcal{F}_v$ so that the properties of $\hat{u}$ match the models for the data and the unknown

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



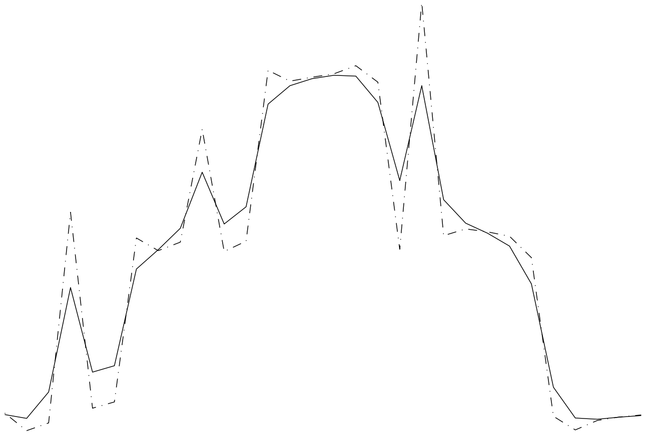
STAIR-CASING

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



EXACT DATA-FIT

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



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

Data (—. — .—), Minimizer (—)

2 Regularity results

We focus on

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

for $\boldsymbol{u \in \mathbb{R}^p}$, $\boldsymbol{v \in \mathbb{R}^q}$

$\boldsymbol{\{G_i\}}$ linear operators $\mathbb{R}^p \to \mathbb{R}^s$, $s \geq 1$ (typically $s = 1$ or $s = 2$, e.g. $G_i \approx \nabla_i$)

If $\varphi'(0) > 0 \quad \Rightarrow \quad \Phi$ is nonsmooth on $\bigcup_i \{u : G_i u = 0\}$

Remind:

$\mathcal{F}_v$ has a (local) minimum at $\hat{u} \;\Rightarrow\; \delta\mathcal{F}_v(\hat{u})(d) = \lim_{t \downarrow 0} \frac{\mathcal{F}_v(u+td) - \mathcal{F}_v(u)}{t} \geq 0, \;\forall d \in \mathbb{R}^p$

Definition: $\boldsymbol{\mathcal{U} : O \to \mathbb{R}^p}$, $\boldsymbol{O \subset \mathbb{R}^q}$ **open, is (strict) local minimizer function if** $\boldsymbol{\forall v \in O}$, $\boldsymbol{\mathcal{F}_v}$ **has a (strict) local minimum at** $\boldsymbol{\mathcal{U}(v)}$

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

2.1 Stability of the minimizers of $\mathcal{F}_v$ *[Durand &Nikolova 06]*

Assumptions: $\varphi : \mathbb{R}_+ \to \mathbb{R}$ is continuous and $\mathcal{C}^{m\geq 2}$ on $\mathbb{R}_+ \setminus \{\theta_1, \theta_2 \ldots\}$, edge-preserving, possibly non-convex and $\mathrm{rank}(A) = p$

A. LOCAL MINIMIZERS

`(knowing local minimizers is important)`

We prove that there is a **closed** subset $\boldsymbol{N} \subset \mathbb{R}^q$ whose Lebesgue measure is $\mathbb{L}^{\boldsymbol{q}}(\boldsymbol{N}) = \mathbf{0}$ such that $\forall \boldsymbol{v} \in \mathbb{R}^{\boldsymbol{q}} \setminus \boldsymbol{N}$, for **every** local minimizer $\hat{\boldsymbol{u}}$ of $\boldsymbol{\mathcal{F}_v}$, there is a $\mathcal{C}^{m-1}$ strict local minimizer function $\boldsymbol{\mathcal{U}} : \boldsymbol{O} \to \mathbb{R}^{\boldsymbol{p}}$, with $O \subset \mathbb{R}^q$—open, such that $v \in O \subset \mathbb{R}^q$ and $\hat{\boldsymbol{u}} = \boldsymbol{\mathcal{U}}(\boldsymbol{v})$.

MORE PRECISIONS

<u>Φ *smooth*</u>

$\forall \boldsymbol{v} \in \mathbb{R}^{\boldsymbol{q}} \setminus \boldsymbol{N}$, every local minimizer $\hat{\boldsymbol{u}}$ of $\boldsymbol{\mathcal{F}_v}$ is strict and $\boldsymbol{\nabla^2 \mathcal{F}_v(\hat{u}) \succ 0}$ (positive definite)

$\boldsymbol{\Phi}$ *piecewise smooth*

Notations: $\hat{\boldsymbol{h}} = \{\boldsymbol{i} : \boldsymbol{G_i \hat{u}} = \boldsymbol{0}\}$ and $\boldsymbol{K_{\hat{h}}} = \{\boldsymbol{w} \in \mathbb{R}^{\boldsymbol{p}} : \boldsymbol{G_i w} = \boldsymbol{0}, \forall \boldsymbol{i} \in \hat{\boldsymbol{h}}\}$

Notice that $\boldsymbol{\mathcal{F}_v}$ is smooth on $\boldsymbol{K_{\hat{h}}}$ near $\hat{\boldsymbol{u}} \in \boldsymbol{K_{\hat{h}}}$

$\forall \boldsymbol{v} \in \mathbb{R}^{\boldsymbol{q}} \setminus \boldsymbol{N}$, every local minimizer $\hat{\boldsymbol{u}}$ of $\boldsymbol{\mathcal{F}_v}$ is strict and fits the Sufficient Condition (SC)

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

B. Global minimizers

We prove that there is a subset $\hat{\boldsymbol{N}} \subset \mathbb{R}^q$ with $\mathbb{L}^{\boldsymbol{q}}(\hat{\boldsymbol{N}}) = \boldsymbol{0}$ and $\text{Int}(\mathbb{R}^q \setminus \hat{N})$ dense in $\mathbb{R}^q$ such that $\forall \boldsymbol{v} \in \mathbb{R}^{\boldsymbol{q}} \setminus \hat{\boldsymbol{N}}$, $\boldsymbol{\mathcal{F}_v}$ has a **unique** global minimizer.

Moreover, there is an open subset of $\mathbb{R}^q \setminus \hat{N}$, dense in $\mathbb{R}^q$, where the global minimizer function $\hat{\mathcal{U}}$ is $\mathcal{C}^{m-1}$-continuous.

2.2 Nonasymptotic bounds on minimizers *[Nikolova 07]*

Classical bounds hold for $\boldsymbol{\beta} \searrow \mathbf{0}$ or $\boldsymbol{\beta} \nearrow \infty$

Assumptions: φ *is* $\mathcal{C}^1$ *on* $\mathbb{R}_+$ *with* $\boldsymbol{\varphi'(t) \geq 0,\ \forall t \geq 0}$ **or** $\boldsymbol{\varphi(t) = \min\{\alpha t^2, 1\}}$

Notations: $\mathbb{1} = [\mathbf{1}, .., \mathbf{1}]$ and $\boldsymbol{G} = [\boldsymbol{G}_1^*, \ldots, \boldsymbol{G}_r^*]^*$ where * means transposed

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

If $\mathrm{rank}(A) = p$ **or** $\varphi'(t) > 0, \forall t > 0$, then every local minimizer $\hat{\boldsymbol{u}}$ of $\mathcal{F}_v$ satisfies

$$\|\boldsymbol{A\hat{u}}\| \leq \|\boldsymbol{v}\|$$

If $\mathrm{rank}(A) = p$, $\varphi'(t) > 0, \forall t > 0$ & $\ker G = \mathrm{span}(\mathbb{1})$, then $\exists \boldsymbol{N} \subset \mathbb{R}^q$ with $\mathbb{L}^{\boldsymbol{q}}(\boldsymbol{N}) = \mathbf{0}$,

$$\forall \boldsymbol{v} \in \mathbb{R}^{\boldsymbol{q}} \setminus \boldsymbol{N}, \quad \|\boldsymbol{A\hat{u}}\| < \|\boldsymbol{v}\|$$

Remind: if $\{G_i u\}_{i=1}^r$ are discrete gradients or 1st-order differences, then $\ker G = \mathrm{span}(\mathbb{1})$

B. The mean of restored data

Usually $\mathbf{mean(noise)}=\mathbf{mean}(Au-v)=\mathbf{0}$.

If $\boldsymbol{A}\mathbb{1}_{\boldsymbol{p}} \propto \mathbb{1}_{\boldsymbol{q}}$ and $\mathbb{1}_{\boldsymbol{q}} \in \mathbf{ker}(\boldsymbol{G})$, then $\mathbf{mean}(\boldsymbol{A}\hat{\boldsymbol{u}} - \boldsymbol{v}) = \mathbf{0}$ $\quad u \in \mathbb{R}^p, \ v \in \mathbb{R}^q$.

However, in general $\mathbf{mean}(\boldsymbol{A}\hat{\boldsymbol{u}} - \boldsymbol{v}) \neq \mathbf{0}$.

C. Residuals for edge-preserving regularization

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

If $\mathrm{rank}(A) = q$, then for every $\boldsymbol{v} \in \mathbb{R}^{\boldsymbol{q}}$, every local minimizer $\hat{\boldsymbol{u}}$ of $\boldsymbol{\mathcal{F}_v}$ satisfies

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

We can always take $\|\varphi'\|_\infty = 1$. Then

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

Surprising?

3 Minimizers under Non-Smooth Regularization

3.1 General case

[Nikolova 97,01,04]

Consider $\mathcal{F}_v(u)=\Psi(u,v)+\beta\sum_{i=1}^{r}\varphi(\|G_i u\|)$ for $\Psi\in\mathcal{C}^{m\geq 2}$, $\varphi\in\mathcal{C}^m(\mathbb{R}_+^*)$, $0<\varphi'(0)\leq\infty$

(Assumptions) If $\hat{u}$ is a local minimizer of $\mathcal{F}_v$, then $\exists\ O\subset\mathbb{R}^q$ open, $v\in O$, $\exists\ \mathcal{U}\in\mathcal{C}^{m-1}$

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

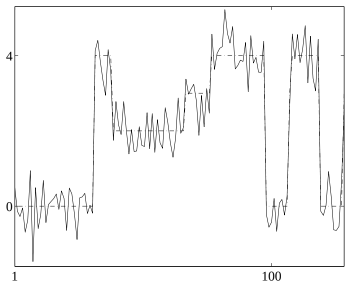
where $\hat{h}=\{i : G_i\hat{u}=0\}$

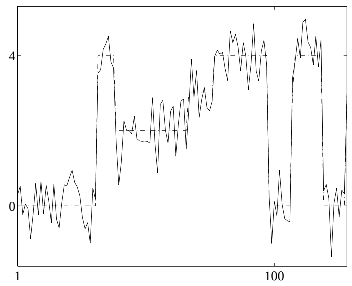
By § 2.1: if $\Psi(u,v)=\|Au-v\|^2$, $\operatorname{rank}(A)=p \;\Rightarrow$ assumptions hold $\forall v\in\mathbb{R}^q\setminus N$

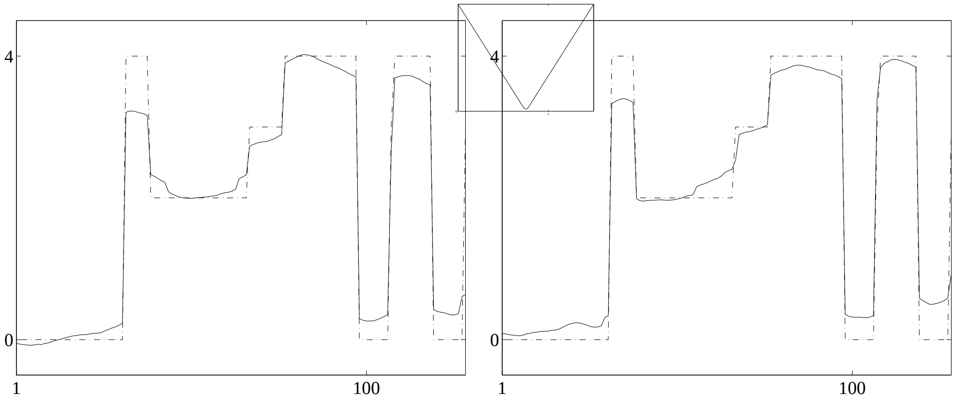
$\hat{h}\subset\{1,..,r\}$ $\qquad$ $\boxed{\mathcal{O}_{\hat{h}}\stackrel{\text{def}}{=}\{v\in\mathbb{R}^q : G_i\mathcal{U}(v)=0,\ \forall i\in\hat{h}\}}$ $\;\Rightarrow\; \mathbb{L}^q(\mathcal{O}_{\hat{h}})>0$

