IPAM Graduate Summer School: Computer Vision, July 2013

Inverse modelling using optimization

to solve imaging tasks

Part II

Mila Nikolova

ENS Cachan, CNRS, France

nikolova@cmla.ens-cachan.fr

Stability of the minimizers of $\mathcal{F}_v$

$$\boxed{\begin{aligned} \mathcal{F}_v(u) &= \|Au - v\|_2^2 + \beta\Phi(u) \\ \Phi(u) &= \sum_i \varphi(\|G_i u\|_2) \end{aligned}}$$

$u \in \mathbb{R}^p$

$v \in \mathbb{R}^q$

$$\begin{cases} \varphi : \mathbb{R}_+ \to \mathbb{R} \\ \varphi \text{ incresing, continuous} \\ \varphi(t) > \varphi(0), \quad \forall t > 0 \end{cases}$$

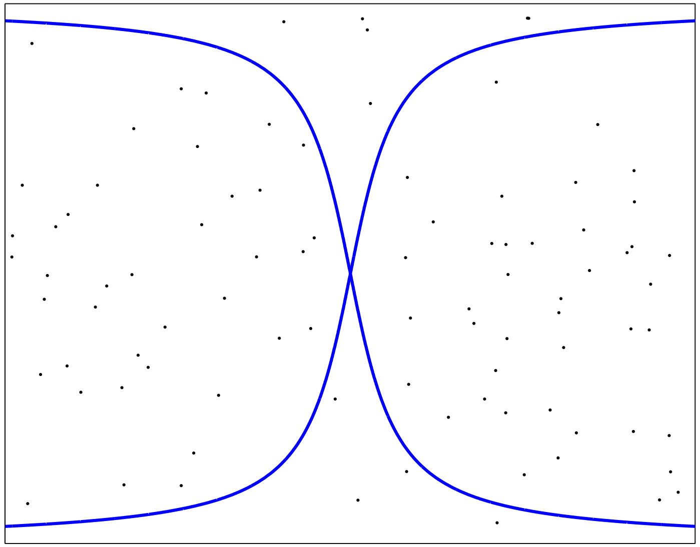
$\{G_i\}$ linear operators $\mathbb{R}^p \to \mathbb{R}^s$, $s \geqslant 1$

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

Systematically: $\ker A \cap \ker G = \{0\}$ $\qquad G \doteq \begin{bmatrix} G_1 \\ G_2 \\ \cdots \end{bmatrix}$

Question **2** Why?

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



- $N = \{(s,t) \ : \ t = \pm \arctan(s)\}$
- N is closed in $\mathbb{R}^2$ and its Lebesgue measure in $\mathbb{R}^2$ is $\mathbb{L}^2(N) = 0$
- $(x,y) =$ random $\mathbb{R}^2$

Question **3** What is the chance that $(x,y) \in N$?

[Durand & Nikolova 06]

Assumptions: φ is piecewise $\mathcal{C}^{m\geqslant 2}$, edge-preserving, possibly non-convex, $\mathrm{rank}(A) = p$

- There is a closed $\boldsymbol{N} \subset \mathbb{R}^q$ with $\mathbb{L}^{\boldsymbol{q}}(\boldsymbol{N}) = \mathbf{0}$ such that $\forall \boldsymbol{v} \in \mathbb{R}^{\boldsymbol{q}} \setminus \boldsymbol{N}$, every (local) minimizer $\hat{\boldsymbol{u}}$ of $\boldsymbol{\mathcal{F}_v}$ is given by $\hat{\boldsymbol{u}} = \boldsymbol{\mathcal{U}}(\boldsymbol{v})$ where $\boldsymbol{\mathcal{U}}$ is a $\mathcal{C}^{m-1}$ (local) minimizer function.

Question **4** Why knowledge on local minimizers is important?

Question **5** Compare $\hat{u}$ and $\mathcal{U}(v+\varepsilon)$ where $\varepsilon \in \mathbb{R}^q$ is small enough.

- $\exists\ \hat{\boldsymbol{N}} \subset \mathbb{R}^q$ with $\mathbb{L}^{\boldsymbol{q}}(\hat{\boldsymbol{N}}) = \mathbf{0}$ such that $\forall \boldsymbol{v} \in \mathbb{R}^{\boldsymbol{q}} \setminus \hat{\boldsymbol{N}}$, $\boldsymbol{\mathcal{F}_v}$ has a unique global minimizer

Question **6** What can happen if $v \in \hat{N}$?

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

Question **7** If $\mathcal{F}_v$ is strictly convex, determine N and $\hat{N}$.

Nonasymptotic bounds on minimizers

[Nikolova 07]

Assumption: φ is piecewise $\mathcal{C}^1$

- φ is strictly increasing **or** $\operatorname{rank}(A) = p$

$$\hat{\boldsymbol{u}} \text{ is a (local) minimizer of } \boldsymbol{\mathcal{F}_v} \quad \Rightarrow \quad \|\boldsymbol{A\hat{u}}\| \leqslant \|\boldsymbol{v}\|$$

- $\|\boldsymbol{\varphi}'\|_\infty < \infty$ (φ is edge-preserving) and $\operatorname{rank}(A) = q \leqslant p$

$$\hat{\boldsymbol{u}} \text{ is a (local) minimizer of } \boldsymbol{\mathcal{F}_v} \quad \Rightarrow \quad \|\boldsymbol{v} - \boldsymbol{A\hat{u}}\|_\infty \leqslant \frac{\boldsymbol{\beta}}{\mathbf{2}} \|\boldsymbol{\varphi}'\|_\infty \, \|(\boldsymbol{AA}^*)^{-1}\boldsymbol{A}\|_\infty \, \|\boldsymbol{G}\|_1$$

$\|\boldsymbol{\varphi}'\|_\infty = \mathbf{1}$ and G – 1^{st} order differences: $\begin{cases} \text{signal} & \Rightarrow \quad \|\boldsymbol{v} - \hat{\boldsymbol{u}}\|_\infty \leqslant \boldsymbol{\beta} \\ \text{image} & \Rightarrow \quad \|\boldsymbol{v} - \hat{\boldsymbol{u}}\|_\infty \leqslant \mathbf{2}\boldsymbol{\beta} \end{cases}$

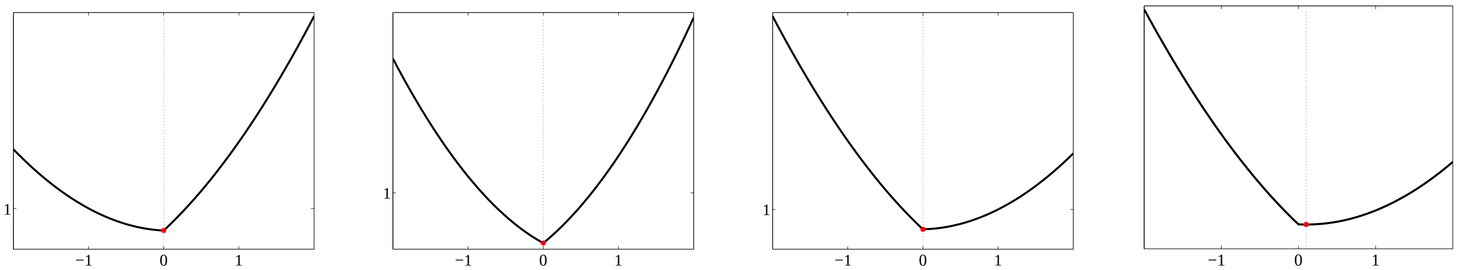
Question **8** If $v = u_o + n$ for n Gaussian noise, is it possible to clean v from this noise by minimizing $\mathcal{F}_v$?

Non-Smooth Energies, Side Derivatives, Subdifferential

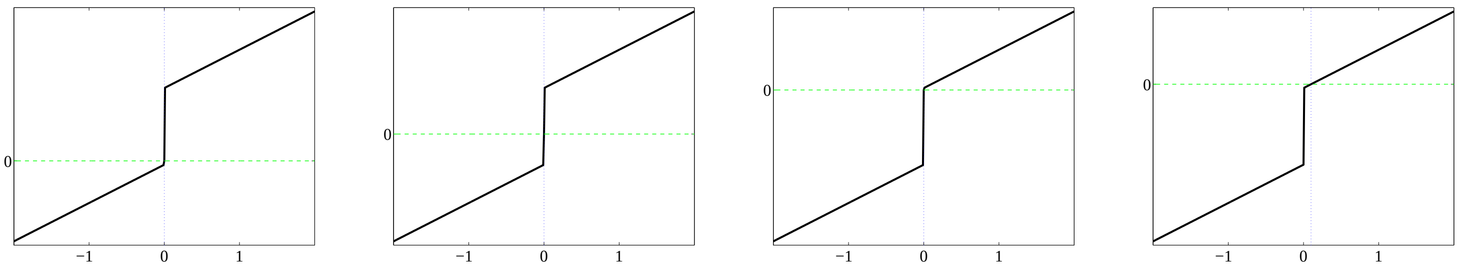
Rademacher's theorem: If $\mathcal{F}_v : \mathbb{R}^p \to \mathbb{R}$ is Lipschitz continuous, then $\mathcal{F}_v$ is differentiable (in the usual sense) almost everywhere in $\mathbb{R}^p$.

A kink is a point u where $\nabla \mathcal{F}_v(u)$ is not defined (in the usual sense).

