Summer School 2014: **Inverse Problem and Image Processing**

Inverse modeling in inverse problems using optimization

Mila Nikolova

CMLA – CNRS, ENS Cachan, France

nikolova@cmla.ens-cachan.fr

http://mnikolova.perso.math.cnrs.fr

Institute of Applied Physics and Computational Mathematics, Old yard

Beijing, 14-17 July, 2014

$$\boxed{\text{object } \boldsymbol{u_o}} \rightarrow \boxed{\begin{matrix}\text{capture}\\ \text{energy}\end{matrix}} \rightarrow \boxed{\begin{matrix}\text{sampling}\\ \text{quantization}\end{matrix}} \rightarrow \boxed{\text{processing}}$$

$$\begin{pmatrix}\text{scene}\\ \text{body}\\ \text{earth}\end{pmatrix} \quad \begin{pmatrix}\text{reflected}\\ \text{or}\\ \text{emitted}\end{pmatrix} \quad \begin{pmatrix}\text{signal}\\ \text{or}\\ \text{image}\end{pmatrix} \qquad \text{data } \boldsymbol{v} \qquad \begin{matrix}\downarrow\\ \text{output } \hat{\boldsymbol{u}}\end{matrix}$$

Mathematical model: $\boldsymbol{v} = \text{Transform}(\boldsymbol{u_o}) \bullet (\text{Perturbations})$

Some transforms: loss of pixels, blur, FT, Radon T., frame T. $(\cdots)$

Processing tasks: $\begin{cases} \hat{\boldsymbol{u}} = \text{recover}(\boldsymbol{u_o}) \\ \hat{\boldsymbol{u}} = \text{objects of interest}(\boldsymbol{u_o}) \end{cases} \quad (\cdots)$

Mathematical tools: PDEs, Statistics, Functional anal., Matrix anal., $(\cdots)$

Example due to R.S.Wilson

$\boldsymbol{u_o}$ (unknown–signal, picture, density map) $\boldsymbol{v}$ (data, degraded) $=$ Transform($\boldsymbol{u_o}$)$\bullet\boldsymbol{n}$ (noise)

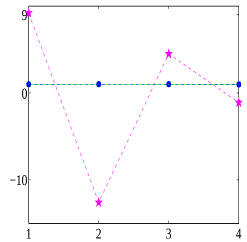
An ill-posed inverse problem

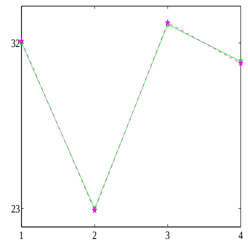
$\boldsymbol{u_o = [\,1 \;\; 1 \;\; 1 \;\; 1\,]^T}$ Transform: $A = \begin{bmatrix} 10 & 7 & 8 & 7 \\ 7 & 5 & 6 & 5 \\ 8 & 6 & 10 & 9 \\ 7 & 5 & 9 & 10 \end{bmatrix}$ $\mathrm{rank}(A) = 4$

- no noise: $\boldsymbol{v = Au_o \;=\; [\,32 \quad 23 \quad 33 \quad 31\,]^T} \;\Rightarrow\; \boldsymbol{\hat{u} = A^{-1}v = u_o}$
- with noise: $\boldsymbol{v = Au_o + n = [\,32.1 \;\; 22.9 \;\; 33.1 \;\; 30.9\,]^T}$

  Least-squares solution: $\boldsymbol{\hat{u} = \underset{u\in\mathbb{R}^4}{\arg\min}\left\{\|Au - v\|^2\right\} = A^{-1}v}$

  $$\Rightarrow \quad \boldsymbol{\hat{u} = [\,9.2 \quad -12.6 \;\; 4.5 \quad -1.1\,]^T}$$





  Tikhonov regularization: $\boldsymbol{\hat{u} = \underset{u\in\mathbb{R}^4}{\arg\min}\, \mathcal{F}_v(u)}$

$$\boldsymbol{\mathcal{F}_v(u) \stackrel{\mathrm{def}}{=} \|Au - v\|^2 + \beta \sum_{i=1}^{3} \big(u[i+1] - u[i]\big)^2}$$

$\beta = 1 \quad \Rightarrow \quad \boldsymbol{\hat{u} = [\;1 \quad 1.01 \quad 1.02 \quad 0.98\,]^T}$

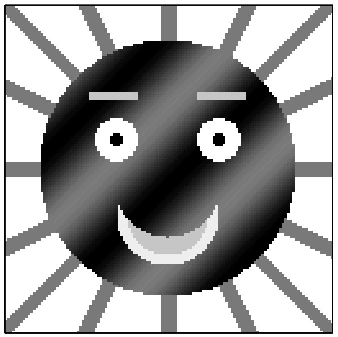
Image/signal processing tasks often require to solve **ill-posed inverse problems**

Out-of-focus picture: $v = a * u_o + \text{noise} = Au_o + \text{noise}$

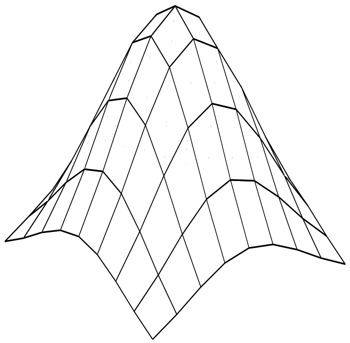
A is ill-conditioned $\equiv$ (nearly) noninvertible

Least-squares solution: $\hat{u} = \arg\min_{u} \left\{ \|Au - v\|^2 \right\}$

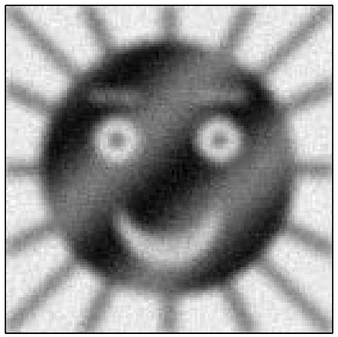
Tikhonov regularization: $\hat{u} := \arg\min_{u} \left\{ \|Au - v\|^2 + \beta \sum_i \|G_i u\|^2 \right\}$ for $\{G_i\} \approx \nabla$, $\beta > 0$