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

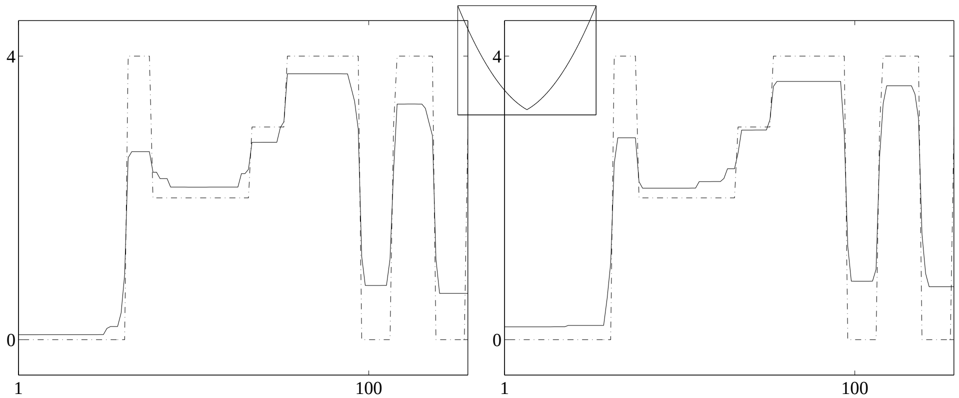
$G_i\approx\nabla_{u_i}\;\Rightarrow\;\hat{h}\equiv$constant regions $\Rightarrow$ **stair-casing:** $\hat{u}_i=\hat{u}_j$ for many neighbors (i,j)



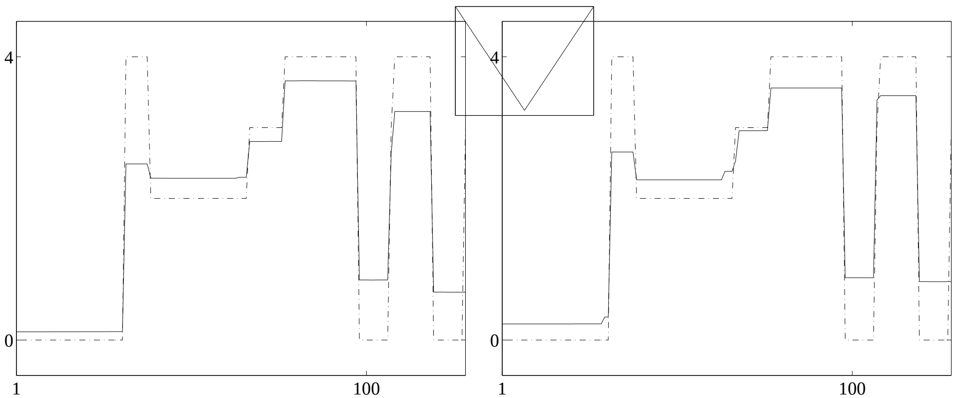




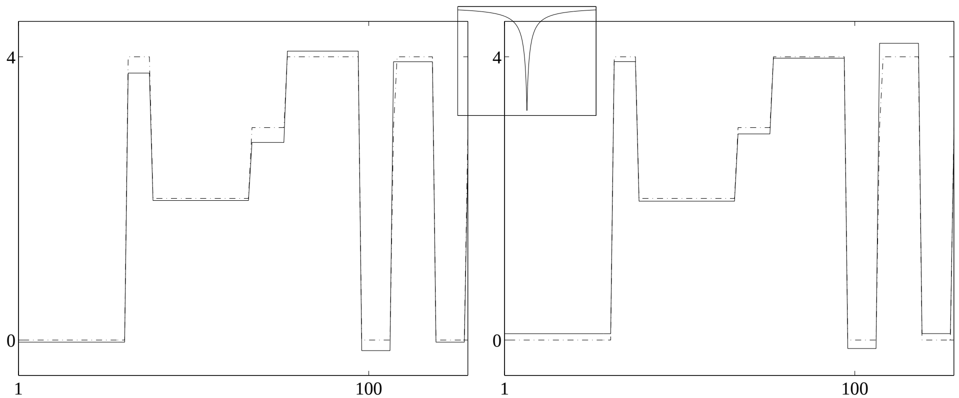
$\varphi(t)=\frac{t^2}{2}$ if $t \leq \alpha$, $\varphi(t)=\alpha t-\frac{\alpha^2}{2}$ if $t > \alpha$ **(smooth)**



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



$\varphi(t) = |t|$



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

$\boldsymbol{\mathcal{O}_{\hat{h}}}$ for Hard Thresholding

$$\boldsymbol{\mathcal{F}_v(u) = \|u - v\|_\alpha^\alpha + \beta\|u\|_0, \quad \alpha \geq 1}$$

$$\mathcal{U}_i(v) = \begin{cases} 0 & \text{if} \quad |v_i|^\alpha \leq \beta \\ v_i & \text{if} \quad |v_i|^\alpha > \beta \end{cases} \quad \Rightarrow \quad \hat{h} = \big\{i : \mathcal{U}_i(v) = 0\big\} = \big\{i : |v_i|^\alpha \leq \beta\big\}$$

$$\boldsymbol{\mathcal{O}_h = \{v \in \mathbb{R}^q : \ |v_i|^\alpha \leq \beta, \forall i \in h, \ |v_i|^\alpha > \beta, \forall i \in h^c\}; \ \mathbb{R}^q = \bigcup_h \mathcal{O}_h}$$

Data yielding K-sparse solutions: $\mathcal{S}_K = \big\{v \in \mathbb{R}^q : \|\hat{u}\|_0 \leq K\big\} = \bigcup_{h:\#h \leq K} \mathcal{O}_h$

3.3 Measuring the sparsest approximations *[Malgouyres & Nikolova 08]*

The most economical way to represent data $\boldsymbol{v}$ in a frame $W=\{w_i\}$ – find $\hat{\boldsymbol{u}}$ that solves

$$\min \|\boldsymbol{u}\|_0 \text{ subject to } \|\textstyle\sum_i \boldsymbol{u}_i \boldsymbol{w}_i - \boldsymbol{v}\| \leq \boldsymbol{\tau}, \quad \tau \geq 0, \ \|.\| \text{ a norm}$$

$\boldsymbol{S_K} \doteq \{\boldsymbol{v} \in \mathbb{R}^{\boldsymbol{q}} : \|\hat{\boldsymbol{u}}\|_0 \leq \boldsymbol{K}\}$ - all data yielding a K-sparse solution.

$\mathcal{D}_\theta \doteq \{v \in \mathbb{R}^q : f(v) \leq \theta\}$ where f is a norm. We derive

$$C_K \left(\frac{\tau}{\theta}\right)^{q-K} \theta^q \left(1 - \delta\frac{\tau}{\theta}\right) - \theta^q \varepsilon(K,\tau,\theta) \leq \mathbb{L}^q(\boldsymbol{S_K} \cap \boldsymbol{\mathcal{D}_\theta}) \leq C_K \left(\frac{\tau}{\theta}\right)^{q-K} \theta^q \left(1 + \delta\frac{\tau}{\theta}\right)$$

C_K depends on $(\|.\|, f, W)$, while δ depends on $(\|.\|, f)$ (explicit formulae for C_K, δ)

Typically $\boldsymbol{\tau} \ll \boldsymbol{\theta}$ and then $\mathbb{L}^{\boldsymbol{q}}(\boldsymbol{S_K} \cap \boldsymbol{\mathcal{D}_\theta}) = \boldsymbol{C_K} \left(\frac{\boldsymbol{\tau}}{\boldsymbol{\theta}}\right)^{\boldsymbol{q-K}} \boldsymbol{\theta}^{\boldsymbol{q}} + \boldsymbol{\theta}^{\boldsymbol{q}}\ \boldsymbol{o}\left(\left(\frac{\boldsymbol{\tau}}{\boldsymbol{\theta}}\right)^{\boldsymbol{q-K}}\right)$

Assumption: data v are uniform on $\mathcal{D}_\theta$, then $\text{Proba}(\|\hat{u}\|_0 \leq K) = \frac{\mathbb{L}^q(S_K \cap \mathcal{D}_\theta)}{\theta^N \mathbb{L}(\mathcal{D}_1)}$

Goal: build a coder—choose $\|.\|$ and W—such that

$$\textbf{Proba}(\|\hat{\boldsymbol{u}}\|_0 \leq \boldsymbol{K}) \textbf{ is large if } \boldsymbol{K} \ll \boldsymbol{q} \textbf{ and small if } \boldsymbol{K} \to \boldsymbol{q}$$

3.4 Recovery of Quasi-Binary images and signals *[Nikolova 98]*

- The sought image is binary $u \in \{0,1\}^p$, data $v = Au + n$

  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

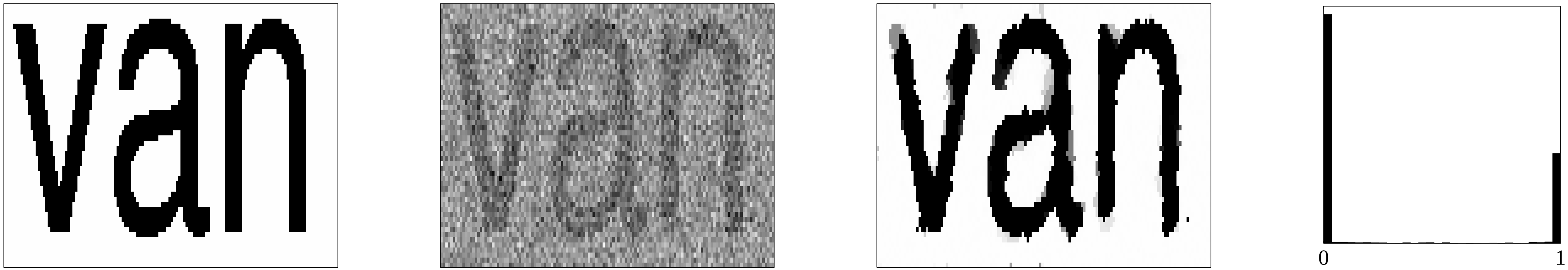
  Graph-cuts : difficult if $A \neq$ identity

- Instead, define continuous-valued quasi-binary minimizers of **convex** $\mathcal{F}_v$:
  discourage nonbinary values & enforce stair-casing

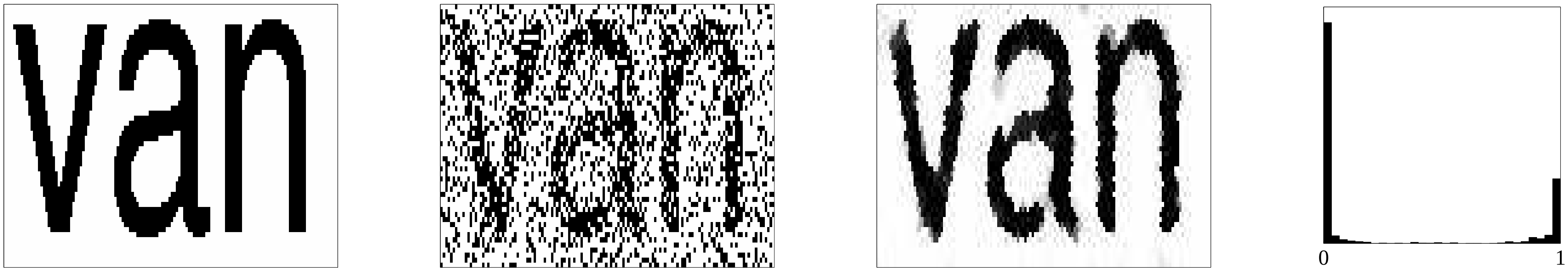
minimize $\mathcal{F}_v$ subject to $u \in [0,1]^p$ (convex constraint)

$$\mathcal{F}_v(u) = \sum_i \big((Au)_i - v_i\big)^2 - \gamma \sum_i \left(u_i - \frac{1}{2}\right)^2 + \beta \sum_{j \in \mathcal{N}_i} |u_i - u_j| , \quad \gamma \lessapprox \lambda_{\min}(A)$$