Example: $\mathcal{F}_v(u) = \dfrac{1}{2}(u-v)^2 + \beta|u|$ for $\beta = 1 > 0$ and $u, v \in \mathbb{R}$



$v = -0.9$ $\quad$ $v = -0.2$ $\quad$ $v = 0.95$ $\quad$ $v = 1.1$



$$\hat{u} = \begin{cases} v + \beta & \text{if} \quad v < -\beta \\ 0 & \text{if} \quad |v| \leqslant \beta \\ v - \beta & \text{if} \quad v > \beta \end{cases}$$

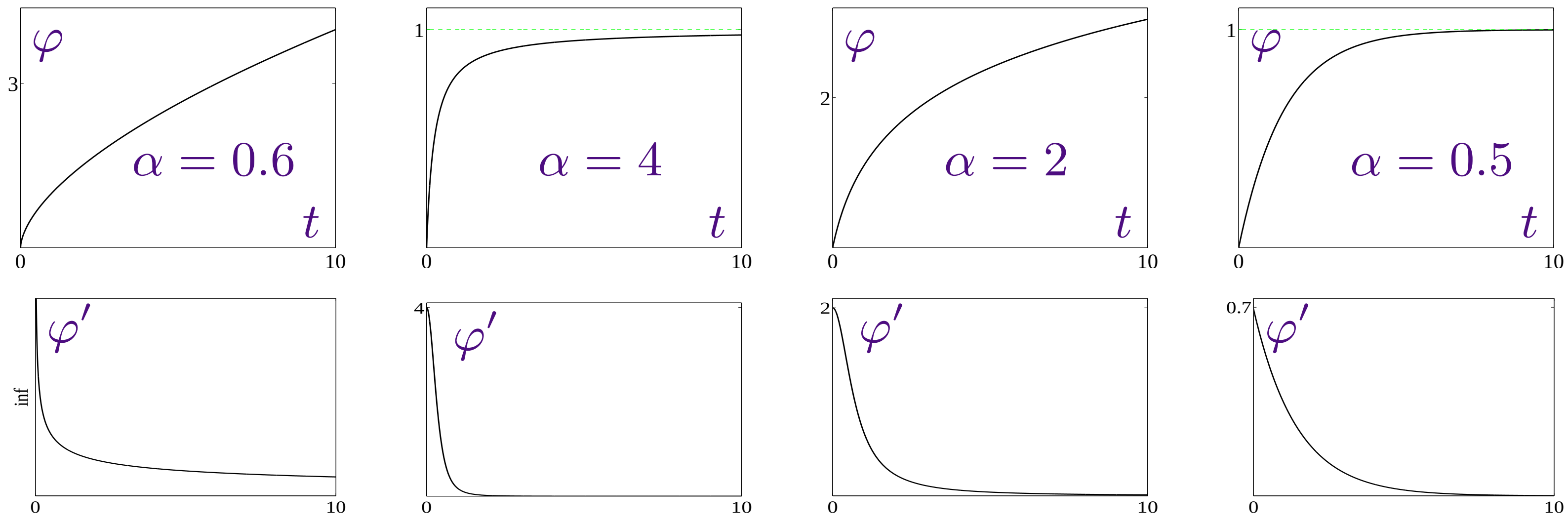
Question **9** What is drawn on the second row?

Question **10** Give a condition for $\mathcal{F}_v$ to have a minimum at $\hat{u}$.

3 Minimizers under Non-Smooth Regularization

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

$$\boldsymbol{\varphi(t)} \ \Big\| \ \boldsymbol{t^\alpha, \ \alpha \in (0,1)} \ \Big| \ \boldsymbol{\frac{\alpha\, t}{\alpha\, t + 1}} \ \Big| \ \boldsymbol{\ln(\alpha t + 1)} \ \Big| \ \boldsymbol{1 - \alpha^t \quad \alpha \in (0,1)} \ \Big| \ (\cdots) \ , \quad \boldsymbol{\alpha > 0}$$



$\varphi(t) = t$ and $G_i u \approx (\nabla u)_i \quad \Rightarrow \quad \mathcal{F}_v(u) = \mathrm{TV}(u)$ (total variation) [Rudin, Osher, Fatemi 92]

[Nikolova 97,01]

Let $\hat{u}$ be a (local) minimizer of $\mathcal{F}_v$. Set $\hat{h} \doteq \{i : G_i\hat{u} = 0\}$

Then $\exists\ O \subset \mathbb{R}^q$ **open,** $\exists\ \mathcal{U} \in \mathcal{C}^{m-1}$ (local) minimizer function so that

$$v' \in O, \quad \hat{u}' = \mathcal{U}(v') \quad \Rightarrow \quad G_i\hat{u}' = 0, \quad \forall\, i \in \hat{h}$$

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

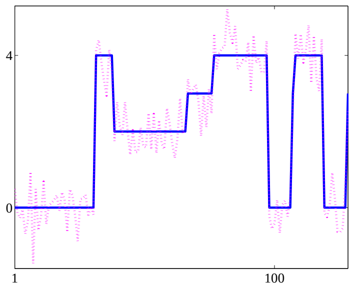
$G_i = \nabla_i \ \Rightarrow\ \hat{u}[i] = \hat{u}[j]$ for many neighbors (i, j) (the "stair-casing" effect)

$G_i u = u[i] \ \Rightarrow$ many samples $\hat{u}[i] = 0$ – highly used in Compressed Sensing

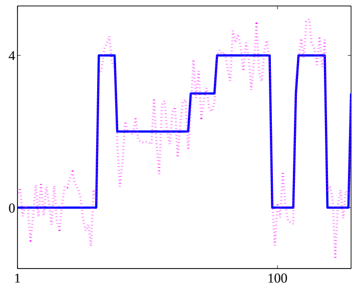
Question **11** What happens if $\{G_i\}$ yield second-order differences?

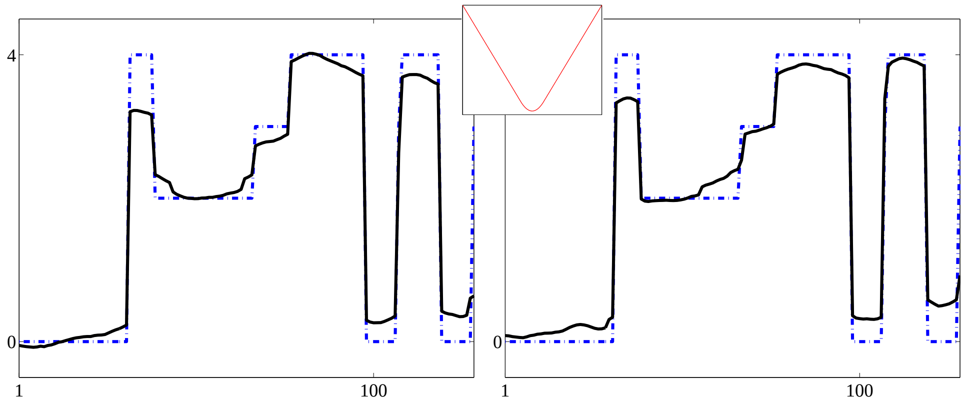
Question **12** Describe the prior that $\hat{u}$ satisfies for a general $\{G_i\}$.

Property <u>fails</u> if $\mathcal{F}_v$ is smooth, except for $v \in N$ where N is closed and $\mathbb{L}^q(N) = 0$.

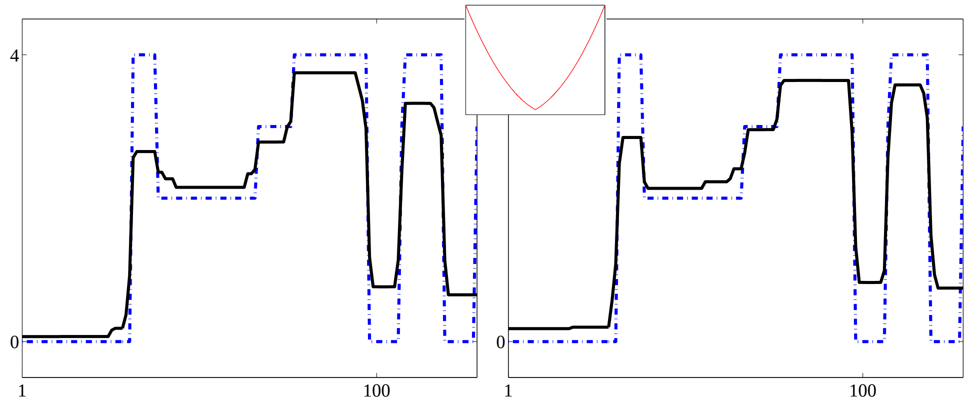


$$\mathcal{F}_v(u) = \|u - v\|^2$$
$$+\beta \sum \varphi(|u[i] - u[i-1]|)$$

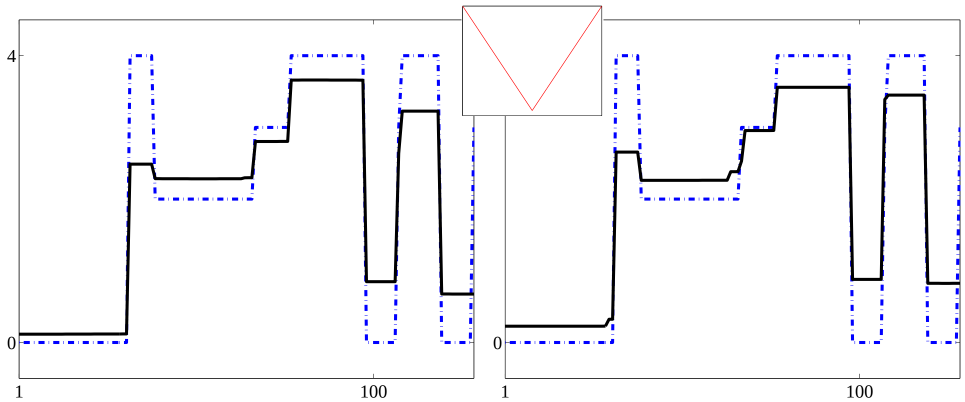




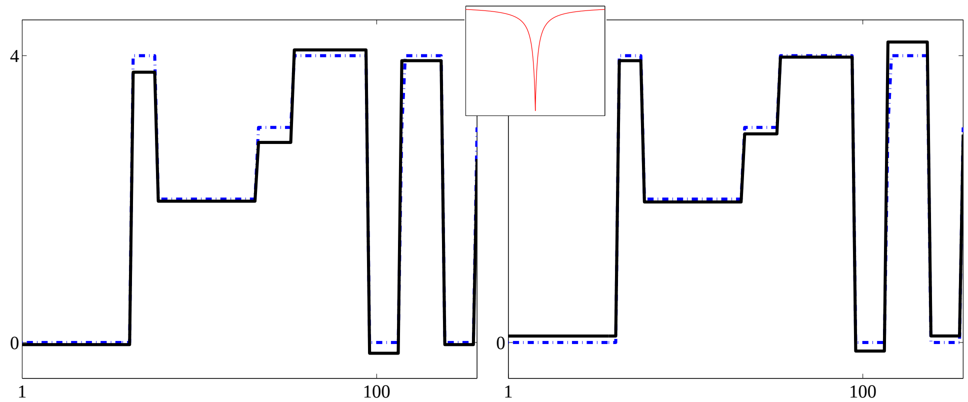
$\varphi(t) = \sqrt{\alpha + t^2}, \quad \varphi'(0) = 0$ **(smooth at 0)**



$\varphi(t) = (t + \alpha \mathrm{sign}(t))^2, \quad \varphi'(0^+) = 2\alpha$



$\varphi(t) = |t|, \quad \varphi'(0^+) = 1$



$\varphi(t) = \alpha|t|/(1 + \alpha|t|), \quad \varphi'(0^+) = \alpha$

Let $u_o \in \mathbb{R}$ and $\mathrm{pdf}(u_o) = \frac{1}{2} e^{-|u_o|}$ (Laplacian distribution)

Question **13** Give $\Pr(u_o = 0)$.