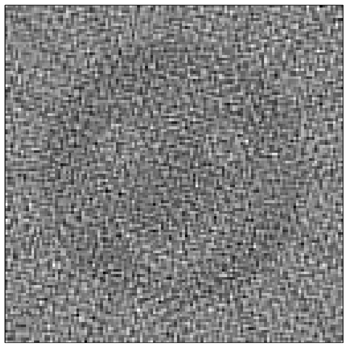
Original u_o



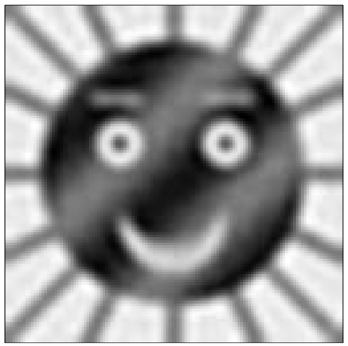
Blur a



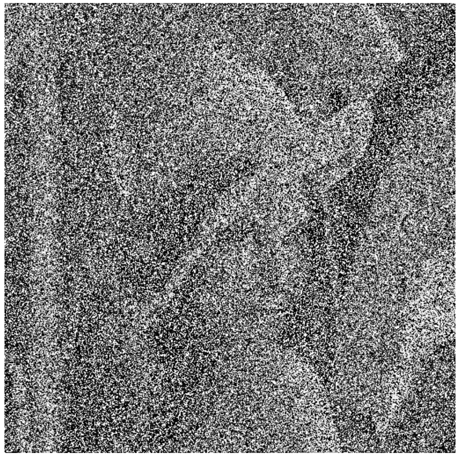
Data v

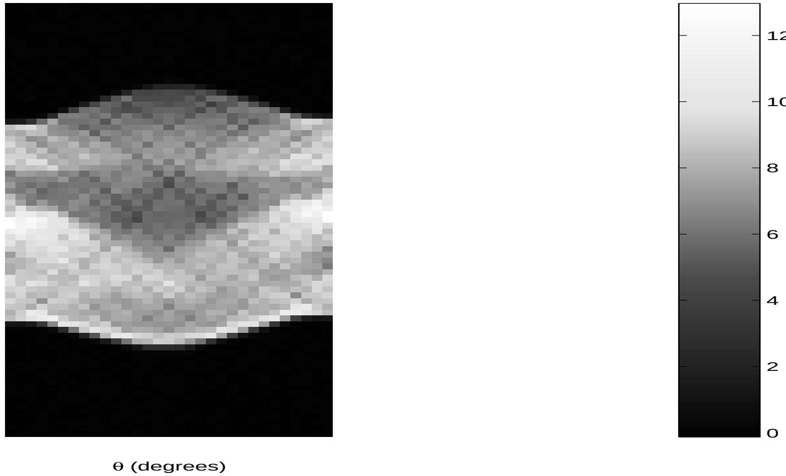


$\hat{u}$: Least-squares



$\hat{u}$: Tikhonov





Impulse noise | Jitter (video) | Radon (tomography)

$\Downarrow$ $\Downarrow$ $\Downarrow$

Formulate your problem as the minimization (maximization) of a functional (an energy) whose solution is the sought after signal/image

Goal of this tutorial: **How to choose your energy $\mathcal{F}_v$?**

Approach: Salient features of the minimizers of classes of energies $\mathcal{F}_v$

Outline

1. Energy minimization methods

$\boldsymbol{u_o}$ (unknown) $\qquad$ $\boldsymbol{v}$ (data) $=$ Transform($\boldsymbol{u_o}$)$\bullet$ (Perturbations)

solution $\hat{\boldsymbol{u}}$

$\hat{\boldsymbol{u}}$ $\nearrow$ close to data production model $\boldsymbol{\Psi(u,v)}$ (data-fidelity)

$\searrow$ coherent with priors and desiderata $\boldsymbol{\Phi(u)}$ (prior – functional, constraint)

Combining models:
$$\hat{\boldsymbol{u}} \quad := \quad \arg\min_{\boldsymbol{u}\in\boldsymbol{\Omega}} \mathcal{F}_v(\boldsymbol{u}) \qquad\qquad (\mathcal{P})$$

$$\mathcal{F}_v(\boldsymbol{u}) \quad := \quad \boldsymbol{\Psi(u,v)} + \boldsymbol{\beta\Phi(u)}, \qquad \boldsymbol{\beta > 0}$$

How to choose ($\mathcal{P}$) to get a good $\hat{\boldsymbol{u}}$?

Applications: Denoising, Segmentation, Deblurring, Tomography, Seismic imaging, Zoom, Superresolution, Compression, Learning, Motion estimation, Pattern recognition $(\cdots)$

The $m \times n$ image u is stored in a $p = mn$-length vector, $\boldsymbol{u} \in \mathbb{R}^{\boldsymbol{p}}$, data $\boldsymbol{v} \in \mathbb{R}^{\boldsymbol{q}}$

$\boldsymbol{\Psi}$ usually models the production of data $\boldsymbol{v}$ $\quad\Rightarrow\quad$ $\Psi = -$ log(Likelihood $(\boldsymbol{v}|\boldsymbol{u})$)

$v = Au_o + n$ for n white Gaussian noise $\quad\Rightarrow\quad$ $\boxed{\boldsymbol{\Psi(u,v) \propto \|Au - v\|_2^2}}$

The information on $\boldsymbol{u}$ we have is implicitly contained in Ψ. It is scarcely enough.

A good prior Φ is needed to solve our task.

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

- Bayesian approach
- Variational approach

Both approaches lead to similar energies

Prior via regularization term $\quad \boldsymbol{\Phi(u) = \sum_i \varphi(\|G_i u\|)}$

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

$\boldsymbol{\{G_i\}}$ — linear operators. Examples : Id, ∇, ∇^2, $\nabla\widetilde{W}$ for $\widetilde{W}$ left inverse of a frame if $u = W$(image)

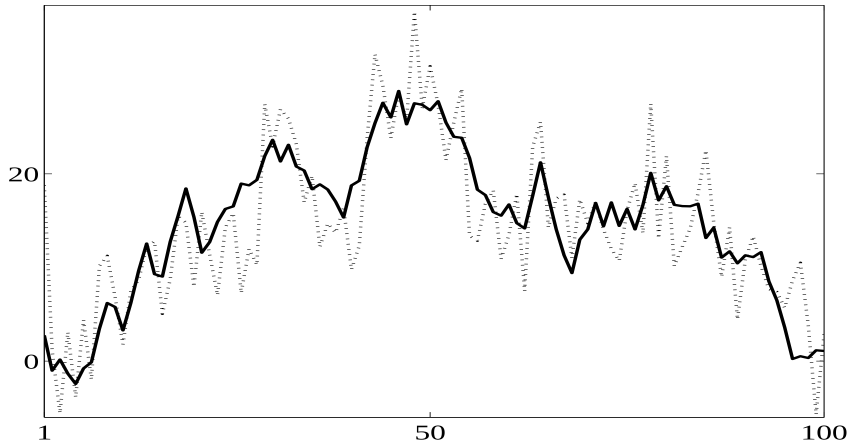
Bayes: U, V random variables, Likelihood $f_{V|U}(v|u)$, Prior $f_U(u) \propto \exp\{-\lambda\Phi(u)\}$

Maximum a Posteriori (MAP) yields the most likely solution $\hat{\boldsymbol{u}}$ given the data $\boldsymbol{V} = \boldsymbol{v}$:

$$\hat{u} = \arg\max_u f_{U|V}(u|v) \quad = \quad \arg\min_u \left(-\ln f_{V|U}(v|u) - \ln f_U(u) \right)$$
$$= \quad \arg\min_u \left(\Psi(u,v) \ + \ \beta\Phi(u) \right) \ = \ \arg\min_u \mathcal{F}_v(u)$$