Applications for blind channel estimation *[Alberge, Nikolova, Duhamel 02, 06]*



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



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

4 Minimizers relevant to non-smooth data-fidelity

4.1 General case

[Nikolova 01,02]

Consider $$\mathcal{F}_v(u)=\sum_{i=1}^{q}\psi(|\langle a_i,u\rangle-v_i|)+\beta\Phi(u) \text{ for } \Phi\in\mathcal{C}^m,\ \psi\in\mathcal{C}^m(\mathbb{R}_+),\quad \boldsymbol{\psi'(0)>0}$$

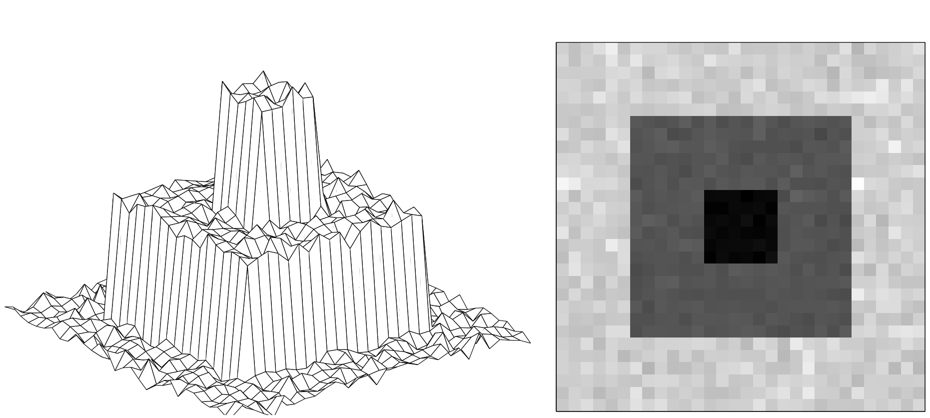
(Assumptions) If $\hat{\boldsymbol{u}}$ is a local minimizer of $\mathcal{F}_v$, then $\exists\ \boldsymbol{O}\subset\mathbb{R}^q$ open, $v\in O$, $\exists\ \boldsymbol{\mathcal{U}}\in\mathcal{C}^{m-1}$

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

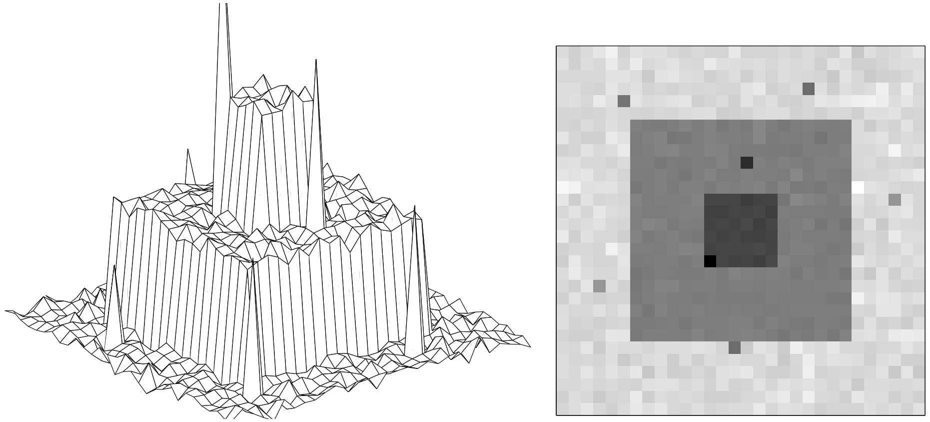
where $\hat{\boldsymbol{h}}=\{\boldsymbol{i}:\langle\boldsymbol{a_i},\hat{\boldsymbol{u}}\rangle=\boldsymbol{v_i}\}$

$$\hat{\boldsymbol{h}}\subset\{1,..,q\} \qquad \boxed{\mathcal{O}_{\hat{h}}\doteq\left\{v\in\mathbb{R}^q:\langle a_i,\mathcal{U}(v)\rangle=v_i,\forall i\in\hat{h}\right\}}\ \Rightarrow\ \mathbb{L}^q(\mathcal{O}_{\hat{h}})>0$$

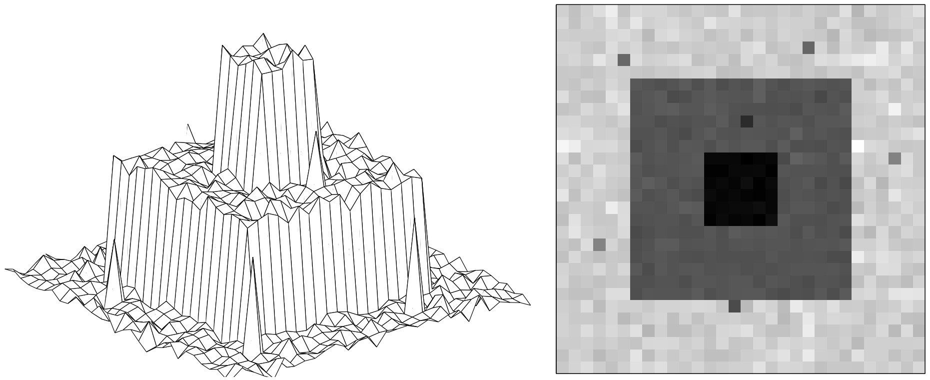
$\Rightarrow$ Noisy data $\boldsymbol{v}$ yield (local) minimizers $\hat{\boldsymbol{u}}$ of $\boldsymbol{\mathcal{F}_v}$ which achieve an **exact fit to data** $\langle\boldsymbol{a_i},\hat{\boldsymbol{u}}\rangle=\boldsymbol{v_i}$ for a certain number of indexes $\boldsymbol{i}$



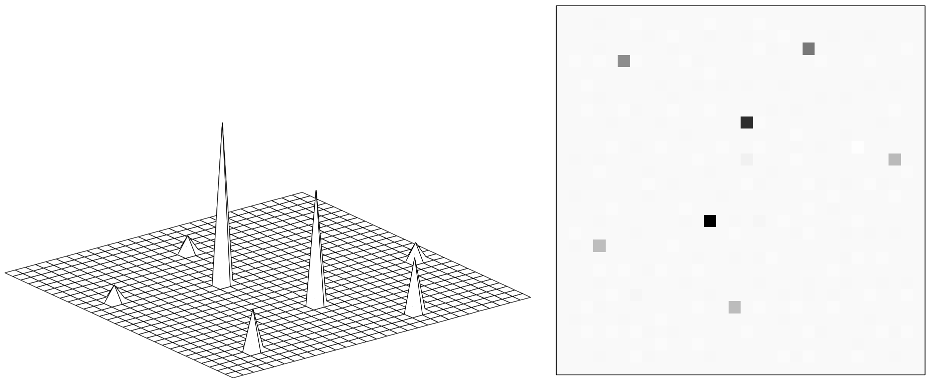
Original u_o



Data $v = u_o$+outliers

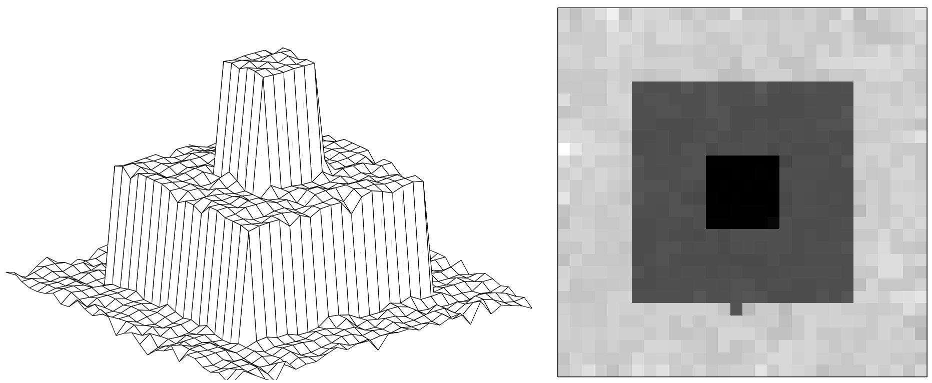


Restoration $\hat{u}$ for $\boldsymbol{\beta = 0.14}$

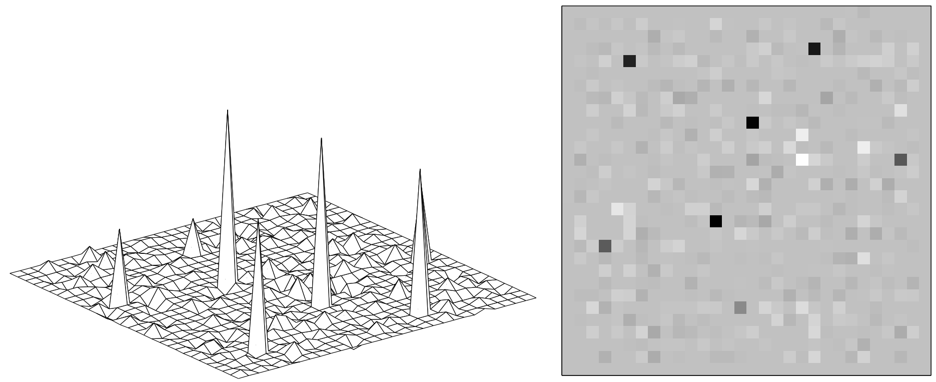


Residuals $v - \hat{u}$

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

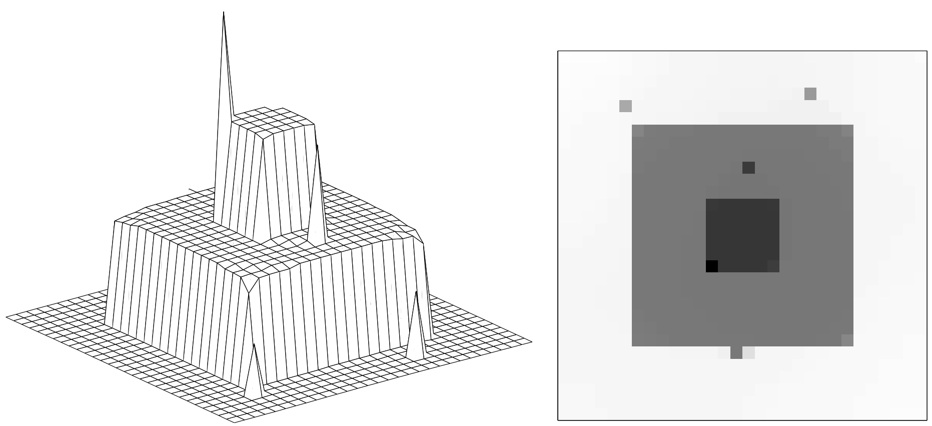


Restoration $\hat{u}$ for $\boldsymbol{\beta = 0.25}$

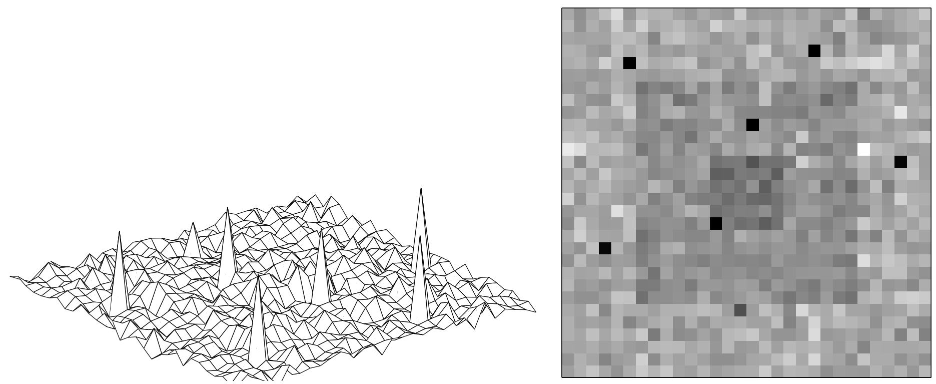


Residuals $v - \hat{u}$

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



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

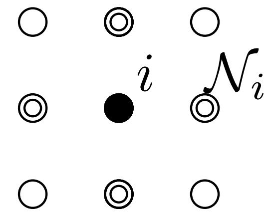


Residuals $v - \hat{u}$

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

$$v \in \mathbb{R}^p \quad \Rightarrow \quad \begin{array}{l} \hat{u} = \arg\min\limits_{u} \mathcal{F}_v(u) \\ \hat{h} = \{i : \hat{u}_i = v_i\} \end{array} \qquad \begin{cases} v_i \ \text{ is regular } \ \text{if } i \in \hat{h} \\ v_i \ \text{ is outlier } \ \text{if } i \in \hat{h}^c \end{cases}$$