Let $v = u_o + n$ where $\mathrm{pdf}(n) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{n^2}{2\sigma^2}}$ (centered Gaussian distribution)

The corresponding MAP energy to recover u_o from v reads as

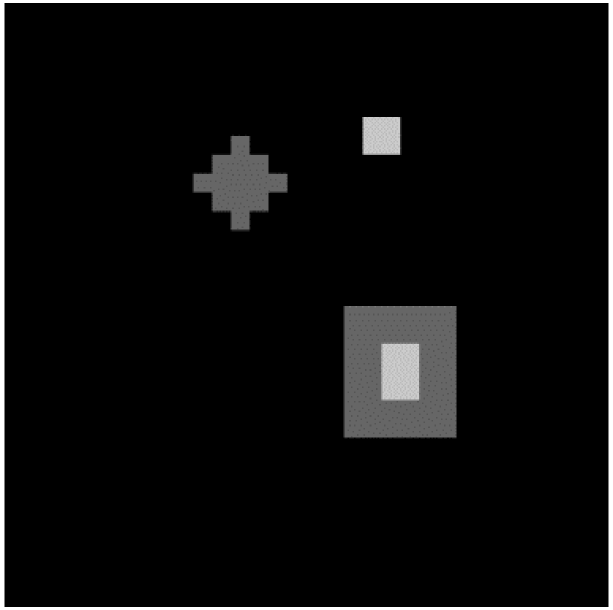
$$\mathcal{F}_v(u) = \frac{1}{2}(u - v)^2 + \beta|u| \quad \text{for} \quad \beta = \frac{1}{\sigma^2}$$

Question **14** Give the minimizer function $\mathcal{U}$ for $\mathcal{F}_v$.

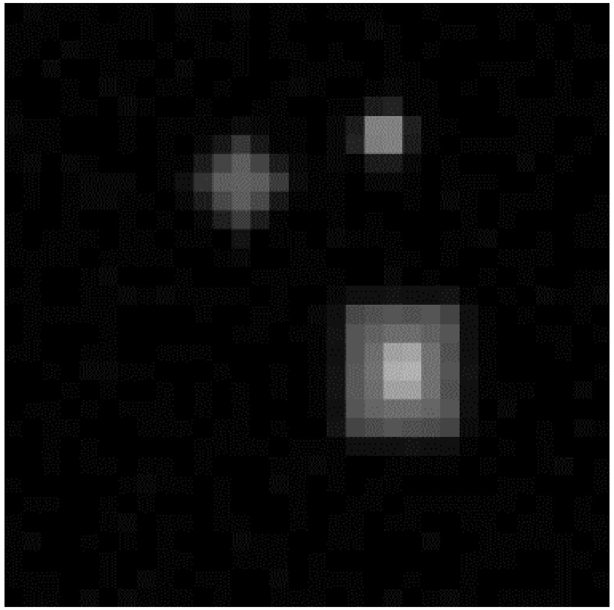
Useful reminder on p. 21.

Question **15** Determine the set $\{\nu \in \mathbb{R} \ : \ \mathcal{U}(\nu) = 0\}$. Comment the result.

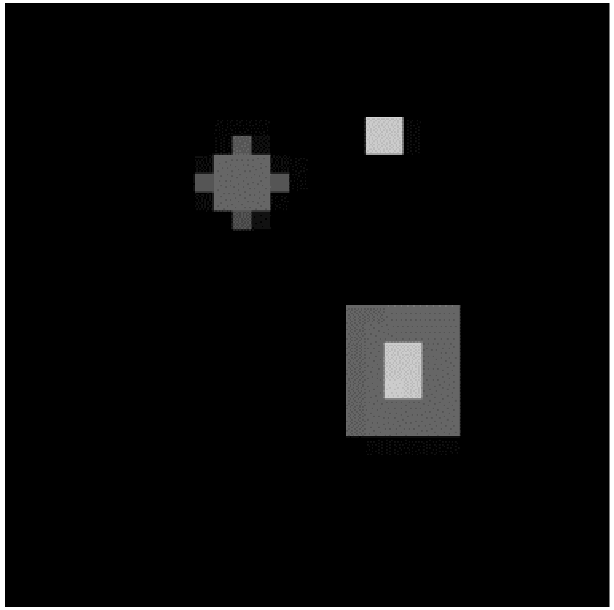
TV energy: $\mathcal{F}_v(u) = \|Au - v\|^2 + \beta \sum_{i=1}^{r} \varphi(\|G_i u\|)$ for $\varphi(t) = t$ and G_i discrete gradient at pixel i



Original

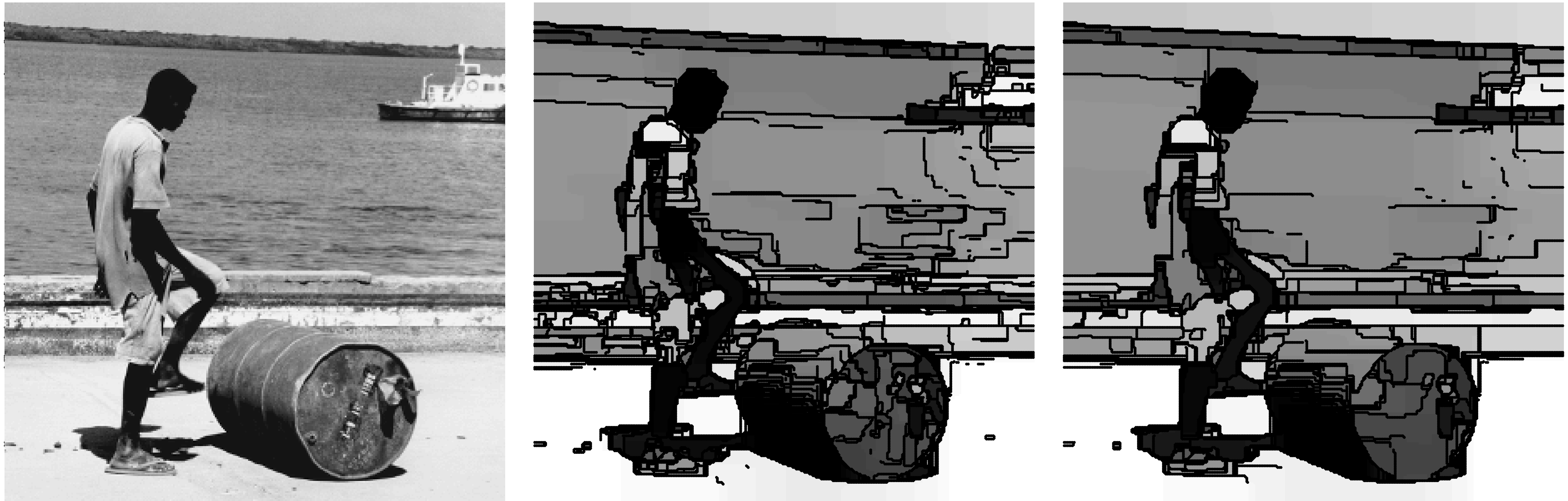


Data



Restored: TV energy

D. C. Dobson and F. Santosa, "Recovery of blocky images from noisy and blurred data", SIAM J. Appl. Math., 56 (1996), pp. 1181-1199.



Minimization of $\mathcal{F}_v(u) = \|u - v\|_2^2 + \beta \mathrm{TV}(u)$, $\quad \beta = 100$ and $\beta = 180$

Here $\quad \boldsymbol{\varphi(t)} = \begin{cases} \mathbf{0} & \text{if} \quad \boldsymbol{t = 0} \\ \mathbf{1} & \text{if} \quad \boldsymbol{t \neq 0} \end{cases}$

Question **16** Compute the global minimizer of $\mathcal{F}_v(u) = (u - v)^2 + \beta\varphi(u)$ for $u, v \in \mathbb{R}$ and $\beta > 0$, according to the value of v.

Question **17** Are there any values of v so that $\mathcal{F}_v$ has more than one global minimizer?

Consider $\boldsymbol{\mathcal{F}_v(u) = \|u - v\|_2^2 + \beta \sum_{i=1}^{p} \varphi(u[i])}$ for $\beta > 0$ and $u, v \in \mathbb{R}^p$.

The global minimizer function $\mathcal{U} : \mathbb{R}^p \to \mathbb{R}^p$ for $\mathcal{F}_v$ has p components which depend on v.

Question **18** Compute each component $\mathcal{U}_i$

Question **19** Let $h \subset \{1, \cdots, p\}$. Determine the subset $\mathcal{O}_h \subset \mathbb{R}^p$ such that if $v \in \mathcal{O}_h$ then the global minimizer $\hat{u}$ of $\mathcal{F}_v$ satisfies $\hat{u}[i] = 0, \ \forall\, i \in h$ and $\hat{u}[i] \neq 0$ if $i \notin h$.

Note that $\quad \displaystyle\sum_{i=1}^{p} \varphi(u[i]) = \#\{i \ : \ u[i] \neq 0\} = \ell_0(u)$

4 Minimizers relevant to non-smooth data-fidelity

General case

[Nikolova 01,02]

$$\boldsymbol{\mathcal{F}_v(u)=\sum_i \psi(|a_i u - v[i]|) + \beta\Phi(u)}, \qquad \Phi\in\mathcal{C}^m,\ \psi\in\mathcal{C}^m(\mathbb{R}_+^*),\quad \boldsymbol{\psi'(0^+)>0}$$

Let $\hat{\boldsymbol{u}}$ be a (local) minimizer of $\mathcal{F}_v$. Set $\boldsymbol{\hat{h}=\{i : a_i\hat{u}=v[i]\}}$.
Then $\exists\ \boldsymbol{O}\subset\mathbb{R}^q$ open, $\exists\ \boldsymbol{\mathcal{U}}\in\mathcal{C}^{m-1}$ (local) minimizer function so that

$$\boldsymbol{v'\in O,\quad \hat{u}'=\mathcal{U}(v')} \quad\Rightarrow\quad \left\{\begin{array}{ll} \boldsymbol{a_i\hat{u}'=v[i],} & \boldsymbol{i\in\hat{h}} \\ \boldsymbol{a_i\hat{u}'\neq v[i],} & \boldsymbol{i\in\hat{h}^c} \end{array}\right.$$

(Local) minimizers $\hat{\boldsymbol{u}}$ of $\boldsymbol{\mathcal{F}_v}$ achieve an **exact fit** to (noisy) data
$\boldsymbol{a_i\hat{u}=v[i]}$ for a certain number of indexes $\boldsymbol{i}$

Property fails if $\mathcal{F}$ is smooth, except for $v\in N$ where N is closed and $\mathbb{L}^q(N)=0$.