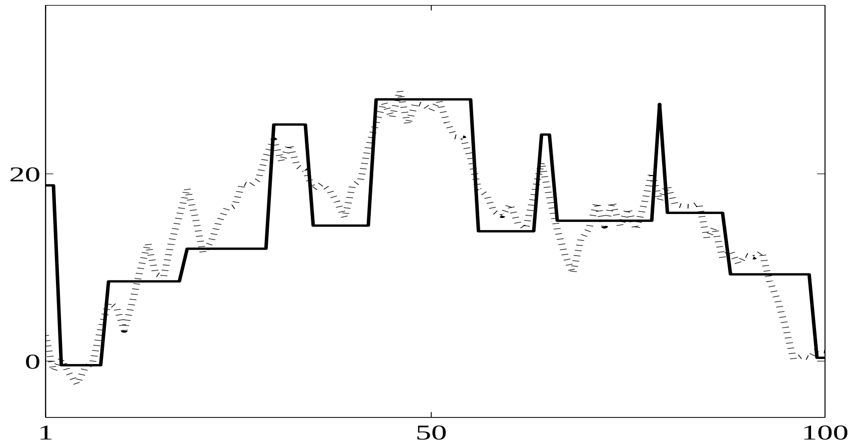
MAP is a very usual way to combine models on data-acquisition and priors

Realist models for data-acquisition $\boldsymbol{f_{V|U}}$ and prior $\boldsymbol{f_U}$ is still an open question



Original $u_o \sim f_U$ (—)

Data $v = u_o$+noise $(\cdots)$, noise $\sim f_{V|U}$



The true MAP $\hat{u}$ (—)

The original u_o $(\cdots)$

`Are you satisfied with the solution?`

- **Minimizer approach** (the core of our tutorials)

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

  Strong results.

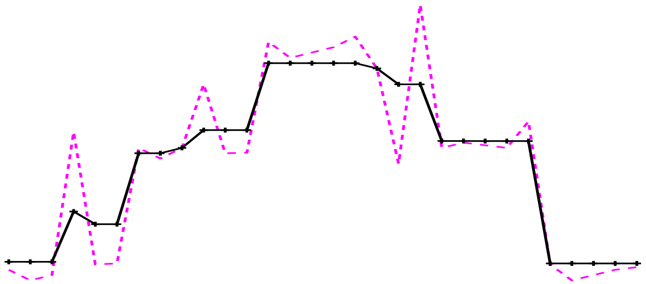
  $\boxed{\textbf{Rigorous tools for modelling}}$

  – Conceive $\mathcal{F}_v$ so that the properties of $\hat{\boldsymbol{u}}$ satisfy your requirements.

(a "chicken and egg" problem?)

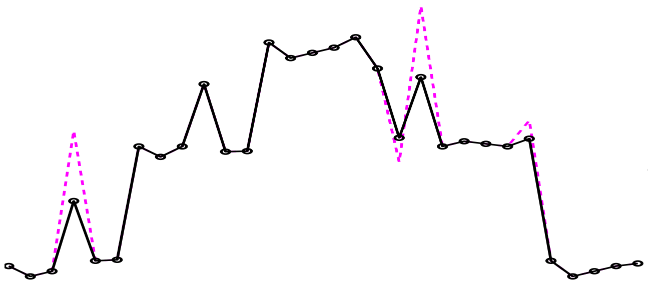
"There is nothing quite as practical as a good theory." *Kurt Lewin*

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



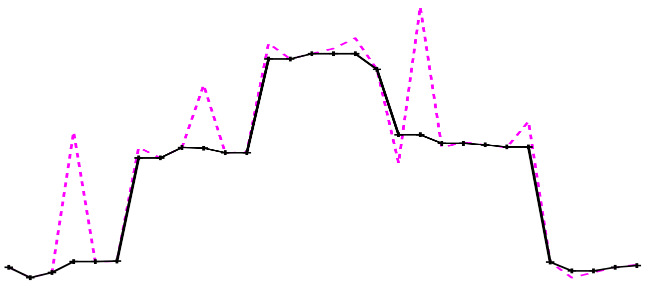
STAIR-CASING

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



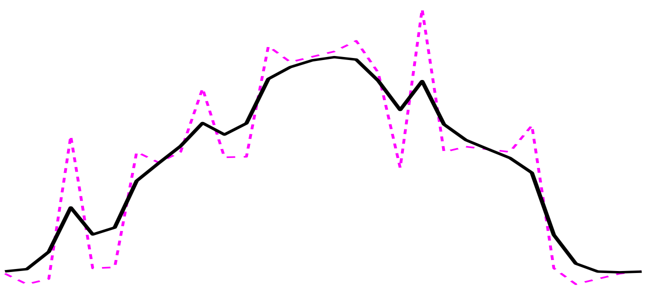
EXACT DATA-FIT

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



BOTH EFFECTS

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



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

Data (— — —), Minimizer (—)

We shall explain why and how to use

Some energy functions

Regularization [Tikhonov, Arsenin 77]: $\mathcal{F}_v(u) = \|Au - v\|^2 + \beta\|Gu\|^2$, $G = I$ or $G \approx \nabla$

Focus on edges, contours, segmentation, labeling

Statistical framework

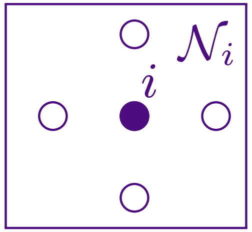
Potts model [Potts 52] (ℓ_0 semi-norm applied to differences):

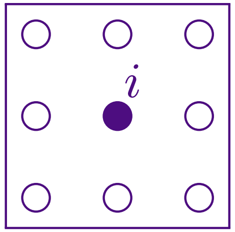
$$\mathcal{F}_v(u) = \Psi(u,v) + \beta \sum_{i,j} \phi(u[i] - u[j]) \quad \phi(t) := \begin{cases} 0 & \text{if} \quad t = 0 \\ 1 & \text{if} \quad t \neq 0 \end{cases}$$

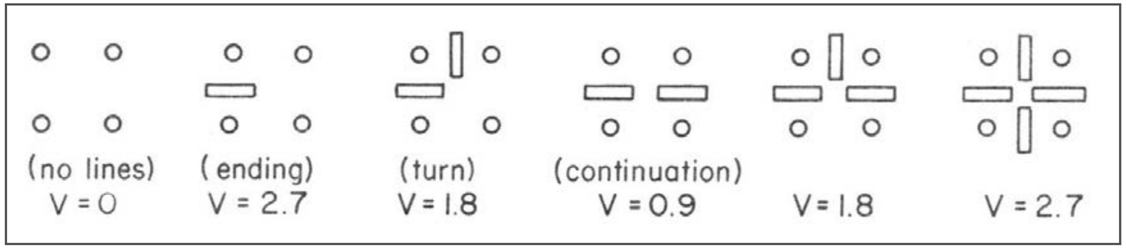
Line process in Markov random field priors [Geman, Geman 84]: $(\hat{u}, \hat{\ell}) = \arg\min_{u,\ell} \mathcal{F}_v(u,\ell)$

$$\mathcal{F}_v(u,\ell) = \Psi(u,v) + \beta \sum_i \Big(\sum_{j \in \mathcal{N}_i} \varphi(u[i] - u[j])(1 - \ell_{i,j}) + \sum_{(k,n) \in \mathcal{N}_{i,j}} \mathrm{V}(\ell_{i,j}, \ell_{k,n}) \Big)$$