Outlier detector: $v \rightarrow \hat{h}^c(v) = \{i : \hat{u}_i \neq v_i\}$

Smoothing: $\hat{u}_i$ for $i \in \hat{h}^c$ = estimate of the outlier

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

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

4.3 Recovery of frame coefficients using ℓ_1 data-fitting *[Durand &Nikolova07,08]*

- Data: $v = u_o +$ noise; if multiplicative noise: $v = \log(\text{data}) = \log(\text{image}) + \log(\text{noise})$

- Noisy frame coefficients $y = Wv = Wu_o +$ noise

- Hard thresholding can keep relevant information if T small $y_{T\,i} \doteq \begin{cases} 0 & \text{if } |y_i| \leq T \\ y_i & \text{if } |y_i| > T \end{cases}$

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

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

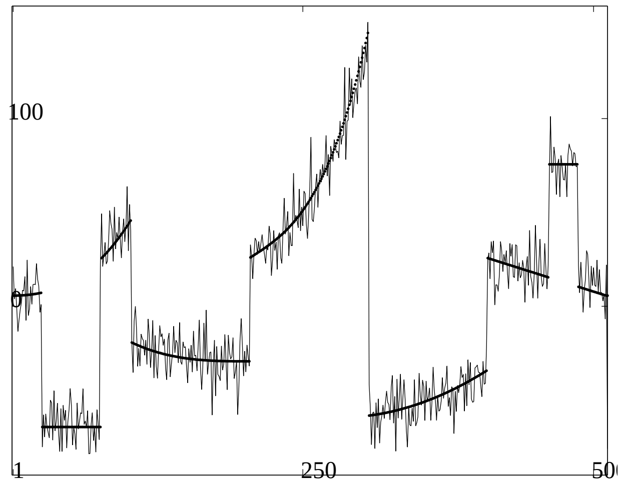
Our choice:

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

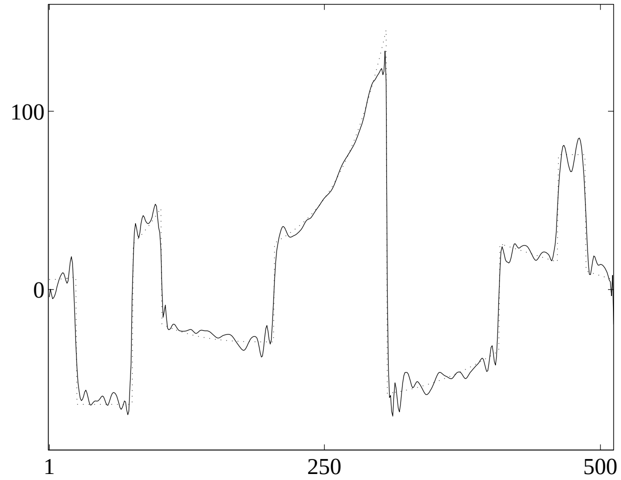
$$\hat{u} = W^* \hat{x}, \quad \text{where } W^* \text{ left inverse}$$

Rationale:

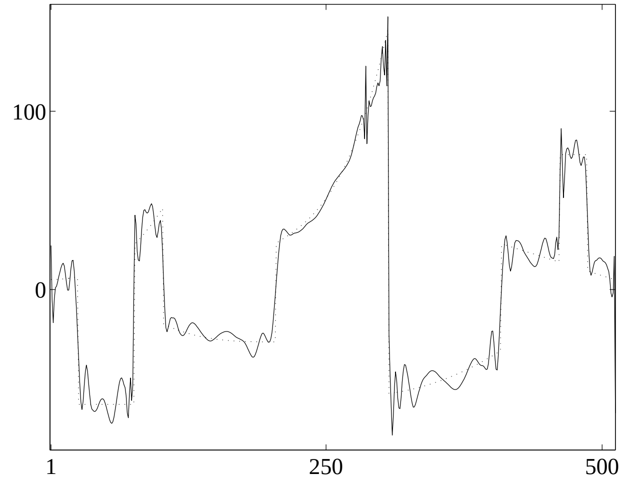
Keep $\hat{x}_i = y_{T\,i}$		Restore $\hat{x}_i \neq y_{T\,i}$		
significant coefs	$y_i \approx (Wu_o)_i$	outliers	$\|y_i\| \gg \|(Wu_o)_i\|$	(frame-shaped artifacts)
thresholded coefs if	$(Wu_o)_i \approx 0$	edge coefs	$\|(Wu_o)_i\| > \|y_{T\,i}\| = 0$	("Gibbs" oscillations)



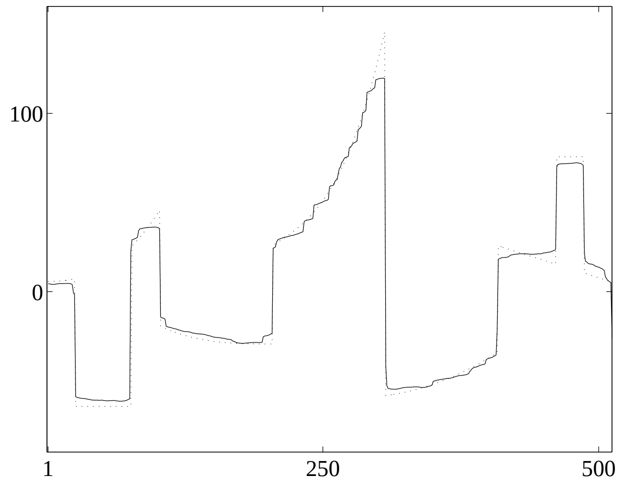
Original and data



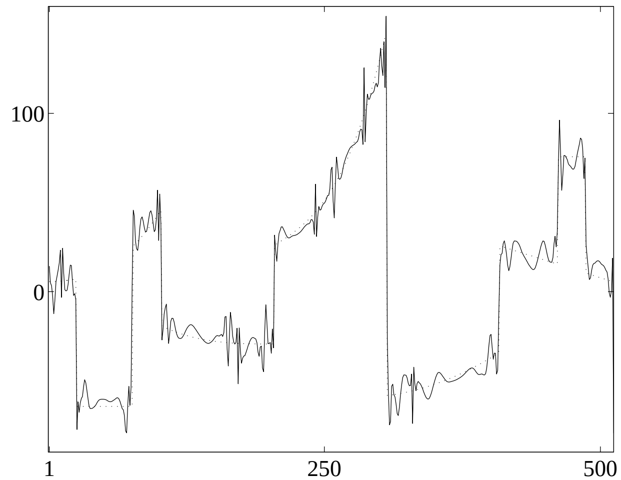
Sure-shrink method



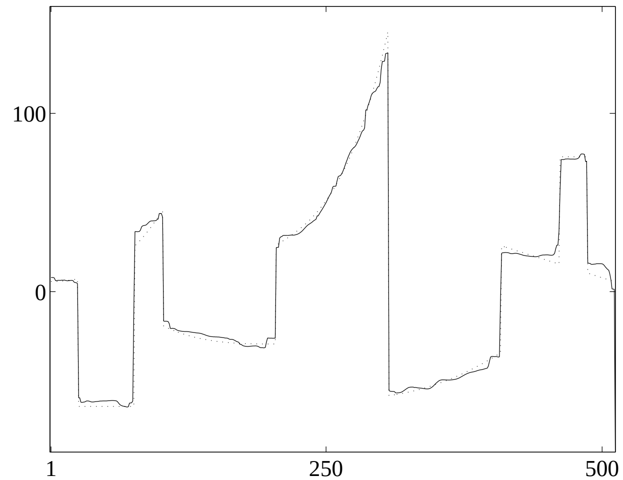
Hard thresholding



Total variation



Data y_T, $T = 23$ $\Longrightarrow$



The proposed method

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

Original | Additive noise | TV | Optimal T=100 | Our data T=50 | Our method

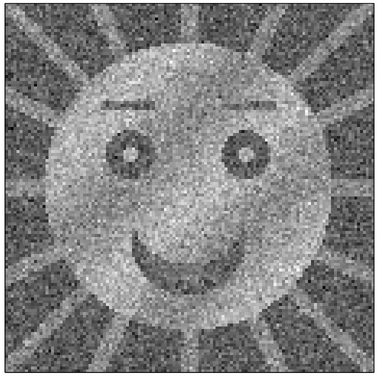
Multiplicative noise | Noise model+TV | Our+Daubechies8 | Our+Contourlets | Original

4.4 Fast Cleaning of Noisy Data From Impulse Noise *[Nikolova 03 and 08]*

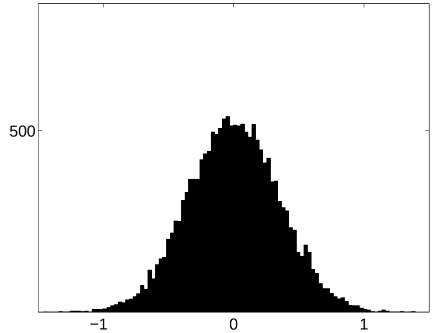
$$\text{minimize} \quad \mathcal{F}_v(u) = \|u - v\|_1 + \beta\|Gu\|^2$$

Very fast minimization scheme *(no line search)*

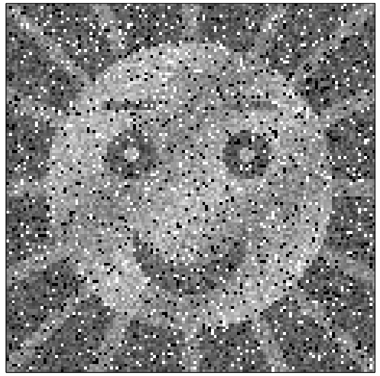
Semi-explicit expressions for the minimizer



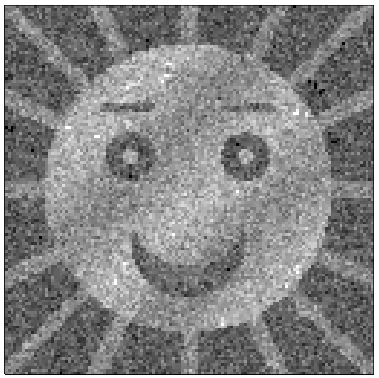
$u_o = u^* + n$



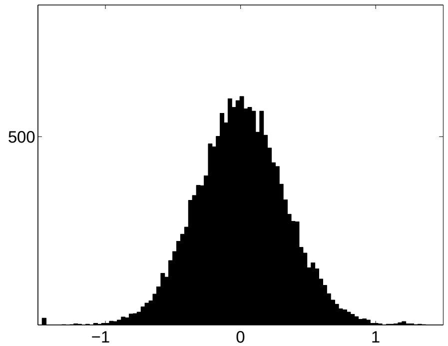
Hist(n)



$v = u_o + \omega$



Proposed method



Hist($\hat{u} - u_o$)

4.5 Fast 2-stage restoration under impulse noise *[Chan, Nikolova et al. 04, 05, 08]*

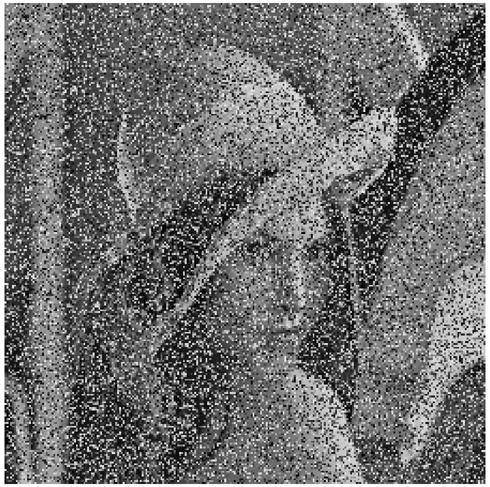
1. Approximate the outlier-detection stage by rank-order filter

   Corrupted pixels $\hat{h}^c = \{i : \hat{v}_i \neq v_i\}$ where $\hat{v}$=Rank-Order Filter (v)

   - Salt &Pepper (SP) noise e.g. by adaptive median
   - Random-Valued (RV) noise e.g. by center-weighted median