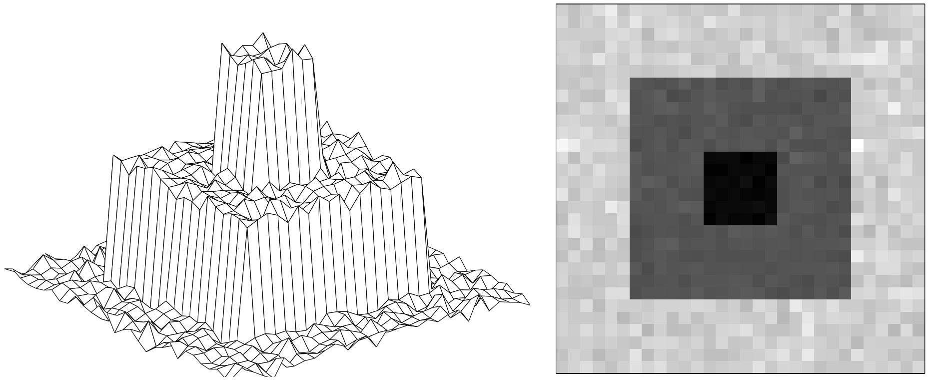
Question **20** Suggest cases when you would like that your minimizer obeys this property.

Question **21** Propose some choices for ψ. Explain.

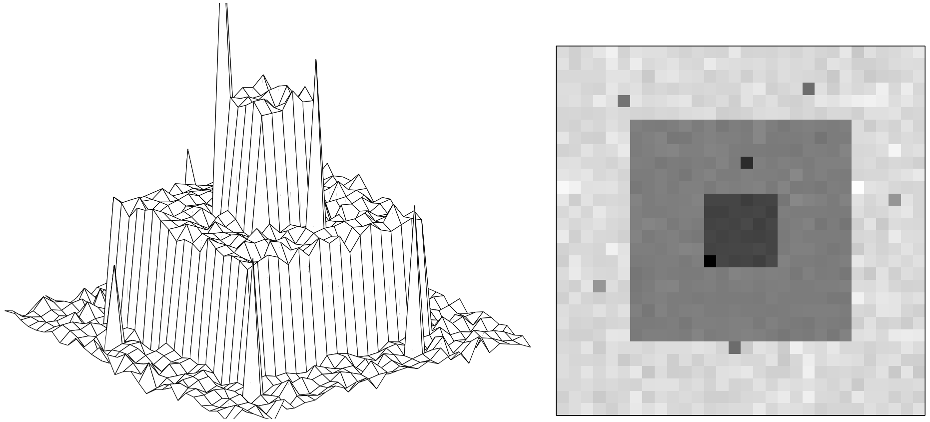
Question **22** Compute the minimizer of $\mathcal{F}_v(u) = |u - v| + \beta u^2$ for $u, v \in \mathbb{R}$ and $\beta > 0$.

Question **23** Explain the relationship between the properties of the minimizer

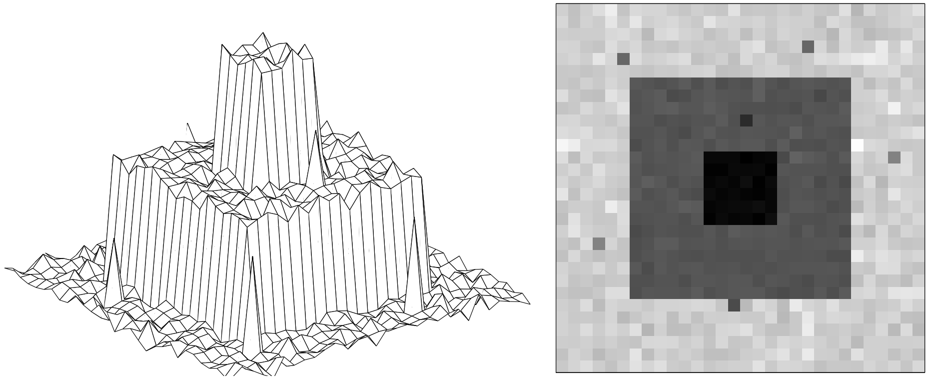
when $\varphi'(0^+) > 0$ and when $\psi'(0^+) > 0$



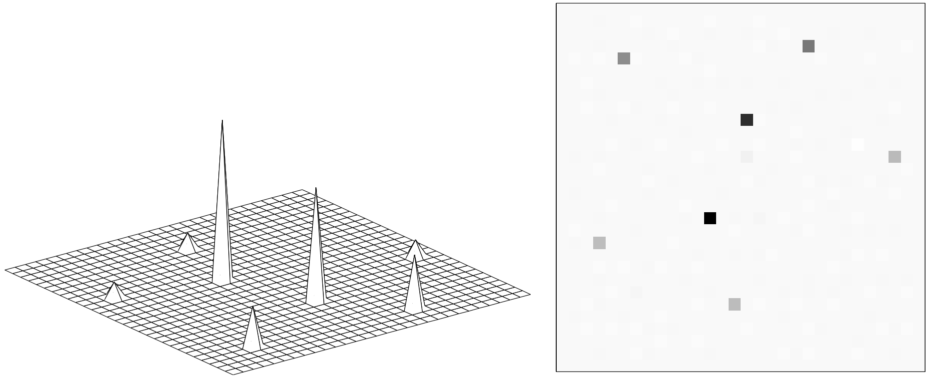
Original u_o



Data $v = u_o$+outliers

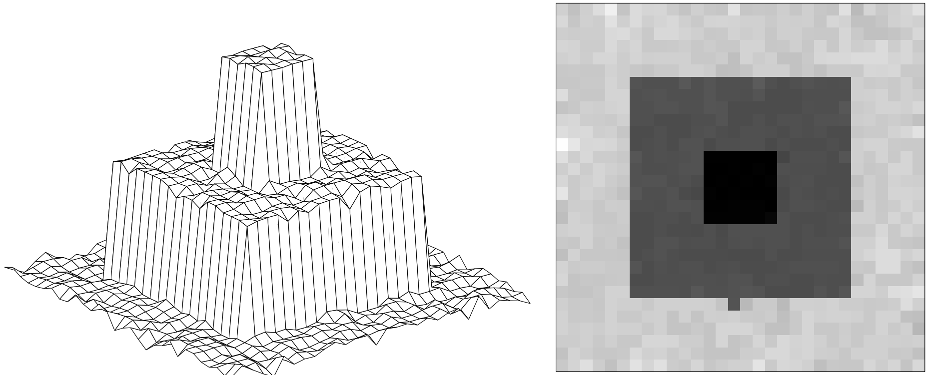


Restoration $\hat{u}$ for $\boldsymbol{\beta = 0.14}$

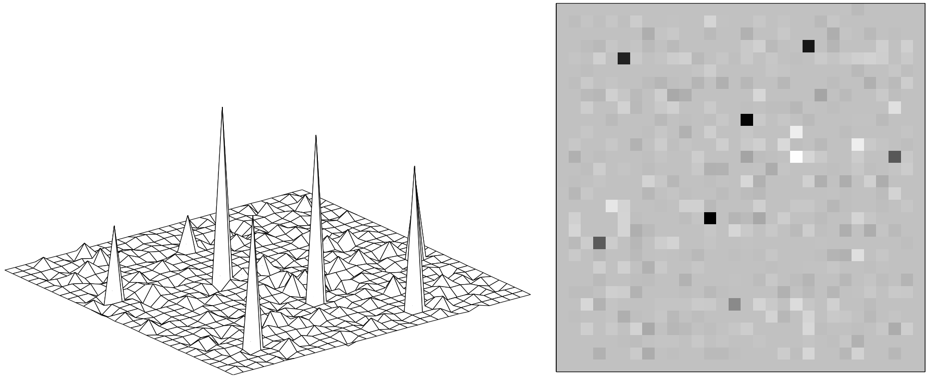


Residuals $v - \hat{u}$

$$\boldsymbol{\mathcal{F}_v(u) = \sum_i |u[i] - v[i]| + \beta \sum_{j \in \mathcal{N}_i} |u[i] - u[j]|^{1.1}}$$

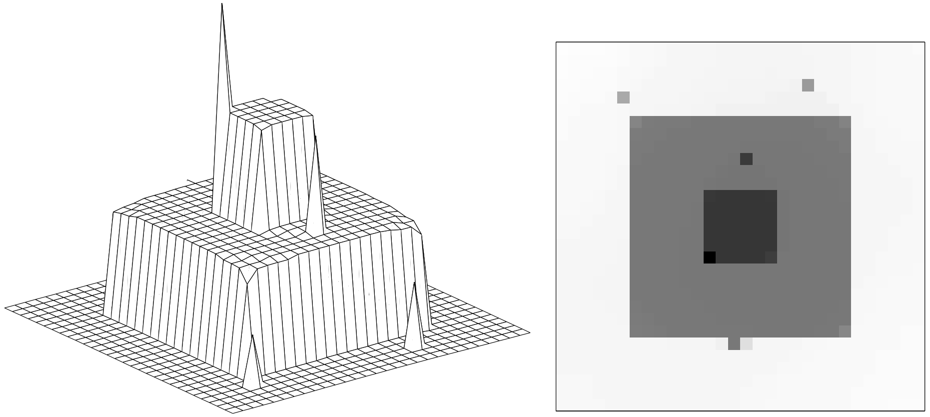


Restoration $\hat{u}$ for $\boldsymbol{\beta = 0.25}$

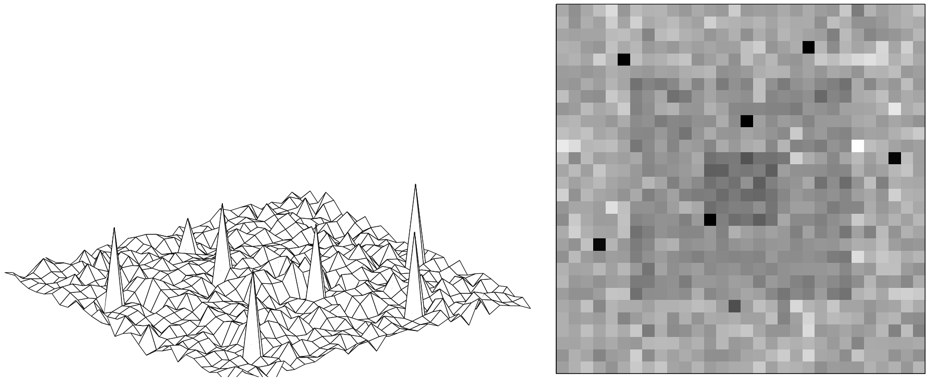


Residuals $v - \hat{u}$

$$\boldsymbol{\mathcal{F}_v(u) = \sum_i \big|u[i] - v[i]\big| + \beta \sum_{j \in \mathcal{N}_i} |u[i] - u[j]|^{1.1}}$$