$[\ell_{i,j} = 0 \Leftrightarrow$ no edge$]$, $[\ell_{i,j} = 1 \Leftrightarrow$ edge between i and $j]$, $\varphi(t) = 1$







GEMAN AND GEMAN: STOCHASTIC RELAXATION, GIBBS DISTRIBUTIONS, AND BAYESIAN RESTORATION

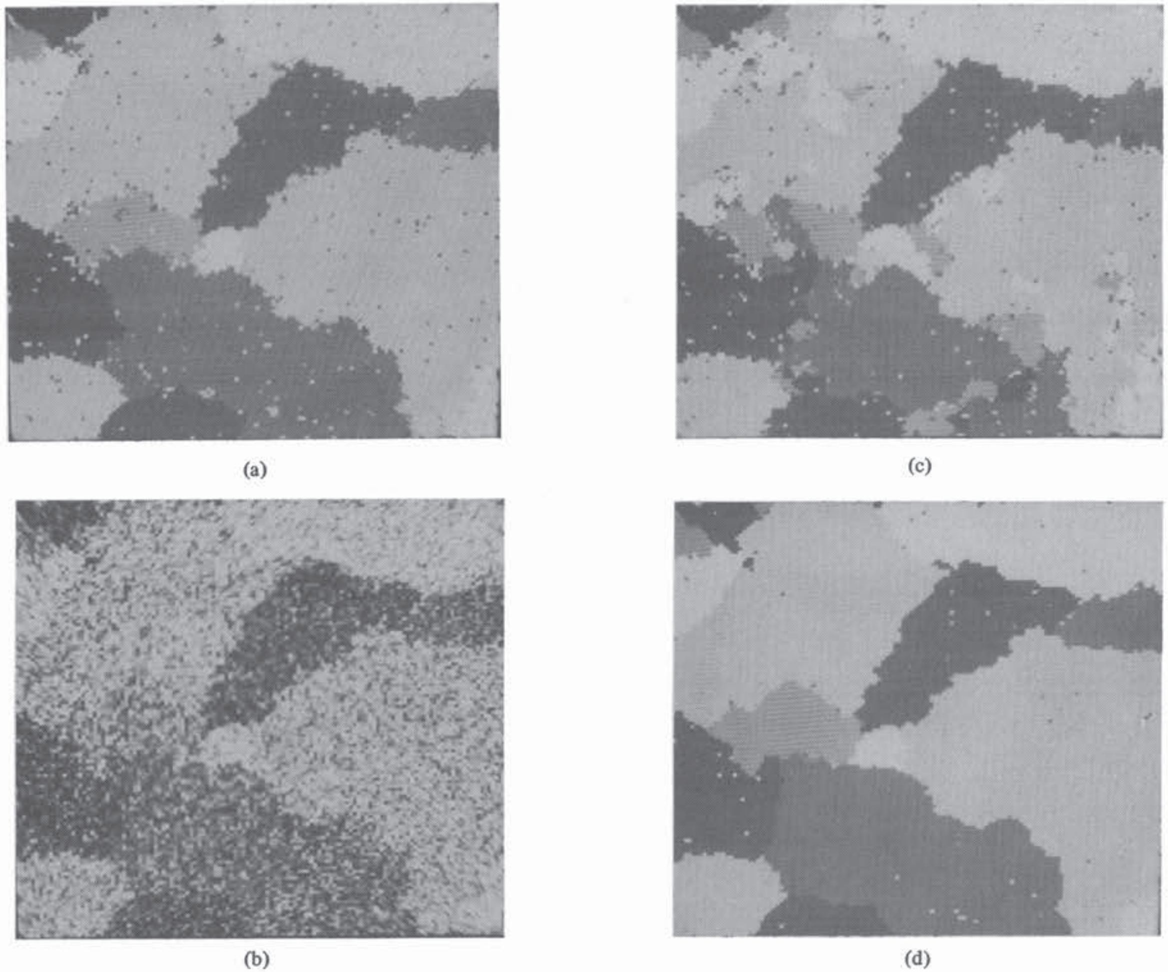


Fig. 2. (a) Original image: Sample from MRF. (b) Degraded image: Additive noise. (c) Restoration: 25 iterations. (d) Restoration: 300 iterations.

Image credits: S. Geman and D. Geman 1984. Restoration with 5 labels using Gibbs sampler

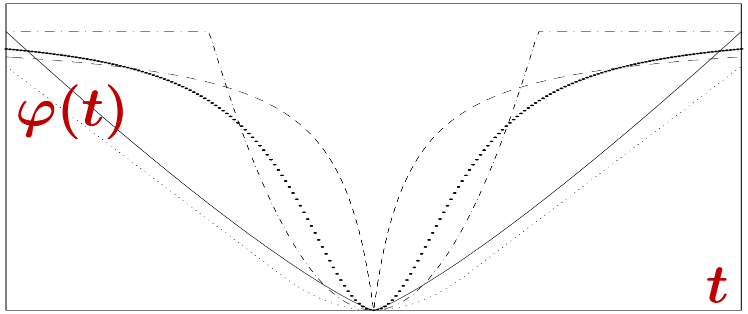
"We make an analogy between images and statistical mechanics systems. Pixel gray levels and the presence and orientation of edges are viewed as states of atoms or molecules in a lattice-like physical system. The assignment of an energy function in the physical system determines its Gibbs distribution. Because of the Gibbs distribution, Markov random field (MRF) equivalence, this assignment also determines an MRF image model." [S. Geman, D. Geman 84]

M.-S. functional [Mumford, Shah 89]: $\mathcal{F}_v(u, L) = \int_\Omega (u-v)^2 dx + \beta \left(\int_{\Omega \setminus L} \|\nabla u\|^2 dx + \alpha |L| \right)$

discrete version: $\Phi(u) = \sum_i \varphi(\|G_i u\|)$, $\varphi(t) = \min\{t^2, \alpha\}$, $\{G_i\} \approx \nabla$

Total Variation (TV) [Rudin, Osher, Fatemi 92]: $\mathcal{F}_v(u) = \|u-v\|_2^2 + \beta\, \mathrm{TV}(u)$

$$\mathrm{TV}(u) = \int \|\nabla u\|_2\, dx \approx \sum_i \|G_i u\|_2$$



Various edge-preserving functions φ to define Φ

φ is edge-preserving if $\lim_{t \to \infty} \frac{\varphi'(t)}{t} = 0$

[Charbonnier, Blanc-Féraud, Aubert, Barlaud 97 ...]