   $\Rightarrow$ improve speed and accuracy

2. Restore $\hat{u}$ (denoising, deblurring) using an edge-preserving variational method subject to $\langle a_i, \hat{u} \rangle = v_i$ for all $i \in \hat{h}$ $\Rightarrow$ Fast optimization, pertinent initialization

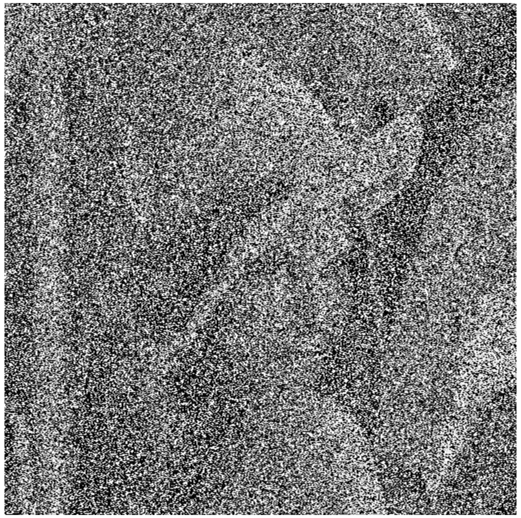


50% RV noise

ACWMF

DPVM

Our method



70 %SP noise(6.7dB)

MED (23.2 dB)

NASM (21.8 dB)

ISM filter (23.4 dB)

Adapt.med.(25.8dB)

Variational (24.6)

Our method(29.3dB)

Original Lena

5. Nonsmooth data-fidelity and regularization

Consequence of §3 and §4: if Φ and Ψ non-smooth, $\begin{cases} G_i \hat{u} = 0 & \text{for} \quad i \in \hat{h}_\varphi \neq \emptyset \\ \langle a_i, \hat{u} \rangle = v_i & \text{for} \quad i \in \hat{h}_\psi \neq \emptyset \end{cases}$

5.1 Binary images by L1 data-fitting and TV

[Chan, Esedoglu, Nikolova06]

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

$$\hat{\Sigma} = \arg\min_{\Sigma} \left\{ \left\| \mathbb{1}_\Sigma - \mathbb{1}_\Omega \right\|_2^2 + \beta \mathrm{TV}(\mathbb{1}_\Sigma) \right\} \qquad \text{nonconvex problem}$$

$$= \arg\min_{\Sigma} \left\{ \text{Surface}(\Sigma \,\Delta\, \Omega) + \beta \text{Per}(\Sigma) \right\} \quad \text{usual techniques (curve evolution, level-sets) fail}$$

(symmetric difference)

We reformulate the problem so that the desired solution minimizes a convex problem

$$\boxed{\mathcal{F}_v(u) = \|u - \mathbb{1}_\Omega\|_1 + \beta \mathrm{TV}(u) \}}$$

Then $\mathcal{F}_v(u)$ is solved for $\hat{u} = \mathbb{1}_{\hat{\Sigma}}$

$\Rightarrow$ **Convex algorithm for finding the global minimum**

5.2 One-step real-time dejittering of digital video

[Nikolova 08]

Image $\boldsymbol{u} \in \mathbb{R}^{r\times c}$, rows $\boldsymbol{u}_i$, pixels $\boldsymbol{u}_i(j)$

Data $\boldsymbol{v}_i(j) = \boldsymbol{u}_i(j + \boldsymbol{d}_i)$, $\boldsymbol{d}_i$ integer, $|\boldsymbol{d}_i| \le M$

Restore $\hat{\boldsymbol{u}} \equiv$ restore $\hat{\boldsymbol{d}}_i$, $1 \le i \le r$

Our approach: restore $\hat{\boldsymbol{u}}_i$ based on $(\hat{\boldsymbol{u}}_{i-1}, \hat{\boldsymbol{u}}_{i-2})$ by $\hat{\boldsymbol{d}}_i = \arg\min_{|\boldsymbol{d}_i|\le N} \mathcal{F}(\boldsymbol{d}_i)$ where

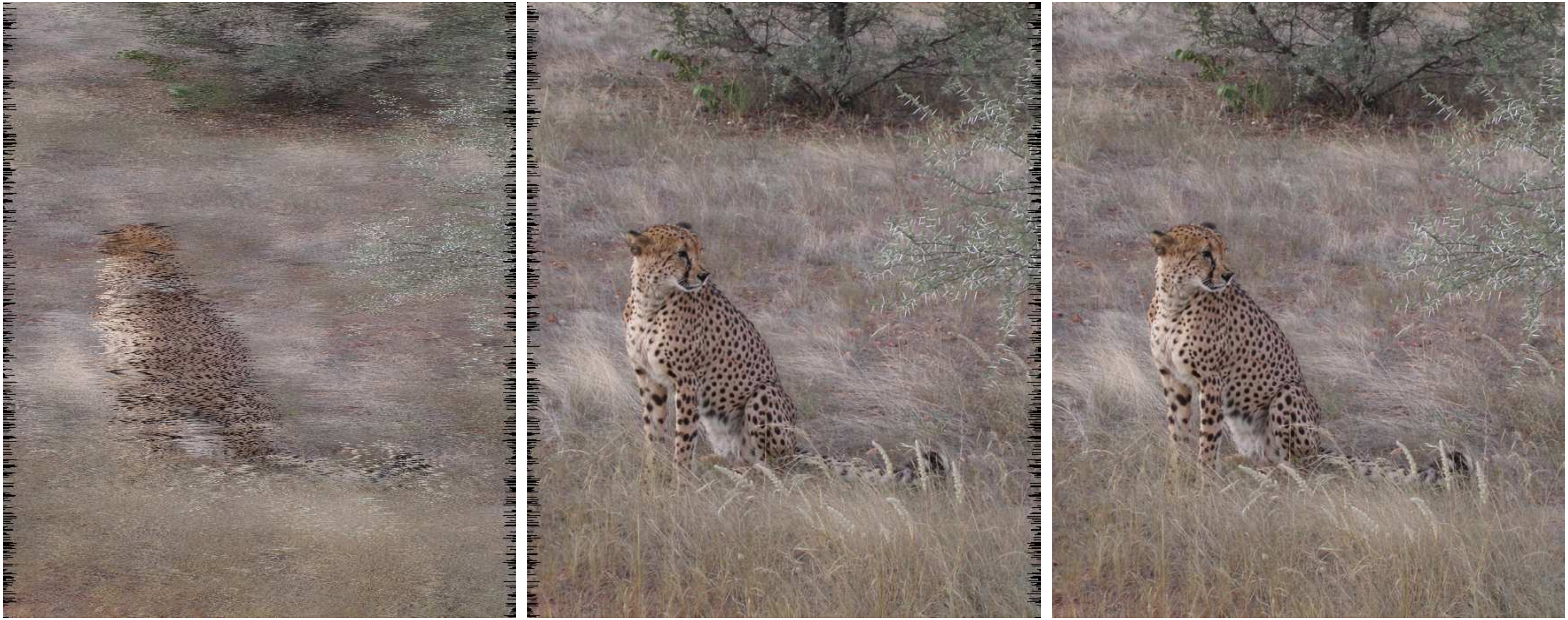
$$\mathcal{F}(\boldsymbol{d}_i) = \sum_{j=N+1}^{c-N} \left|\boldsymbol{v}_i(j+\boldsymbol{d}_i) - 2\hat{\boldsymbol{u}}_{i-1}(j) + \hat{\boldsymbol{u}}_{i-2}(j)\right|^{\alpha}, \quad \alpha \in (0.5, 1], \quad N > M$$

piece-wise linear model for the columns of the image

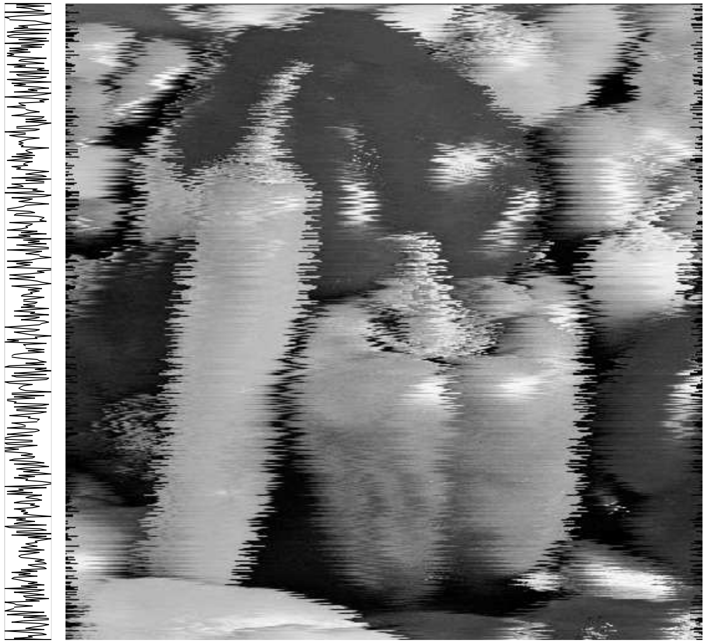
Jittered, $[-20, 20]$ | $\alpha = 1$ | Jitter: $6\sin\left(\frac{n}{4}\right)$ | $\boldsymbol{\alpha = 1} \equiv$ Original

Jittered $\{-8,\ldots,8\}$ Original image $\alpha = 1$ Zooms

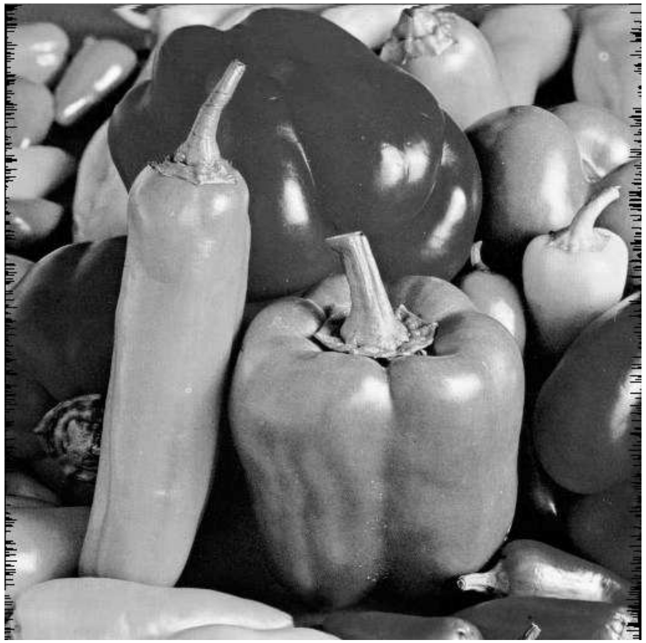
(512×512)Jitter$M=6$ $\boldsymbol{\alpha}\in\{\mathbf{1},\,\frac{\mathbf{1}}{\mathbf{2}}\}$=Original Lena (256×256) Jitter $\{-6,..,6\}$ $\boldsymbol{\alpha}\in\{\mathbf{1},\,\frac{\mathbf{1}}{\mathbf{2}}\}$



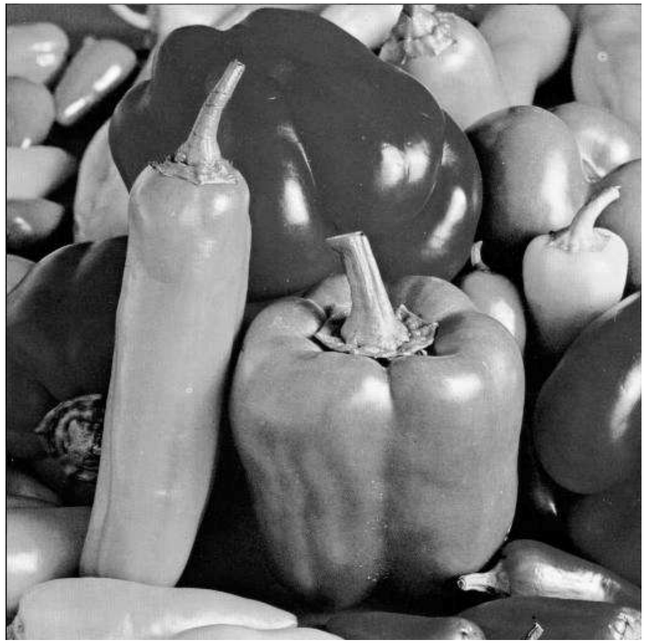
Jitter {-15,..,15} $\alpha = 1,\ \alpha = 0.5$ Original image