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

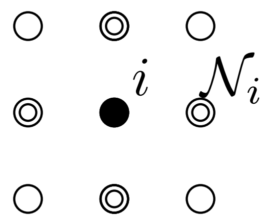


Residuals $v - \hat{u}$

TV-like energy: $\boldsymbol{\mathcal{F}_v(u) = \sum_i (u[i] - v[i])^2 + \beta \sum_{j \in \mathcal{N}_i} |u[i] - u[j]|}$

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*

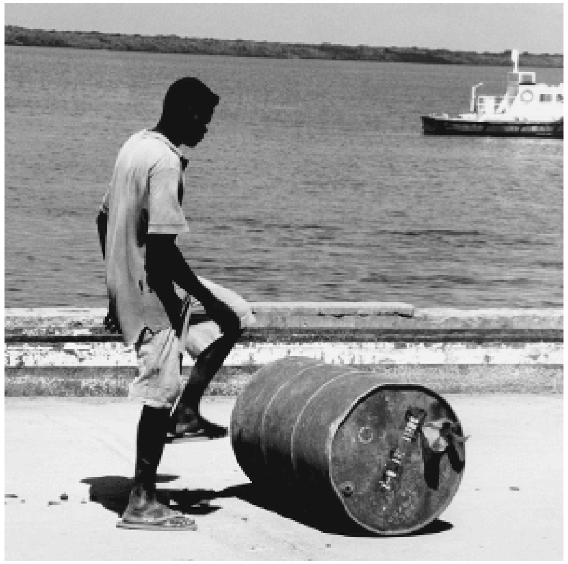
Assumptions:
$$\begin{cases} \text{data } v \text{ contain uncorrupted samples } v[i] \\ v[i] \text{ is outlier if } |v[i] - v[j]| \gg 0, \quad \forall j \in \mathcal{N}_i \end{cases}$$

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

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

Recovery of frame coefficients using ℓ_1 data-fitting

- Data: $v = u_o +$ noise
- Frame coefficients: $y = Wv = Wu_o +$ noise $\qquad \widetilde{W}$ =left inverse of W
- Hard thresholding $\quad y_T[i] \doteq \begin{cases} 0 & \text{if } |y[i]| \leqslant T \\ y[i] & \text{if } |y[i]| > T \end{cases}$

  keeps relevant information if T small
- $\tilde{u} = \widetilde{W} y_T$ — Gibbs oscillations and wavelet-shaped artifacts
- Hybrid energy methods—combine fitting to y_T with prior $\Phi(u)$

[Bobichon, Bijaoui 97], [Coifman, Sowa 00], [Durand, Froment 03]...

[Durand, Nikolova 07]

Desiderata: $\mathcal{F}_y$ convex and

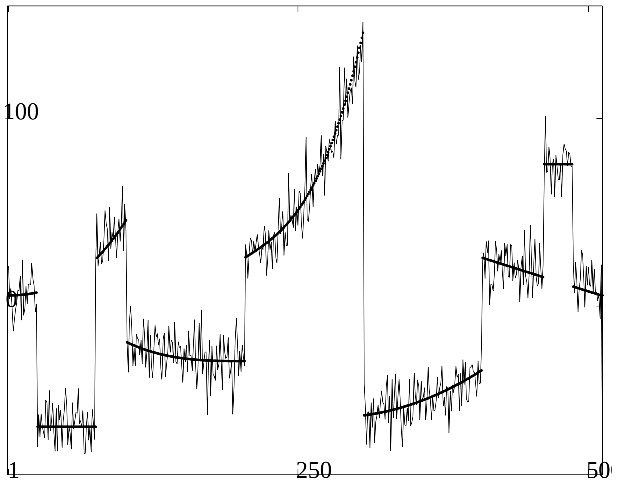
Keep $\hat{x}[i] = y_T[i]$	Restore $\hat{x}[i] \neq y_T[i]$
significant coefs: $y[i] \approx (Wu_o)[i]$	outliers: $\lvert y[i]\rvert \gg \lvert(Wu_o)[i]\rvert$ (frame-shaped artifacts)
thresholded coefs: $(Wu_o)[i] \approx 0$	edge coefs: $\lvert(Wu_o)[i]\rvert > \lvert y_T[i]\rvert = 0$ ("Gibbs" oscillations)

Then:

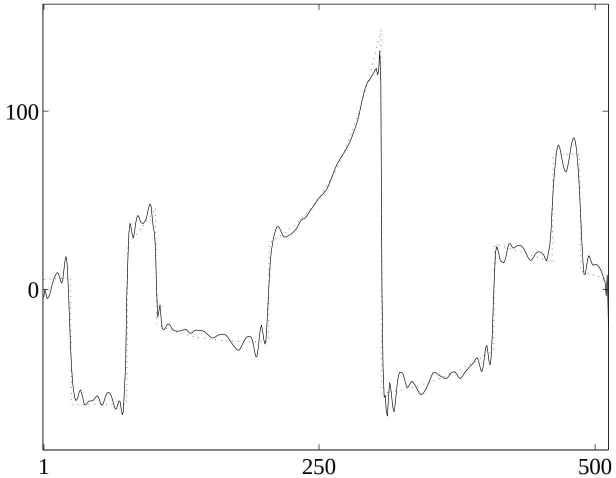
$$\boxed{\begin{array}{c}\text{minimize} \quad \mathcal{F}_y(x) = \sum_i \lambda_i \big|(x - y_T)[i]\big| + \int_\Omega \varphi(|\nabla \widetilde{W} x|) \quad \Rightarrow \quad \hat{x} \\ \hat{u} = \widetilde{W}\hat{x} \ \text{ for } \ \widetilde{W} \ \text{ left inverse, } \varphi \text{ edge-preserving}\end{array}}$$

Question **24** Explain why the minimizers of $\mathcal{F}_y$ fulfill the desiderata.

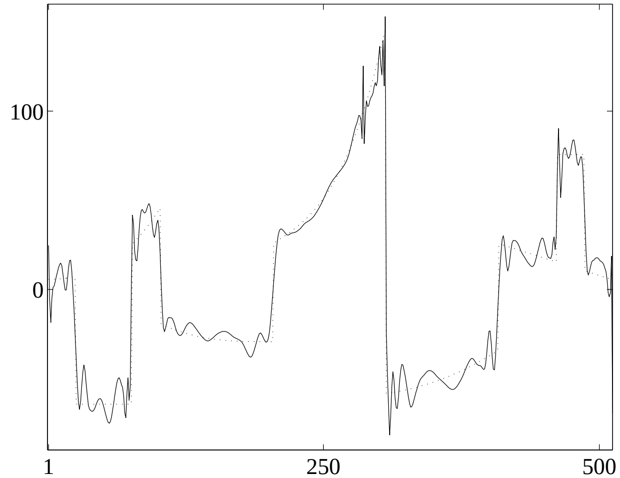
Question **25** Any open questions?



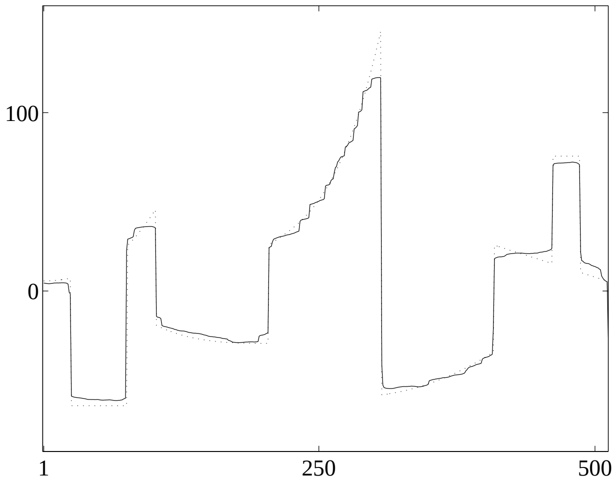
Original and data



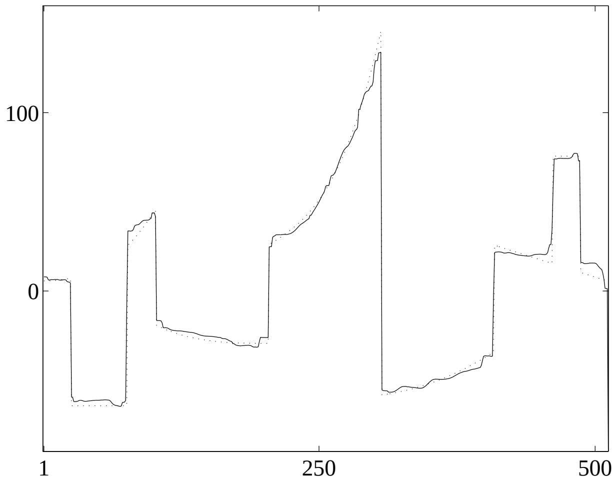
Sure-shrink method



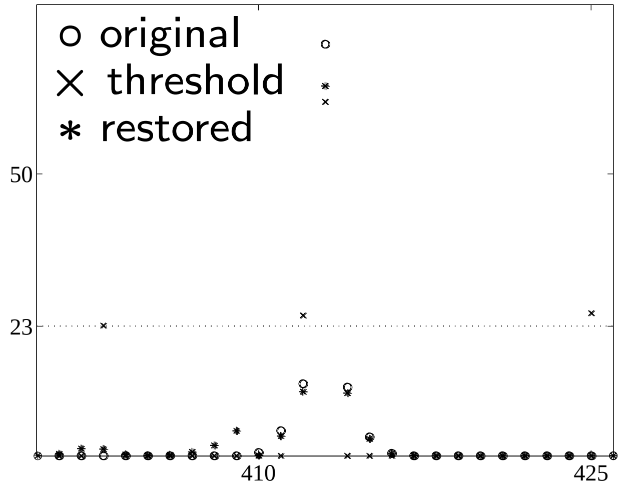
Hard thresholding



Total variation



The proposed method



Magnitude of coefficients

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

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

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

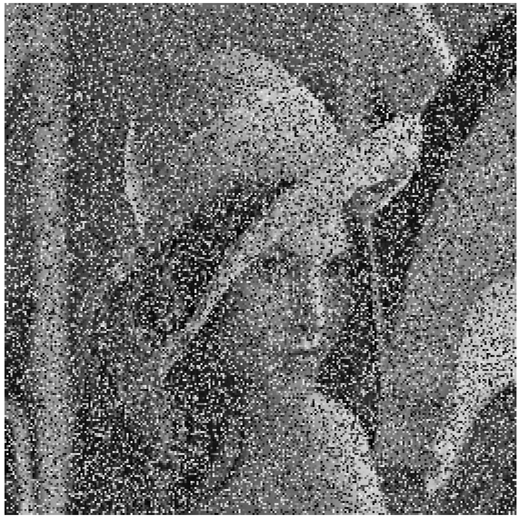
   (e.g. adaptive or center-weighted median)