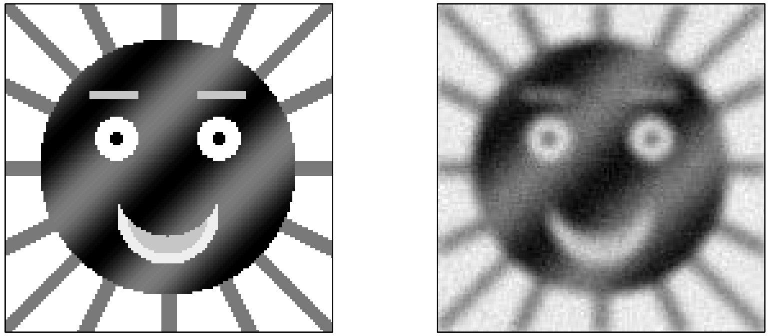
Minimizer approach

ℓ_1- Data fidelity [Nikolova 02]: $\mathcal{F}_v(u) = \|Au - v\|_1 + \beta\Phi(u)$

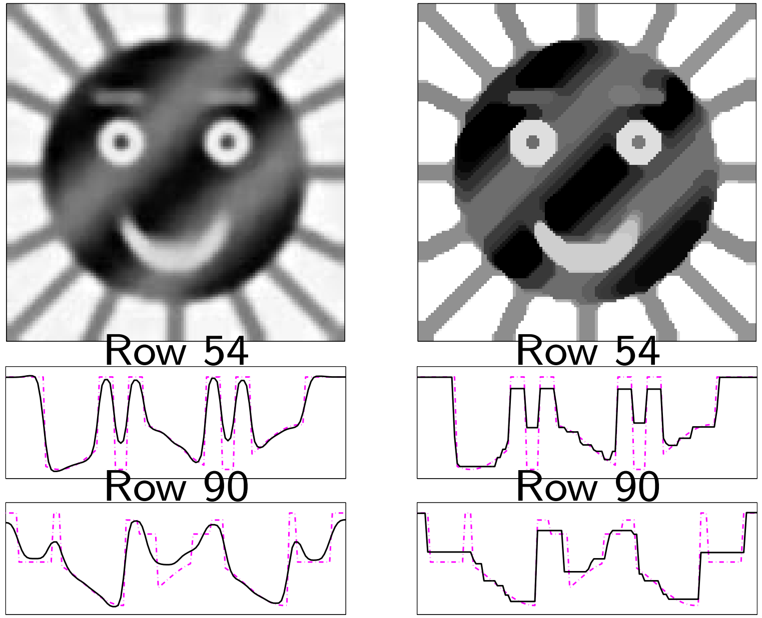
$L_1 - \mathrm{TV}$ model [T. Chan, Esedoglu 05]: $\mathcal{F}_v(u) = \|u-v\|_1 + \beta\, \mathrm{TV}(u)$

CPU time ! Computers ↑↑

Original u_o Data $v = a * u_o + n$



$\varphi(t) = |t|^{\alpha \in (1,2)}$ $\varphi(t) = |t|$

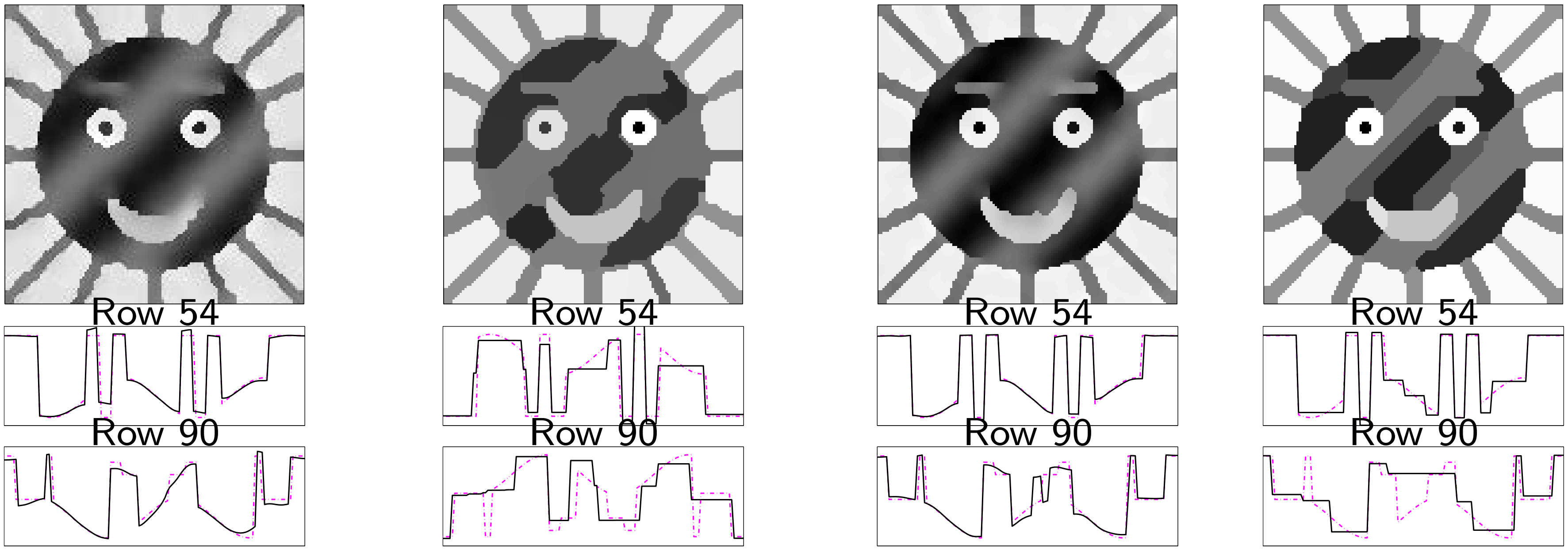


φ convex

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

φ smooth at 0   **φ nonsmooth at 0**

$\varphi(t) = \alpha t^2/(1+\alpha t^2)$ $\varphi(t) = \alpha|t|/(1+\alpha|t|)$ $\varphi(t) = \min\{\alpha t^2, 1\}$ $\varphi(t) = 1 - 1\!\!1_{(t=0)}$



nonconvex

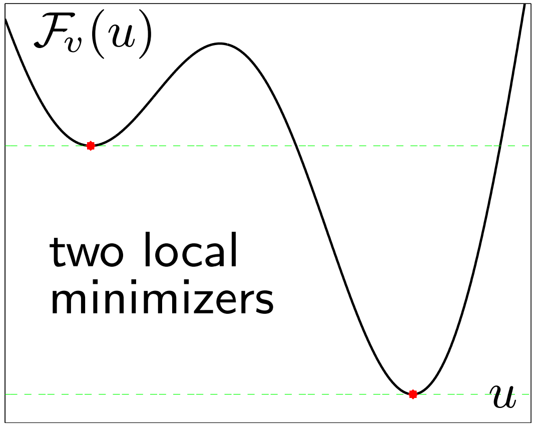
Summer School 2014: **Inverse Problem and Image Processing**