Jitter {-10,..,10}



$\alpha = 0.5$



Original image

Comparison with Smooth Energies *[Nikolova 04]*

We show that if $\mathcal{F}_v(u) = \Psi(u,v) + \beta\Phi(u)$ is **smooth** $\Rightarrow$ **for a.e.** $v \in \mathbb{R}^q$

$$G_i\hat{u} \neq 0, \ \forall i \quad \text{and} \quad \langle a_i, \hat{u}\rangle \neq v_i, \ \forall i$$

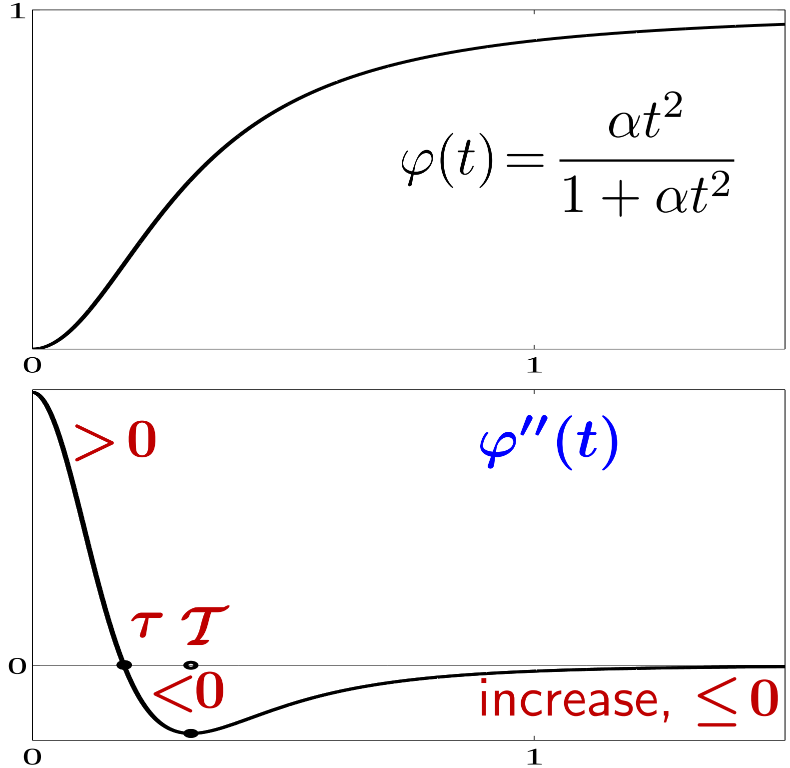
7 Non-convex regularization

[Nikolova 04]

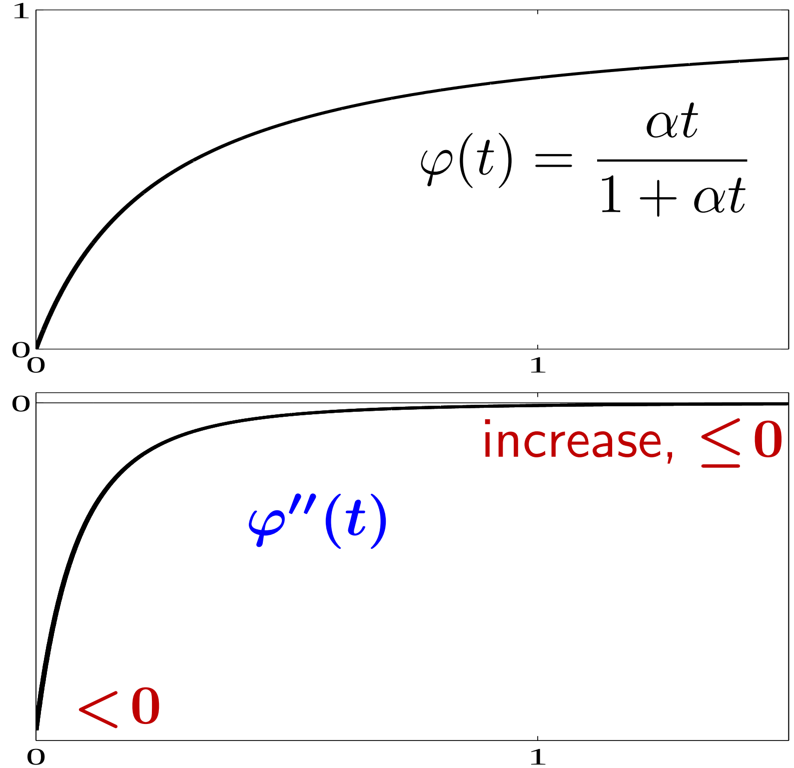
$$\mathcal{F}_v(u) = \|Au - v\|^2 + \beta \sum_{i=1}^{r} \varphi(\|G_i u\|)$$

Standard assumptions on φ: $\mathcal{C}^2$ on $\mathbb{R}_+$ and $\lim_{t\to\infty} \varphi''(t) = 0$, as well as:

$\varphi'(0) = 0$ (Φ is smooth)



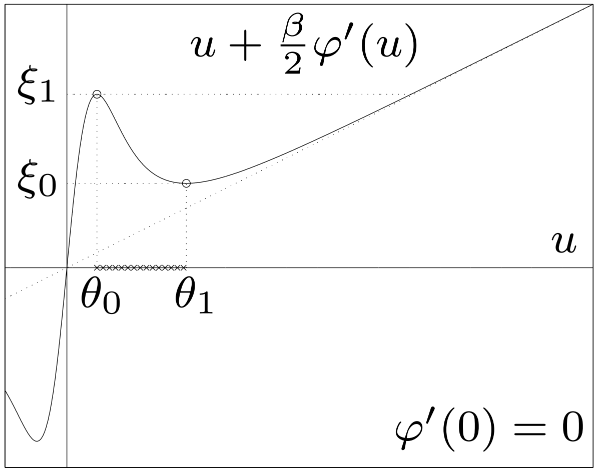
$\varphi'(0) > 0$ (Φ is nonsmooth)

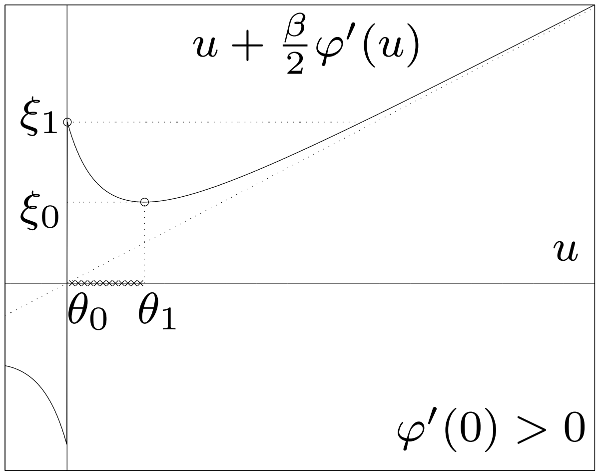


6.1 Either shrinkage or enhancement of differences

Illustration on $\mathbb{R}$

$$\mathcal{F}_v(u) = (u - v)^2 + \beta\varphi(|u|), \quad u, v \in \mathbb{R}$$





No local minimizer in (θ_0, θ_1)

$\exists\ \xi_0 > 0, \quad \exists\ \xi_1 > \xi_0$

$|v| \leq \xi_1 \Rightarrow |\hat{u}_0| \leq \theta_0$

strong smoothing

$|v| \geq \xi_0 \Rightarrow |\hat{u}_1| \geq \theta_1$

loose smoothing

$\exists\ \xi \in (\xi_0, \xi_1)$ $\quad |v| \leq \xi \Rightarrow$ global minimizer $= \hat{u}_0$ (strong smoothing)

$\quad |v| \geq \xi \Rightarrow$ global minimizer $= \hat{u}_1$ (loose smoothing)

For $v = \xi$ the global minimizer jumps from $\hat{u}_0$ to $\hat{u}_1$ $\equiv$ decision for an "edge"

Since [Geman^{2}1984] various nonconvex Φ to produce minimizers with smooth regions and sharp edges

Sketch : if $\hat{\boldsymbol{u}}$ is a (local) minimizer of $\boldsymbol{\mathcal{F}_v}$ then $\exists \boldsymbol{\theta_0} \geq \mathbf{0}$, $\exists \boldsymbol{\theta_1} > \boldsymbol{\theta_0}$ such that

$\widehat{\boldsymbol{h}}_0 = \{\boldsymbol{i} : \|\boldsymbol{G_i}\hat{\boldsymbol{u}}\| \leq \boldsymbol{\theta_0}\}$ $\widehat{\boldsymbol{h}}_1 = \{\boldsymbol{i} : \|\boldsymbol{G_i}\hat{\boldsymbol{u}}\| \geq \boldsymbol{\theta_1}\}$ with $\widehat{\boldsymbol{h}}_0 \cup \widehat{\boldsymbol{h}}_1 = \{\mathbf{1}, \ldots, \boldsymbol{r}\}$

homogeneous regions **edges** $G_i : i \in \{1, \ldots, r\}$

(A) φ nonconvex and $\varphi'(0) = 0$, GG^* invertible, $\beta > K(A, G, \varphi)$ then $\exists \theta_0 \in (\tau, \mathcal{T})$ and $\exists \theta_1 > \mathcal{T} > \theta_0$ such that either $\|G_i\hat{u}\| \leq \theta_0$ or $\|G_i\hat{u}\| \geq \theta_1$, $\forall i$

(B) $\varphi(t) = \min\{\alpha t^2, 1\}$ and $\mathcal{F}_v$ has a global minimum at $\hat{u}$ then $\exists\ \Gamma_i \in (0, 1)$ so that

either $\left|\hat{u}_{i+1} - \hat{u}_i\right| \leq \frac{\Gamma_i}{\sqrt{\alpha}}$ or $\left|\hat{u}_{i+1} - \hat{u}_i\right| \geq \frac{1}{\sqrt{\alpha}\Gamma_i}$

(C) $\boldsymbol{\varphi'(0) > 0}$ (Φ nonsmooth and nonconvex), $\beta > K(A, G, \varphi)$, then $\boldsymbol{\exists \theta_1 > 0}$ such that

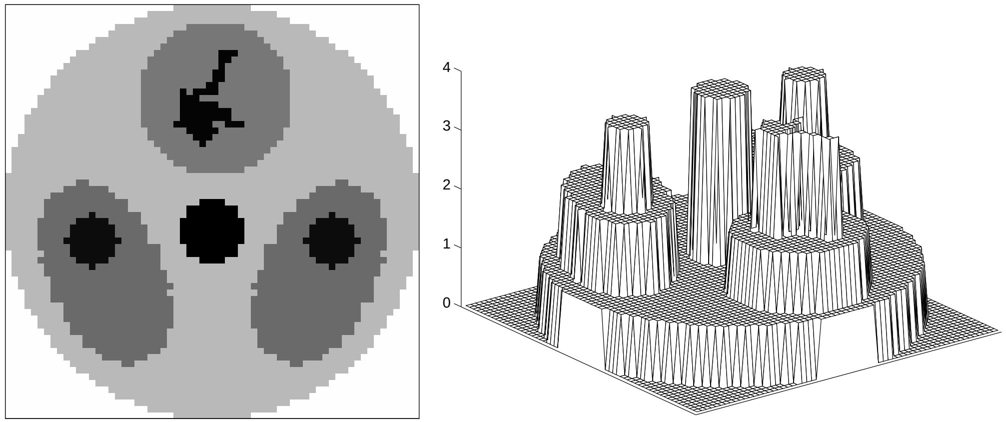
either $\|\boldsymbol{G_i}\hat{\boldsymbol{u}}\| = \mathbf{0}$ or $\|\boldsymbol{G_i}\hat{\boldsymbol{u}}\| \geq \boldsymbol{\theta_1}$, $\forall \boldsymbol{i}$ $\Rightarrow$ fully segmented image, high sparsity

(D) $(\boldsymbol{\ell_0})$: $\varphi(0) = 0$, $\varphi(t) = 1, t \neq 0$ and $\mathcal{F}_v$ has a global minimum at $\hat{u}$ then $\exists\ \Gamma_i$:

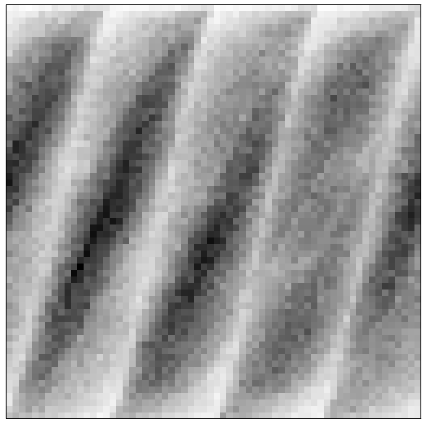
either $\hat{u}_{i+1} = \hat{u}_i$ or $\left|\hat{u}_{i+1} - \hat{u}_i\right| \geq \frac{\sqrt{\beta}}{\Gamma_i}$

Explicit formula for Γ_i, bounds for θ_0 and θ_1

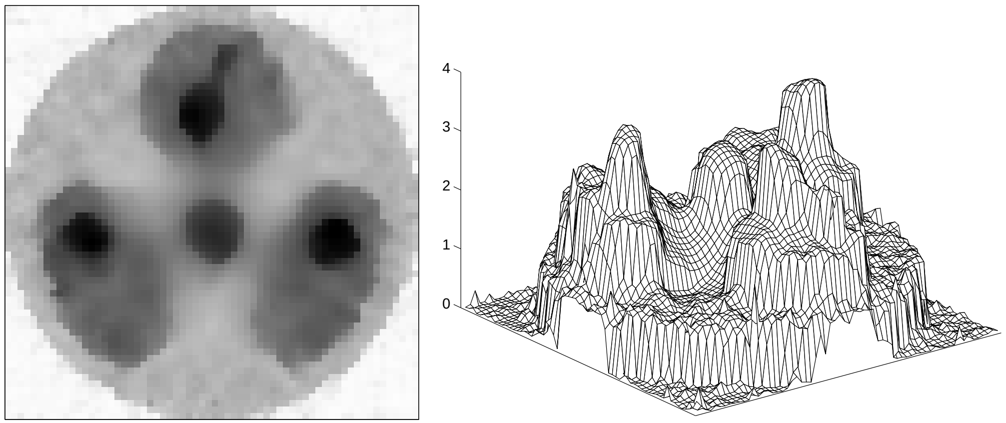
Image Reconstruction in Emission Tomography



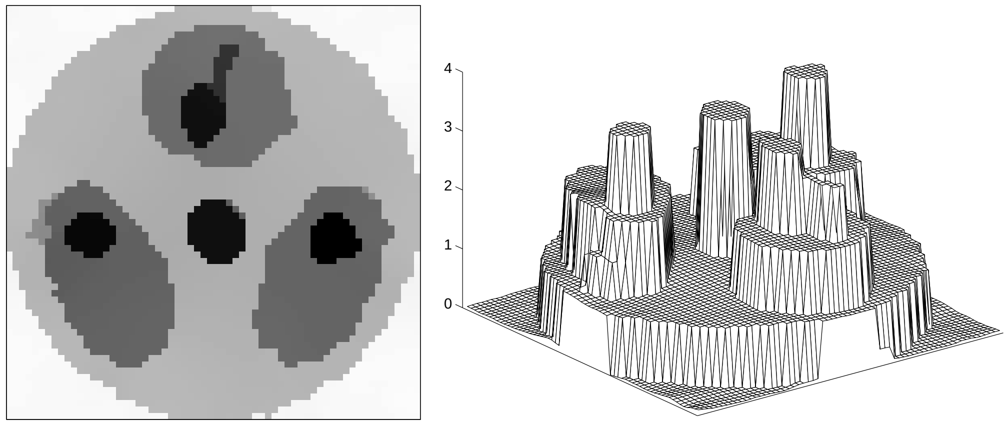
Original phantom



Emission tomography simulated data



φ is smooth (Huber function)



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

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

6.2 Selection for the global minimizer

Additional assumptions: $\|\varphi\|_\infty < \infty$, $\{G_i\}$—1^{st}-order differences, A^*A invertible

$$\mathbb{1}_{\Sigma i} = \begin{cases} \mathbf{1} & \text{if } \boldsymbol{i} \in \boldsymbol{\Sigma} \subset \{\mathbf{1}, .., \boldsymbol{p}\} \\ \mathbf{0} & \text{else} \end{cases}$$

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

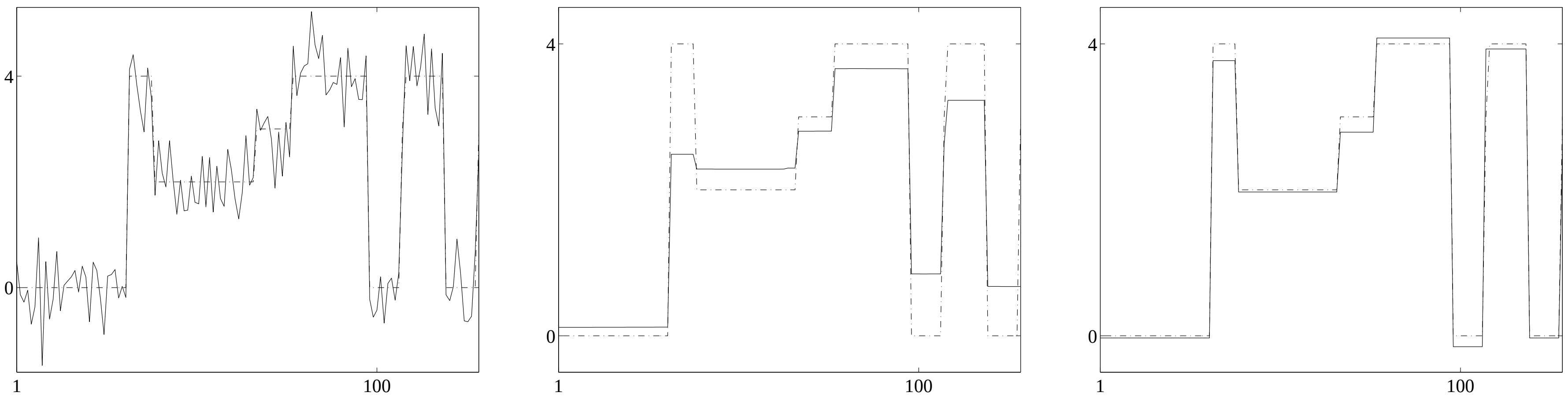
Data: $\boldsymbol{v} = \boldsymbol{\xi}\ \boldsymbol{A}\ \mathbb{1}_{\boldsymbol{\Sigma}} = \boldsymbol{A u_o}$

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

In each case we exhibit $\exists\ \boldsymbol{\xi_0} > \mathbf{0}$, $\exists\ \boldsymbol{\xi_1} > \boldsymbol{\xi_0}$:

- *Case (A) - Φ smooth:* $\xi < \xi_0 \Rightarrow \hat{\boldsymbol{u}}$—smooth, $\xi > \xi_1 \Rightarrow \hat{\boldsymbol{u}}$—perfect edges
- $\varphi(t) = \min\{\alpha t^2,\ 1\}$ $\begin{cases} \xi < \xi_0 \ \Rightarrow\ \hat{\boldsymbol{u}} = \text{regularized least-squares (no edges)} \\ \xi > \xi_1 \ \Rightarrow\ \hat{\boldsymbol{u}} = \boldsymbol{u_o} \end{cases}$
- *Case (C) - Φ nonsmooth:* $\begin{cases} \xi < \xi_0 \ \Rightarrow\ \hat{\boldsymbol{u}} = \text{constant} \\ \xi > \xi_1 \ \Rightarrow\ \hat{\boldsymbol{u}} = \boldsymbol{c}\ \boldsymbol{u_o},\quad c < 1,\ \lim_{\xi\to\infty} c = 1 \text{ if } \Sigma \text{ connected} \end{cases}$
- $(\boldsymbol{\ell_0})$: $\varphi(0)=0,\ \varphi(t)=1, t\neq 0$: $\begin{cases} \xi < \xi_0 \ \Rightarrow\ \hat{\boldsymbol{u}} = \text{constant} \\ \xi > \xi_1 \ \Rightarrow\ \hat{\boldsymbol{u}} = \boldsymbol{u_o} \end{cases}$

6.3 Comparison with Convex Edge-Preserving Regularization



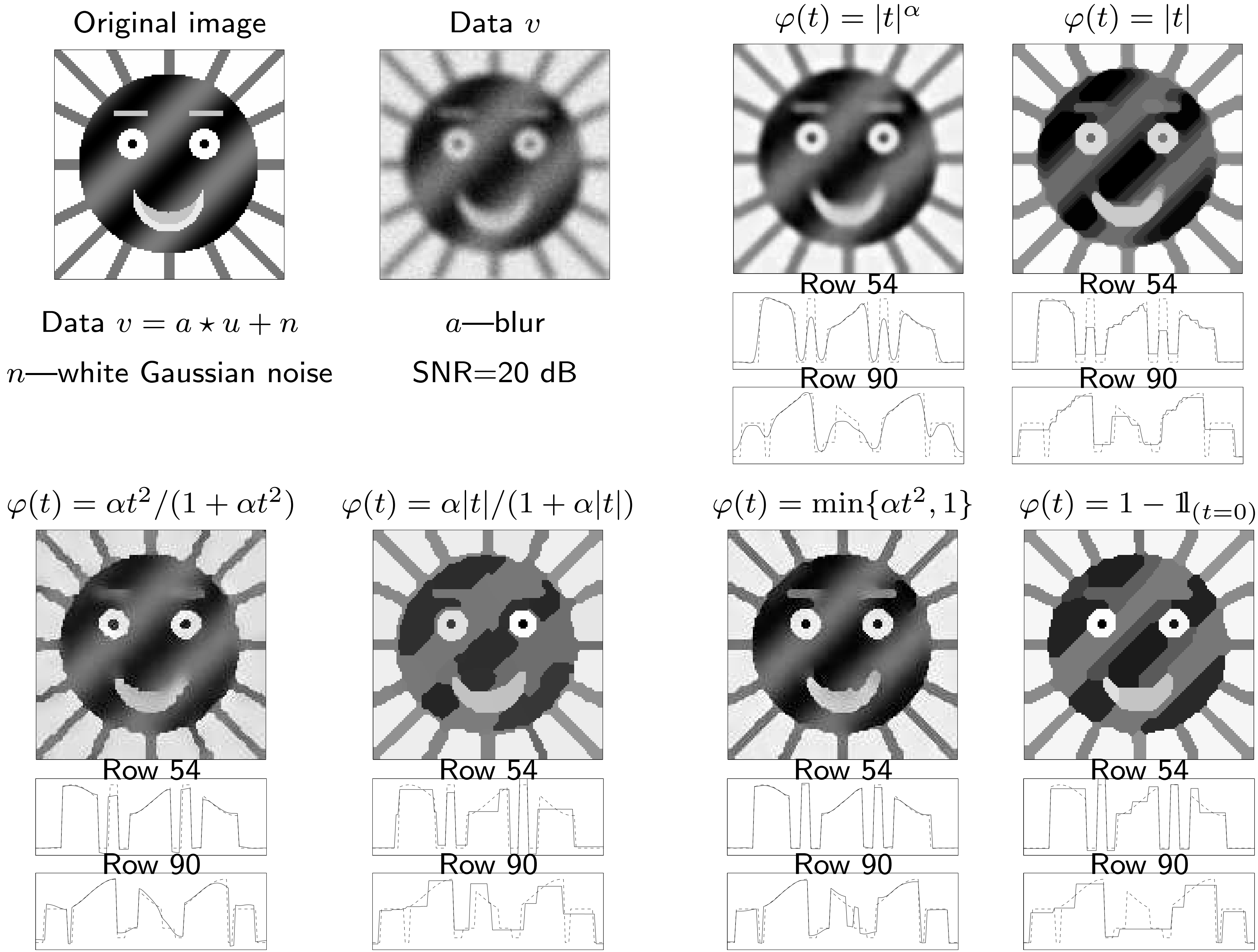
Data $v = u_o + n$ $\qquad$ $\varphi(t) = |t|$ $\qquad$ $\varphi(t) = \alpha|t|/(1 + \alpha|t|)$

- If $\mathcal{F}_v$ is convex, then $\|G_i\hat{u}\|$ can take any value on $\mathbb{R}$

  TV (non-smooth) creates constant zones, a fortiori these are separated by edges

- Edge-recovery using a non-convex φ is fundamentally different: it relies on the competition between different local minima corresponding to different edge configurations. At a v where the global minimum jumps from one local minimum to another, $v \to \mathcal{U}(v)$ is discontinuous.