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

   $\Rightarrow$ improve speed and accuracy

2. Restore $\hat{u}$ (denoise, deblur) using an edge-preserving energy method

   subject to $a_i \hat{u} = v[i]$ for all $i \in \hat{h}$

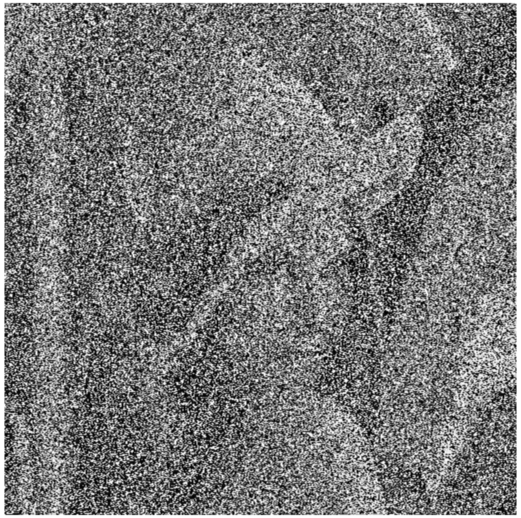


50% RV noise

ACWMF

DPVM

Our method



70 %SP noise(6.7dB)

Adapt.med.(25.8dB)

Our method(29.3dB)

Original Lena

L. Bar, A. Brook, N. Sochen and N. Kiryati,
"Deblurring of Color Images Corrupted by Impulsive Noise",
IEEE Trans. on Image Processing, 2007

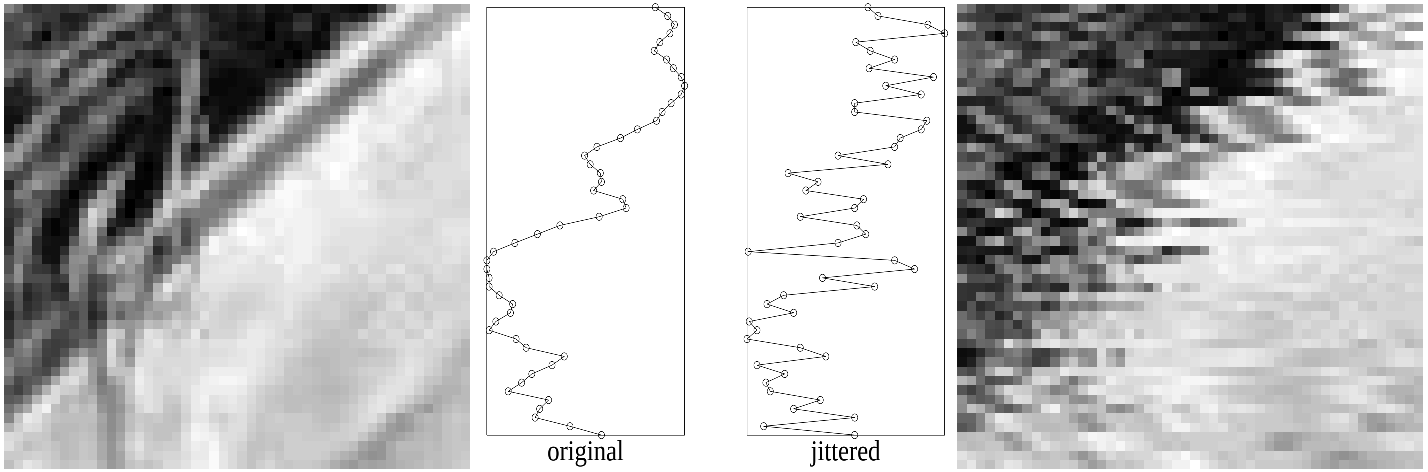
$$\mathcal{F}_v(u) = \|Au - v\|_1 + \beta\Phi(u)$$

blurred, noisy (r.-v.)

zoom - restored

One-step real-time dejittering of digital video [Nikolova 09]

- Image $\boldsymbol{u} \in \mathbb{R}^{m \times n}$, rows $\boldsymbol{u_i}$, its pixels $\boldsymbol{u_i}[\boldsymbol{j}]$
- Data $v_i[j] = u_i[j + d_i]$, d_i integer, $|d_i| \leqslant M$, typically $M \leqslant 20$.
- Restore $\hat{\boldsymbol{u}} \equiv$ restore $\hat{\boldsymbol{d}}_i$, $\boldsymbol{1} \leqslant \boldsymbol{i} \leqslant \boldsymbol{m}$



Original (b) One column Jittered

(b) The same column in the original (left) and in the jittered (right) image

The gray-values of the columns of natural images can be seen as large pieces of 2^{nd} (or 3^{rd}) order polynomials which is false for their jittered versions.

Each column $\hat{\boldsymbol{u}}_i$ is restored using $\hat{d}_i = \arg \min_{|d_i| \leqslant N} \mathcal{F}(d_i)$

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

Question **26** Explain why the minimizers of $\mathcal{F}$ can solve the problem as stated.

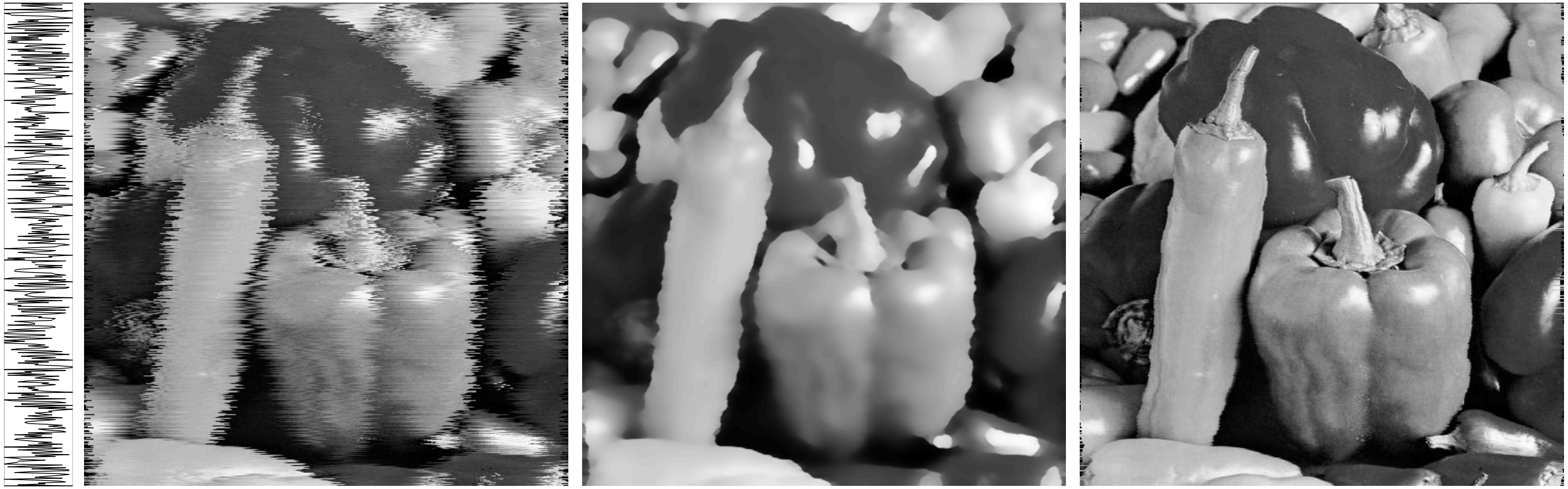
Question **27** What changes if $\alpha = 1$ or if $\alpha = 0.5$?

Question **28** Is it easy to solve the numerical problem?

A Monte-Carlo experiment shows that in almost all cases, $\boldsymbol{\alpha = 0.5}$ is the best choice.

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

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



Jitter Jittered Image Bayesian TV Bake & Shake

Original Our: α=0.5 Our: Error $u_o - \hat{u}$

[Kokaram98, Laborelli03, Shen04, Kang06, Scherzer11]

Comparison with Smooth Energies

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

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

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

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

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

Question **29** What are the consequences if one approximates a nonsmooth energy by a smooth energy?

Let $u \in \mathbb{R}^p$ and $v \in \mathbb{R}^q$.

Consider that $A \in \mathbb{R}^{q\times p}$ and $G \in \mathbb{R}^{r\times p}$ satisfy $\ker(A) \cap \ker(G) = \{0\}$.

$$\mathcal{F}_v(u) = \|Au - v\|_2^2 + \beta\|Gu\|_2^2 \quad \text{for} \quad \beta > 0$$

Question **30** Calculate $\nabla\mathcal{F}_v(u)$.

Question **31** Determine the minimizer function $\mathcal{U}$.

Let $G_i \in \mathbb{R}^{1\times p}$ denote the ith row of G.

Question **32** Characterize the set $\mathcal{K} = \{\nu \in \mathbb{R}^p \ : \ G_i\,\mathcal{U}(\nu) = 0\}$.

Let $a_i \in \mathbb{R}^{1\times p}$ denote the ith row of A.

Question **33** Characterize the set $\mathcal{L} = \{\nu \in \mathbb{R}^p \ : \ a_i\,\mathcal{U}(\nu) = \nu[i]\}$.