Tutorial: **Inverse modeling in inverse problems using optimization**

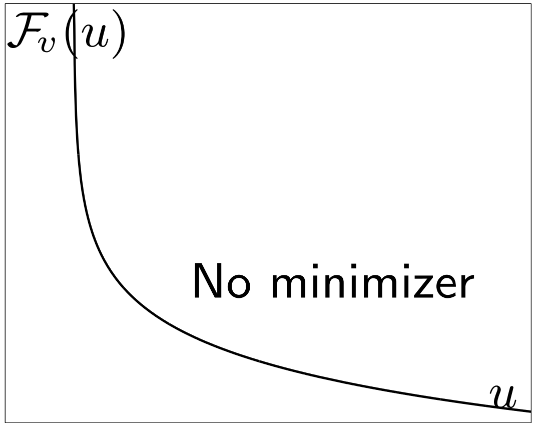
Outline

2 Regularity Results

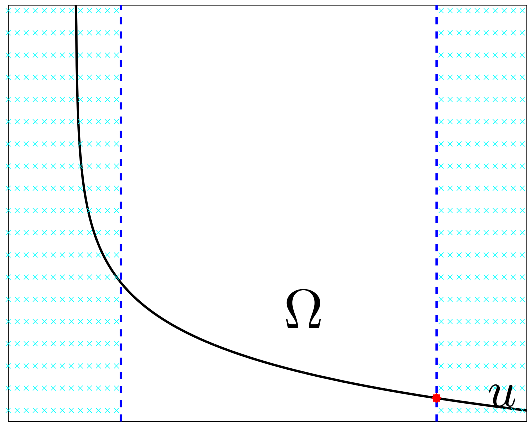
Optimization problems



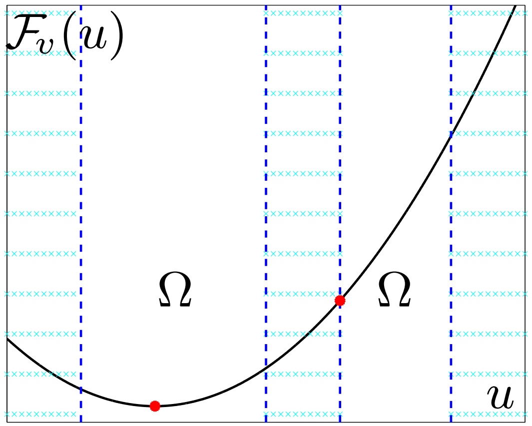
$\mathcal{F}_v$ nonconvex



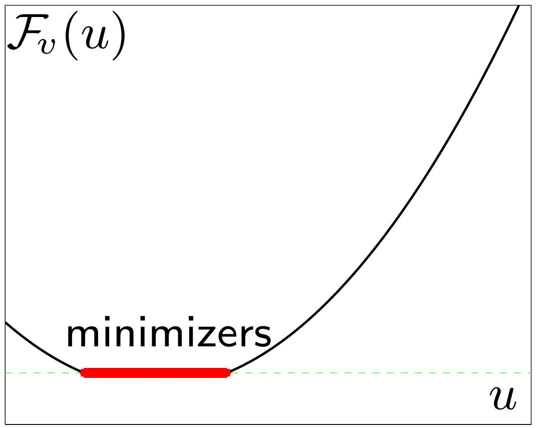
$\mathcal{F}_v$ convex non coercive $\Omega = \mathbb{R}$



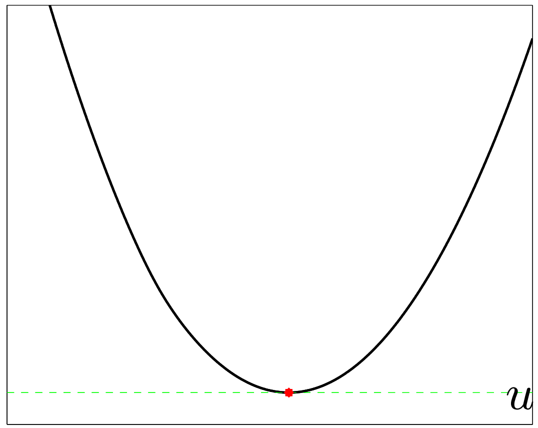
$\mathcal{F}_v$ convex non coercive Ω compact



$\mathcal{F}_v$ strictly convex, Ω nonconvex



$\mathcal{F}_v$ non strictly convex



$\mathcal{F}_v$ strictly convex on $\mathbb{R}$

$$\mathcal{F}_v : \Omega \to \mathbb{R} \qquad \Omega \subset \mathbb{R}^p$$

- Set of optimal solutions $\widehat{U} = \{\hat{u} \in \Omega \; : \; \mathcal{F}_v(\hat{u}) \leqslant \mathcal{F}_v(u) \quad \forall \, u \in \Omega\}$

  $\widehat{U} = \{u\}$ if $\mathcal{F}_v$ strictly convex

  $\widehat{U} \neq \varnothing$ if $\mathcal{F}_v$ coercive of if $\mathcal{F}_v$ continuous and Ω compact

  Otherwise – check

  (e.g. see if $\mathcal{F}_v$ is asymptotically level stable [Auslender, Teboulle 03])

- Nonconvex problems:

  Algorithms may get trapped in local minima

  A "good" local minimizer can be satisfying

  Global optimization – difficult, but progress, e.g. [Robini, Reissman JGO 13]

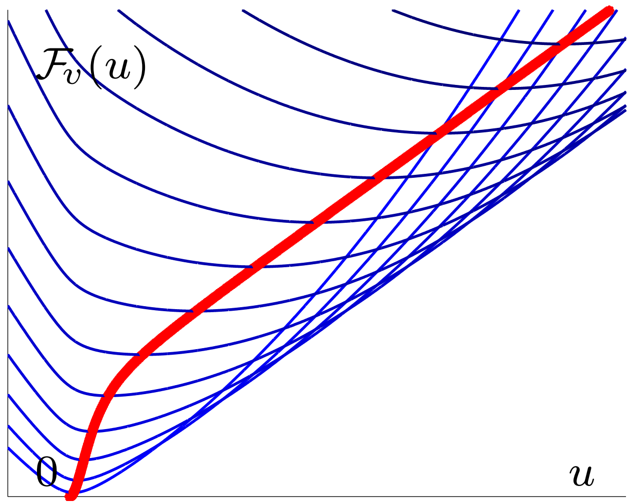
  Convex relaxation methods, see, e.g., [Yuan, Bae, Tai CVPR 10]

- Attention to numerical errors

Definition: $\boldsymbol{\mathcal{U}} : \boldsymbol{O} \to \mathbb{R}^{\boldsymbol{p}}$, $\;O \subset \mathbb{R}^q$ open, is a (local) minimizer function for

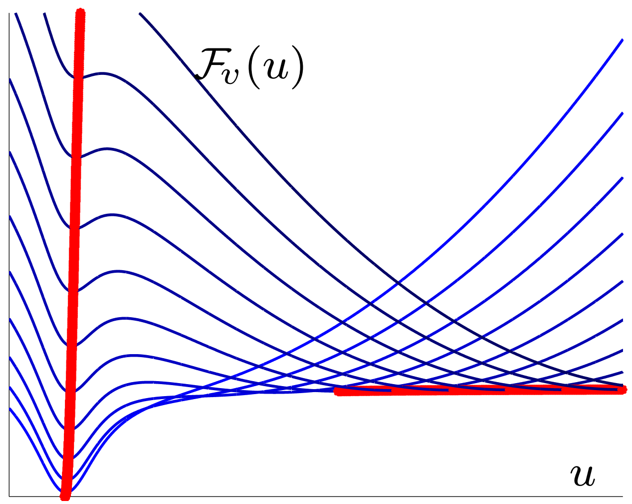
$$\boldsymbol{\mathcal{F}_O} := \{\boldsymbol{\mathcal{F}_v} \;:\; \boldsymbol{v} \in \boldsymbol{O}\} \text{ if } \boldsymbol{\mathcal{F}_v} \text{ has a strict (local) minimum at } \boldsymbol{\mathcal{U}}(\boldsymbol{v}),\ \forall\ \boldsymbol{v} \in \boldsymbol{O}$$

Minimizer functions – an useful tool to analyze the properties of minimizers...