Illustration of all properties


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

7 Models are distorted by Bayesian MAP

[Nikolova 07]

7.1. MAP estimators to combine noisy data and prior

Bayesian approach: U, V random variables, events $U = u$, $V = v$

Likelihood $f_{V|U}(v|u)$, Prior $f_U(u) \propto \exp\{-\lambda\Phi(u)\}$, Posterior $f_{U|V}(u|v) = f_{V|U}(v|u) f_U(u) \frac{1}{Z}$

MAP $\hat{u}$ = the most likely solution given the recorded data $V = v$:

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

MAP is the most frequent way to combine models on data-acquisition and priors

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

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

In practice one needs that:

$$\begin{cases} U \sim f_U \\ AU - V \sim f_N \end{cases} \Rightarrow \begin{cases} f_{\hat{U}} \approx f_U \\ f_{\hat{N}} \approx f_N, \quad \hat{N} \approx A\hat{U} - V \end{cases}$$

Our analytical results show that both models ($f_{V|U}$ and f_U) are deformed in a MAP estimate

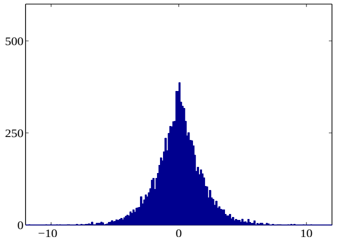
7.2 Example: MAP shrinkage *[Simoncelli99, Belge-Kilmer00, Moulin-Liu00, Antoniadis02]*

- Noisy wavelet coefficients $y = Wv = Wu_o + n$, $n \sim \mathcal{N}(0, \sigma^2 I)$, True coefficients $x = Wu_o$
- Prior: x_i are i.i.d., $\sim f(x_i) = \frac{1}{Z} e^{-\lambda |x_i|^\alpha}$ (Generalized Gaussian, GG), $f_X(x) = \prod_i f(x_i)$

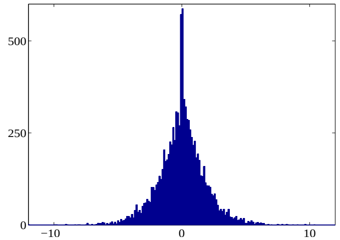
  Experiments have shown that $\alpha \in (0, 1)$ for many real-world images
- MAP restoration $\Leftrightarrow$ $\hat{x}_i = \arg\min_{t \in \mathbb{R}} \left((t - y_i)^2 + \lambda_i |t|^\alpha \right), \quad \forall i$

$(\alpha, \lambda, \sigma)$ fixed—10 000 independent trials:

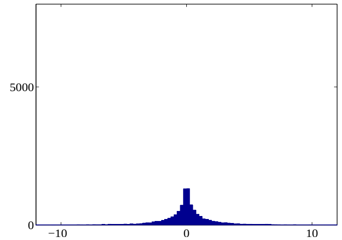
(1) sample $x \sim f_X$ and $n \sim \mathcal{N}(0, \sigma^2)$, (2) form $y = x + n$, (3) compute the **true** MAP $\hat{x}$



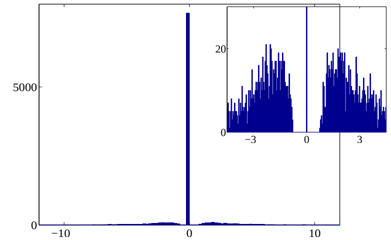
GG,$\alpha = 1.2, \lambda = \frac{1}{2}$



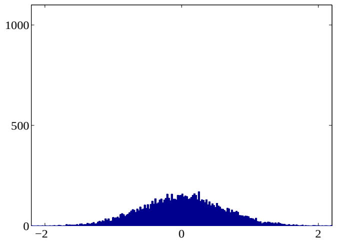
The true MAP $\hat{x}$



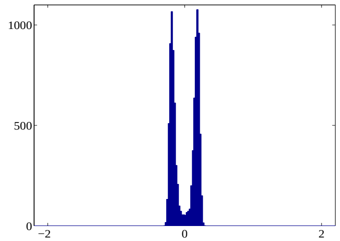
GG, $\alpha = \frac{1}{2}$, $\lambda = 2$



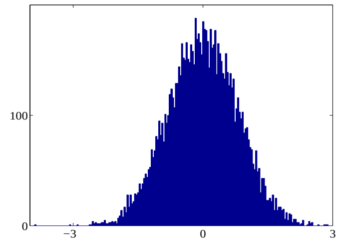
True MAP $\hat{x}$



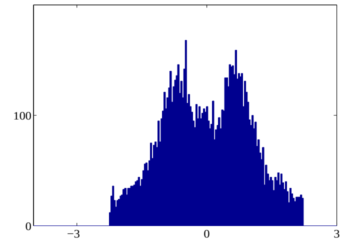
Noise $\mathcal{N}(0, \sigma^2)$



Noise estimate



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



Noise estimate

7.3. Theoretical explanations

$$\boldsymbol{V} = \boldsymbol{AU} + \boldsymbol{N} \text{ and } \boldsymbol{f_{U|V}} \textbf{ continuous } \Rightarrow \begin{cases} \Pr(G_i u = 0) = 0, \quad \forall i \\ \Pr(\langle a_i u = v_i) = 0, \quad \forall i \\ \Pr(\theta_0 < \|G_i u\| < \theta_1) > 0, \ \theta_0 < \theta_1, \ \forall i \end{cases}$$

The analytical results on $\hat{\boldsymbol{u}} = \arg\min_{\boldsymbol{u}} \mathcal{F}_{\boldsymbol{v}}(\boldsymbol{u})$ =MAP *(sections 3, 4 and 7)* yield:

- $\boldsymbol{f_U}$ non-smooth at 0 $\Leftrightarrow$ $\boldsymbol{\varphi'(0) > 0}$

  $$\mathsf{v} \in \mathcal{O}_{\hat{\mathsf{h}}} \Rightarrow \left[\mathsf{G_i}\hat{\mathsf{u}} = 0, \forall \mathsf{i} \in \hat{\mathsf{h}}\right] \Rightarrow \Pr(\mathsf{G_i}\hat{\mathsf{u}} = 0, \forall \mathsf{i} \in \hat{\mathsf{h}}) \geq \Pr(\mathsf{v} \in \mathcal{O}_{\hat{\mathsf{h}}}) = \int_{\mathcal{O}_{\hat{\mathsf{h}}}} \mathsf{f_V}(\mathsf{v})\mathsf{dv} > \mathbf{0}$$

  The **effective** *prior model corresponds to images and signals such that $G_i\hat{u} = 0$ for a certain number of indexes i. If $\{G_i\}$=first-order, it amounts to locally constant images and signals.*

- $\boldsymbol{f_N}$ nonsmooth at 0 $\Leftrightarrow$ $\boldsymbol{\psi'(0) > 0}$

  $$v \in \mathcal{O}_{\hat{h}} \Rightarrow \left[\langle a_i, \hat{u}\rangle = v_i \quad \forall i \in \hat{h}\right] \Rightarrow \Pr\left(\langle a_i, \hat{u}\rangle = v_i, \forall i \in \hat{h}\right) \geq \Pr(V \in \mathcal{O}_{\hat{h}}) > 0$$

  For all $i \in \hat{h}$, the prior has no influence on the solution and the noise remains intact.
  The **Effective** *model corresponds to impulse noise on the data.*

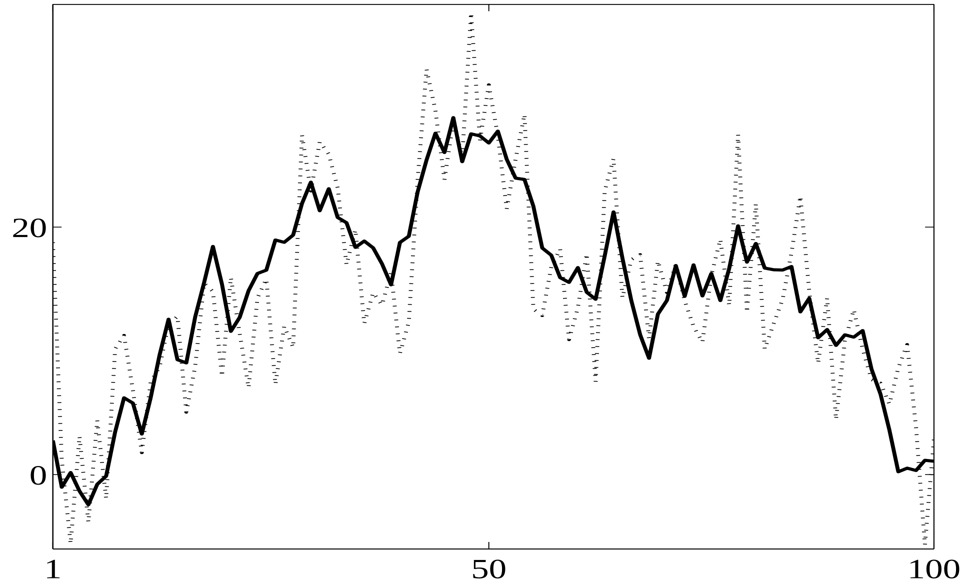
- $\boldsymbol{-\ln f_U}$ nonconvex $\Leftrightarrow$ $\boldsymbol{\varphi}$ nonconvex $\Rightarrow$ $\Pr\left(\theta_0 < \|\mathsf{G_i}\hat{\mathsf{U}}\| < \theta_1\right) = 0, \ \forall \mathsf{i}$

  Effective *prior: differences are either smaller than θ_0 or larger than θ_1.*

- $-\ln f_U$ nonconvex, nonsmooth at 0 $\Rightarrow \Pr\left(\|G_i\hat{u}\|=0\right) > 0,\ \Pr\left(0<\|G_i\hat{u}\|<\theta_1\right)=0$

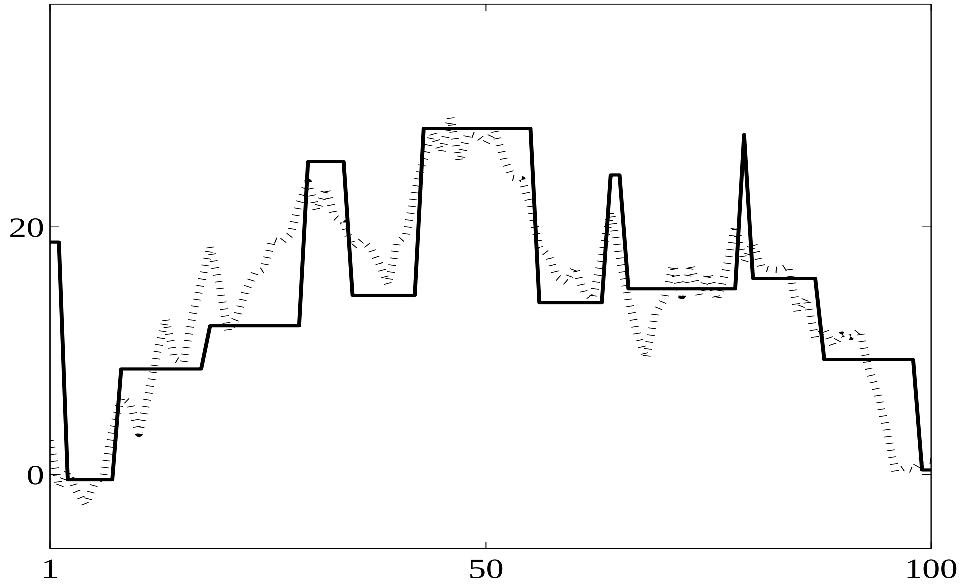
  If $\{G_i\}$—first-order—**effective** prior for signals and images composed of constant zones separated by edges higher than θ_1.

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



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 $(\cdots)$.

Knowing the true distributions, with the true parameters, is not enough.

Combining models remains an open problem

8. Open questions

- Analyzing the properties of the minimizers in connection with the shape of the energy yields strong results
- Such results provide a powerful tool for rigorous modeling
- Specialized energies can be conceived
- Minimization methods can take advantage from the known features of solutions
- **Goal : conceive solutions that match pertinent models**
- **Open field for research**

  What "features" and what "properties" ?

  Quantify features involving randomness

Papers available on http://www.cmla.ens-cachan.fr/~nikolova/