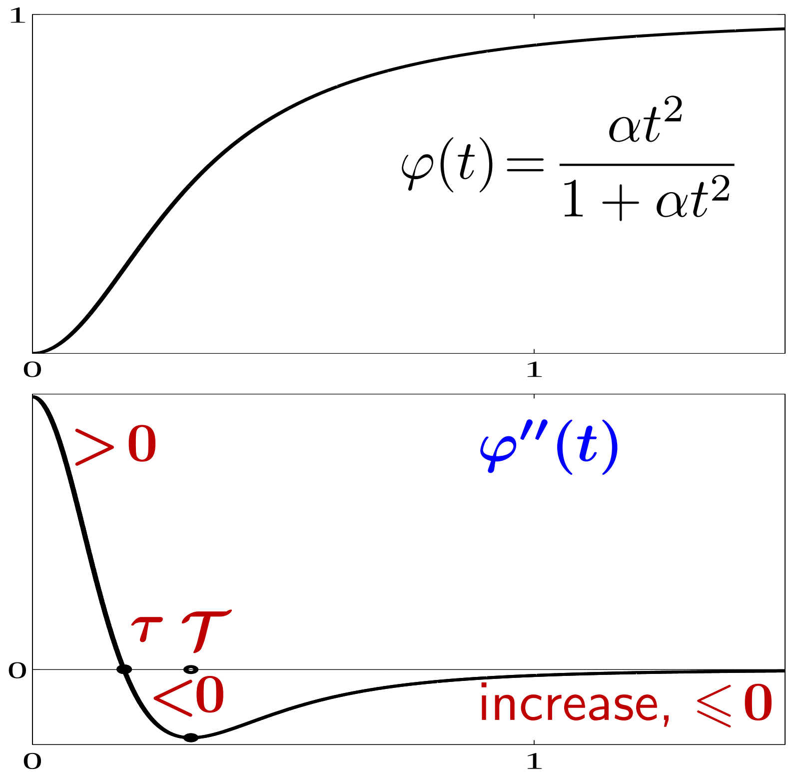
5 Nonconvex Regularization: Why Edges are Sharp?

[Nikolova 04, 10]

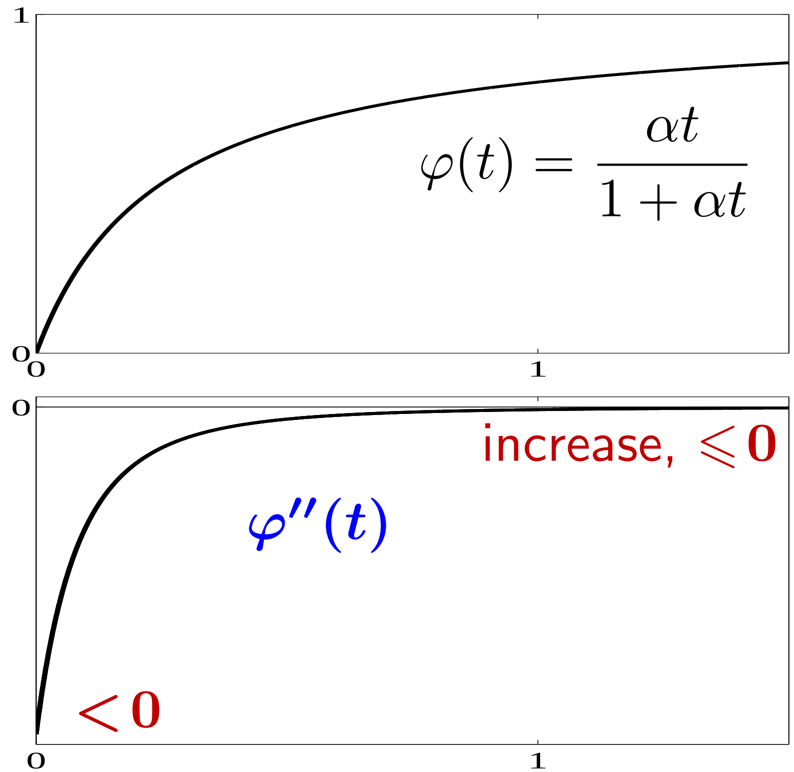
$$\boxed{\mathcal{F}_v(u) = \|Au - v\|^2 + \beta \sum_{i \in J} \varphi(\|G_i u\|)} \qquad J = \{1, \cdots, r\}$$

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)



Sharp edge property

There exist $\boldsymbol{\theta_0 \geqslant 0}$ and $\boldsymbol{\theta_1 > \theta_0}$ such that any (local) minimizer $\hat{\boldsymbol{u}}$ of $\boldsymbol{\mathcal{F}_v}$ satisfies

$$\text{either} \quad \|\boldsymbol{G_i \hat{u}}\| \leqslant \boldsymbol{\theta_0} \quad \text{or} \quad \|\boldsymbol{G_i \hat{u}}\| \geqslant \boldsymbol{\theta_1} \qquad \forall\, \boldsymbol{i \in J}$$

$$\begin{array}{lcll} \widehat{\boldsymbol{h}}_{\boldsymbol{0}} & = & \{\boldsymbol{i} : \|\boldsymbol{G_i \hat{u}}\| \leqslant \boldsymbol{\theta_0}\} & \text{homogeneous regions} \\ \widehat{\boldsymbol{h}}_{\boldsymbol{1}} & = & \{\boldsymbol{i} : \|\boldsymbol{G_i \hat{u}}\| \geqslant \boldsymbol{\theta_1}\} & \text{edges} \end{array}$$

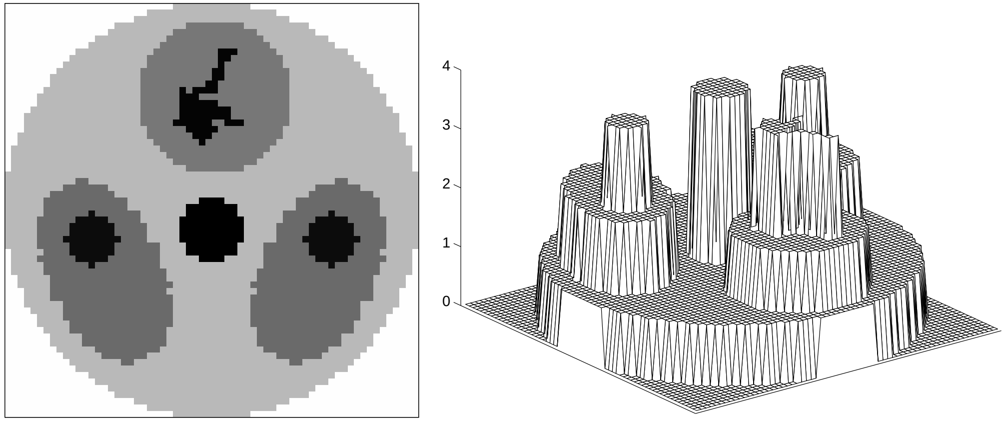
When β increases, then θ_0 decreases and θ_1 increases.

In particular

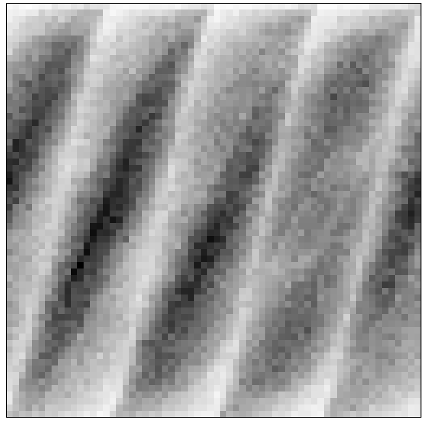
$$\boxed{\boldsymbol{\varphi'(0^+) > 0 \quad \Rightarrow \quad \theta_0 = 0} \quad \text{fully segmented image} \quad (G_i\hat{u} = 0,\ \forall i \in \widehat{h}_0)}$$

Question **34** Explain the prior model involved in $\mathcal{F}_v$ when φ is nonconvex with $\varphi'(0) = 0$ and with $\varphi'(0^+) > 0$.

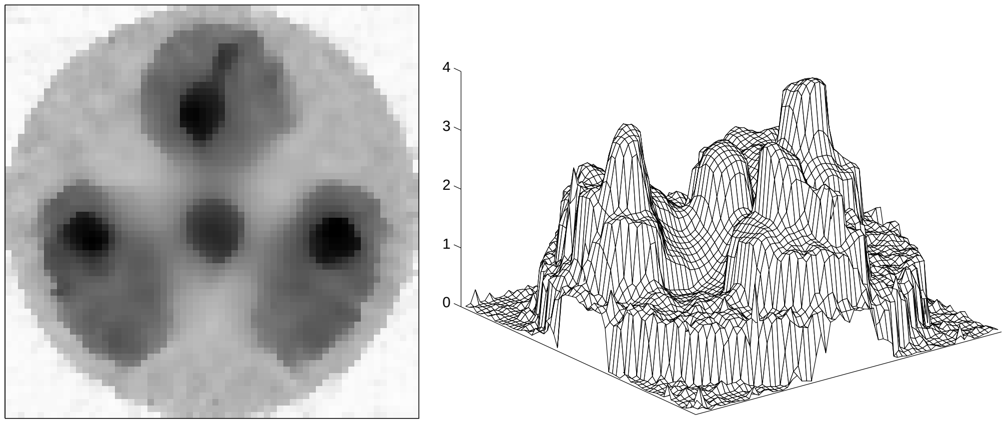
Image Reconstruction in Emission Tomography



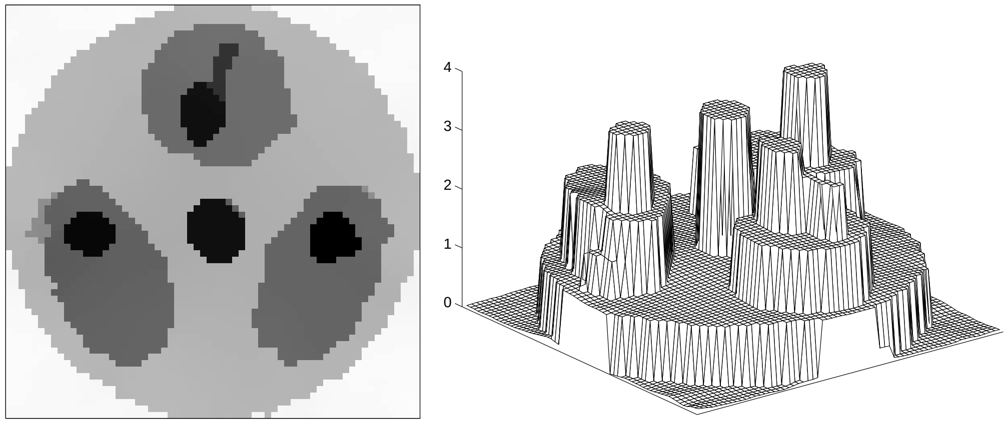
Original phantom



Emission tomography simulated data



φ is smooth (Huber function)



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

Reconstructions using $\mathcal{F}_v(u) = \Psi(u, v) + \beta \sum_{j \in \mathcal{N}_i} \varphi(|u[i] - u[j]|)$, $\quad \Psi =$ smooth, convex

$$\mathcal{F}_v(u) = (u-v)^2 + \beta\varphi(u) \quad u, v \in \mathbb{R} \quad \beta > 0$$

- Assumption: $\beta > -\dfrac{2}{\min\limits_{t\in\mathbb{R}} \varphi''(t)}$ (if $\varphi'(0^+) > 0$ then $\min\limits_{t\in\mathbb{R}} \varphi''(t) = \varphi''(0^+)$).