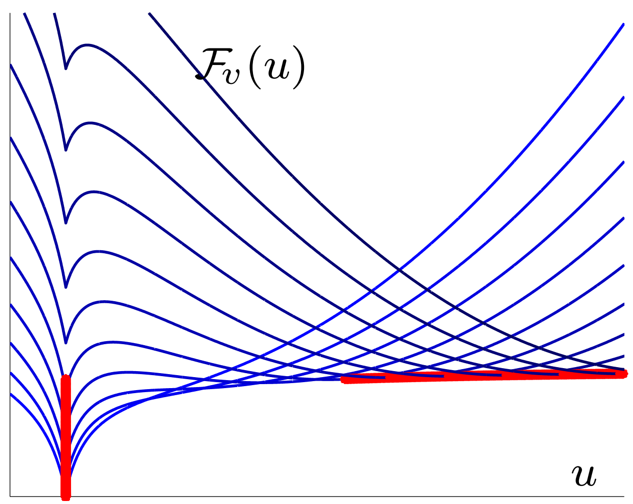
$\mathcal{F}_v(u) = (u-v)^2 + \beta\sqrt{\alpha + u^2}$

minimizer function ( )



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

local minimizer functions ( )



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

global minimizer function ( )

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

Question **1** What these plots reveal about the local / global minimizer functions?

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

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

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

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

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

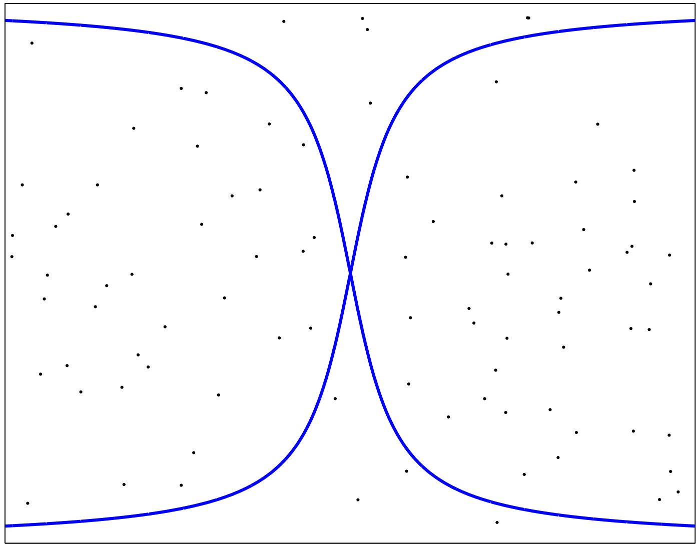
$$\varphi'(0^+) > 0 \quad \Rightarrow \quad \Phi \text{ is nonsmooth on } \bigcup_i \{u : G_i u = 0\}$$

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

Recall:

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

$\mathcal{F}_v$ 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 **2** `What is the chance that` $(x,y) \in N$?

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

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

A. Local minimizers

`(knowing local minimizers is important)`

There is a closed $\boldsymbol{N} \subset \mathbb{R}^q$ with Lebesgue measure $\mathbb{L}^{\boldsymbol{q}}(\boldsymbol{N}) = \boldsymbol{0}$ 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 **3** `For` $v \in \mathbb{R}^q \setminus N$`, compare` $\mathcal{U}(v)$ `and` $\mathcal{U}(v + \varepsilon)$ `where` $\varepsilon \in \mathbb{R}^q$ `is small enough.`

B. Global minimizers

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

Question **4** `What is the chance that` $v \in \hat{N}$`?  What can happen if` $v \in \hat{N}$`?`

Nonasymptotic bounds on minimizers [Nikolova 07]

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

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

- φ is strictly increasing **<u>or</u>** $\mathrm{rank}(A) = p$

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

- $\|\varphi'\|_\infty =$ constant $\quad$ (φ is edge-preserving) $\;$ and $\mathrm{rank}(A) = q \leqslant p$

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

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

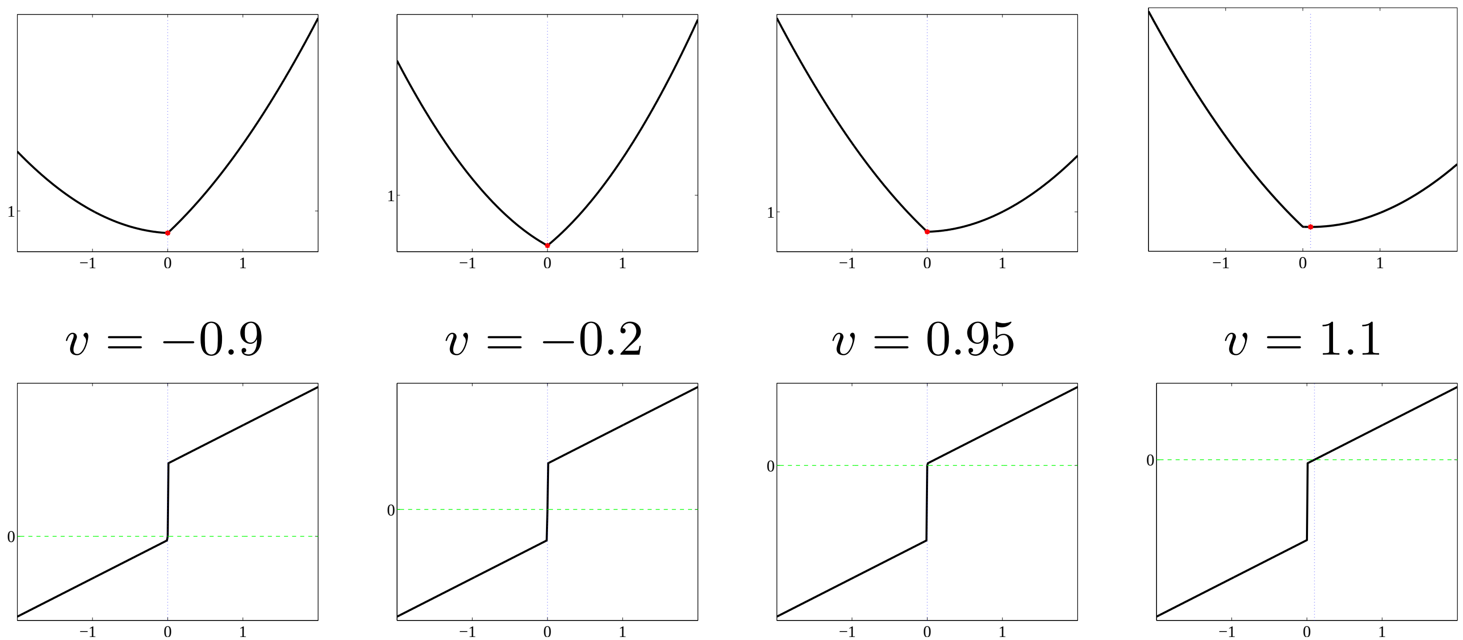
Question **5** `If` $v = u_o + n$ `for` n `Gaussian noise, is it possible to clean` v `from this noise by minimizing` $\mathcal{F}_v$`?` `(See` Ψ `on p. 8.)`