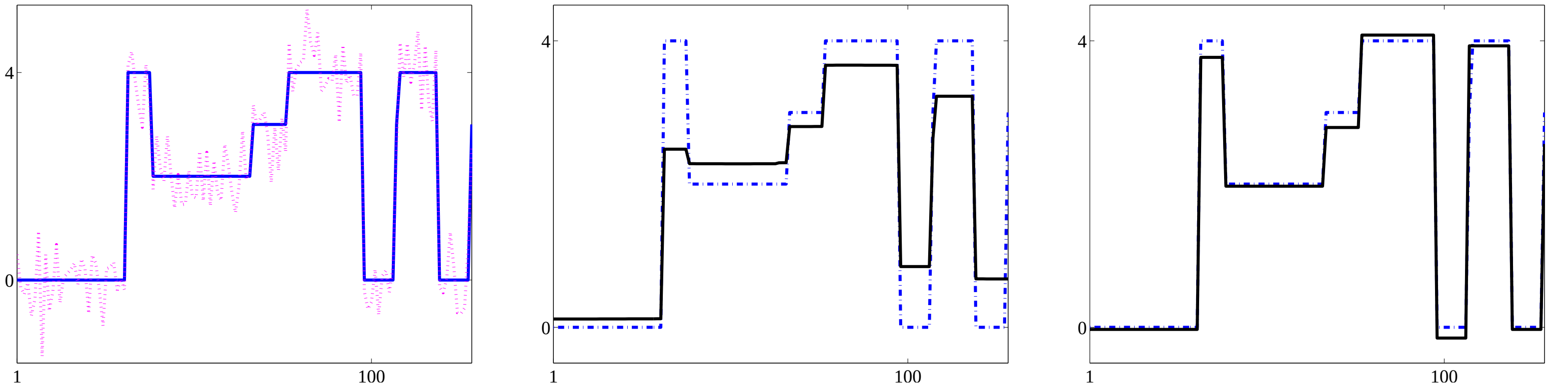
Question **35** Determine the sign of β, i.e. > 0 or < 0.

- $C_\beta \doteq \left\{ t \in (0, +\infty) \;:\; \varphi''(t) < -\dfrac{2}{\beta} \right\}$
- Recall: $\mathcal{F}_v$ has a (local) minimum at a $\hat{u}$ where $\mathcal{F}_v$ is twice differentiable if and only if

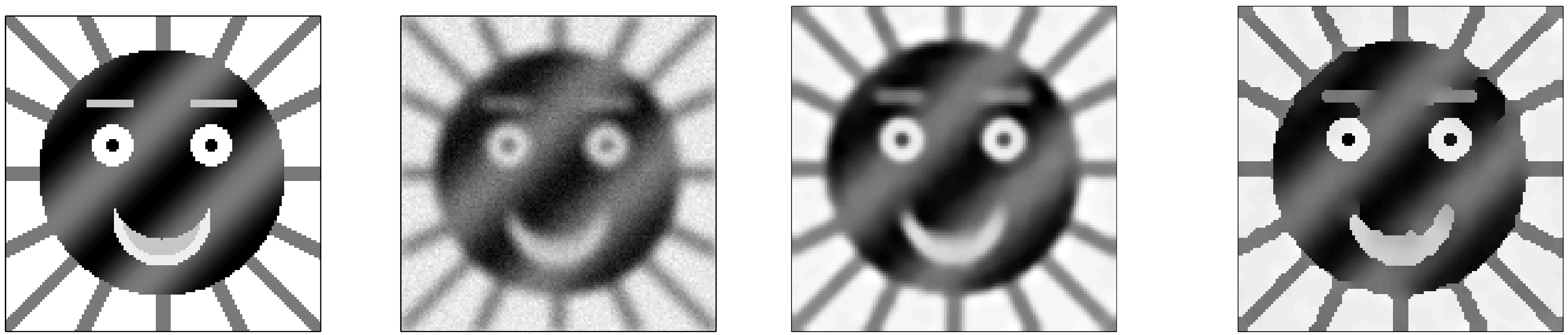
$$\mathcal{F}_v'(\hat{u}) = 0 \quad \text{and} \quad \mathcal{F}_v''(\hat{u}) \geqslant 0$$

Question **36** Show that $\forall\, v \in \mathbb{R}$, if $\hat{u}$ is a (local) minimizer of $\mathcal{F}_v$, then $|\hat{u}| \notin C_\beta$.

Comparison with Convex Edge-Preserving Regularization

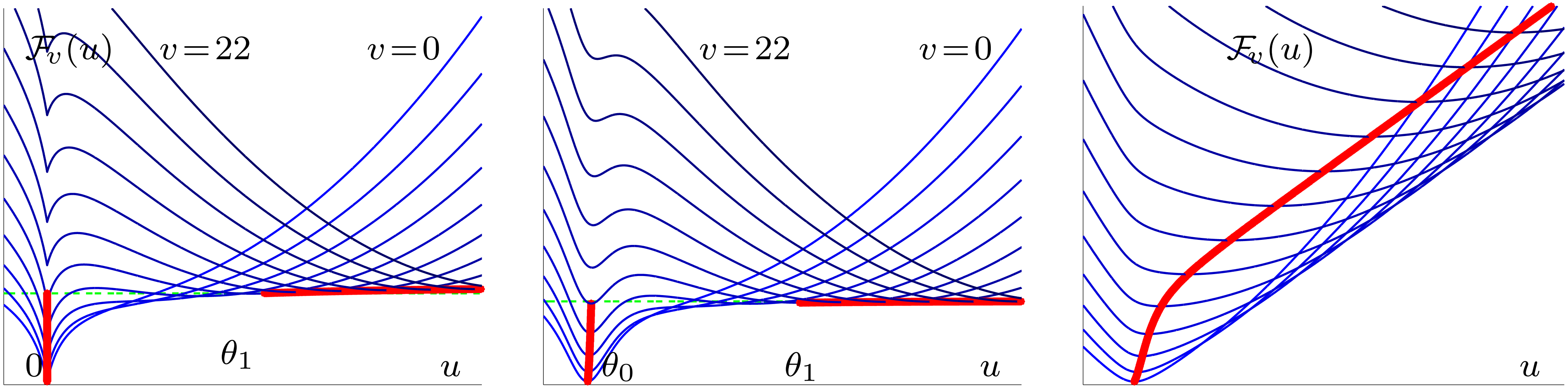


Data $v = u_o + n$ | $\varphi(t) = |t|$ | $\varphi(t) = \alpha|t|/(1 + \alpha|t|)$



original | data | $\varphi(t) = |t|^{1.4}$ | $\varphi(t) = \min\{\alpha t^2, 1\}$

Question **37** Why edges are sharper when φ is nonconvex?



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

global function (●●●) $\qquad$ global minimizer functions (●●●) $\qquad$ unique minimizer function (●●●)

Each blue curve curve: $\boldsymbol{u} \to \boldsymbol{\mathcal{F}_v(u)}$ for $\boldsymbol{v} \in \{0, 2, \cdots\}$

Question **38** How to describe the global minimizer when v increases?

6. Nonsmooth data-fidelity and regularization

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

The L_1-TV energy

T. F. Chan and S. Esedoglu, "Aspects of Total Variation Regularized L^1 Function Approximation", SIAM J. on Applied Mathematics, 2005

$$\mathcal{F}_v(u) = \|u - \mathbb{1}_\Omega\|_1 + \beta \int_{\mathbb{R}^d} \|\nabla u(x)\|_2 \, dx \quad \text{where} \quad \mathbb{1}_\Omega(x) \doteq \begin{cases} 1 & \text{if} \quad x \in \Omega \\ 0 & \quad\text{else} \end{cases}$$

- $\exists\, \hat{u} = \mathbb{1}_\Sigma$ $\quad$ (Ω convex $\Rightarrow$ $\Sigma \subset \Omega$ and $\hat{u}$ unique for almost every $\beta > 0$)
- contrast invariance: if $\hat{u}$ minimizes for $v = \mathbb{1}_\Omega$ then $c\hat{u}$ minimizes $\mathcal{F}_{cv}$

  the contrast of image features is more important than their shapes
- critical values β^* $\begin{cases} \beta < \beta^* & \Rightarrow \quad \text{objects in } \hat{u} \text{ with good contrast} \\ \beta > \beta^* & \Rightarrow \quad \text{they suddenly disappear} \end{cases}$

  $\Rightarrow$ data-driven scale selection

Binary images by L1 − TV

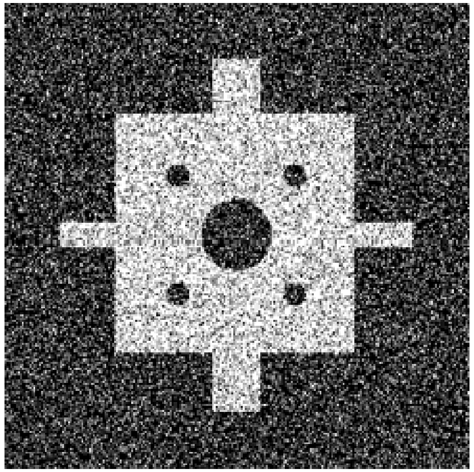
[T. Chan, S. Esedoglu, Nikolova 06]

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

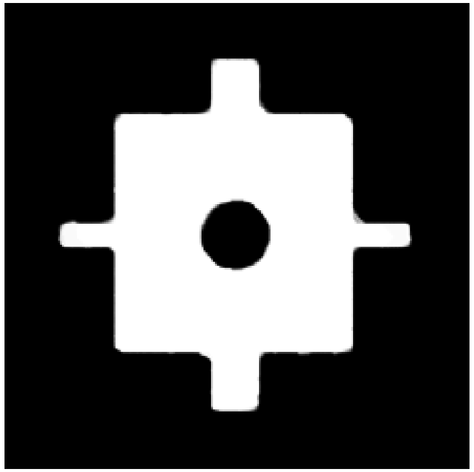
$$\hat{\Sigma} = \arg\min_{\Sigma} \left\{ \left\| \mathbb{1}_\Sigma - \mathbb{1}_\Omega \right\|_2^2 + \beta \mathrm{TV}(\mathbb{1}_\Sigma) \right\} \qquad \text{nonconvex problem} \qquad (\star)$$

usual techniques (curve evolution, level-sets) fail

$\hat{\Sigma}$ solves $(\star)$ $\Leftrightarrow$ $\hat{\boldsymbol{u}} = \mathbb{1}_{\hat{\Sigma}}$ minimizes $\left\| u - \mathbb{1}_\Omega \right\|_1 + \beta\, \mathrm{TV}(u)$ **(convex)**



Data



Restored