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 **6** `Comment the minimizers on the 1st row. What is drawn on the 2nd row?`

Summer School 2014: **Inverse Problem and Image Processing**

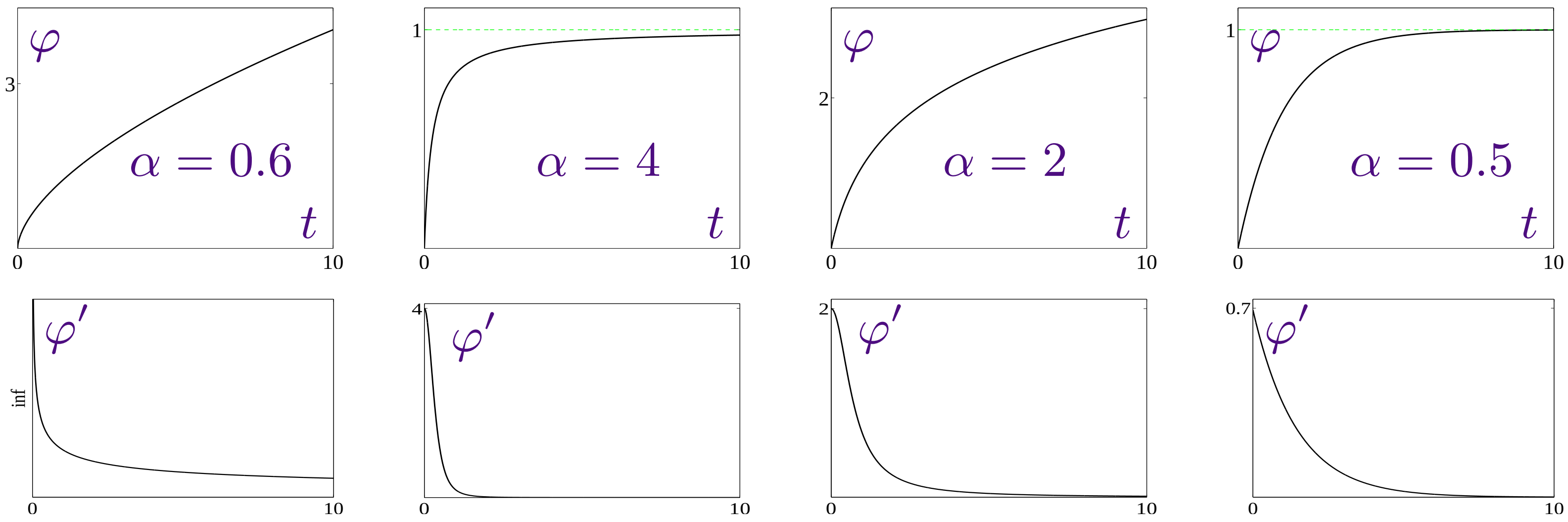
Tutorial: **Inverse modeling in inverse problems using optimization**

Outline

3 Minimizers under Non-Smooth Regularization

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

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



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

General case $\mathcal{F}_v(u)=\Psi(u,v)+\beta\sum_{i=1}^{r}\varphi(\|G_iu\|)$ $\Psi\in\mathcal{C}^{m\geqslant 2}, \varphi'(0^+)>0$ [Nikolova 97,00]

Let $\hat{u}$ be a (local) minimizer of $\mathcal{F}_v$. Set $\hat{h}:=\{i : G_i\hat{u}=0\}$
Then $\exists$ $O\subset\mathbb{R}^q$ **open,** $\exists\,\mathcal{U}\in\mathcal{C}^{m-1}$ (local) minimizer function so that

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

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

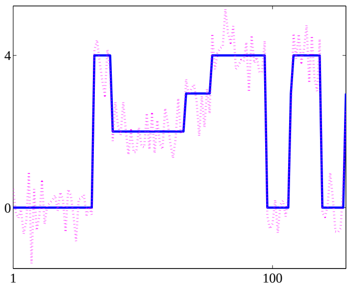
Data v yield (local) minimizers $\hat{u}$ of $\mathcal{F}_v$ such that
$G_i\hat{u}=0$ for a set of indexes $\hat{h}$

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

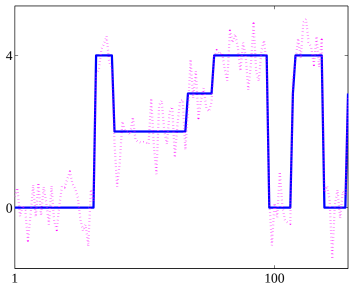
$G_iu=u[i]$ $\Rightarrow$ many samples $\hat{u}[i]=0$ – highly used in Compressed Sensing

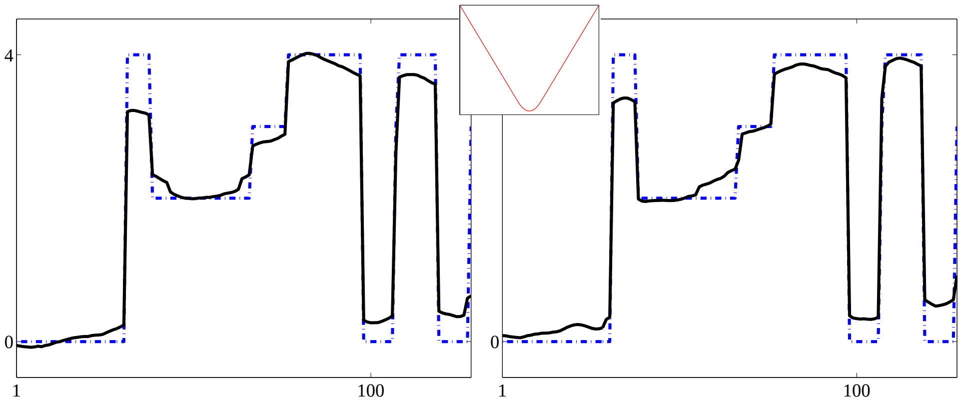
Question **7** `What happens if` $\{G_i\}$ `yield second-order differences?`

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

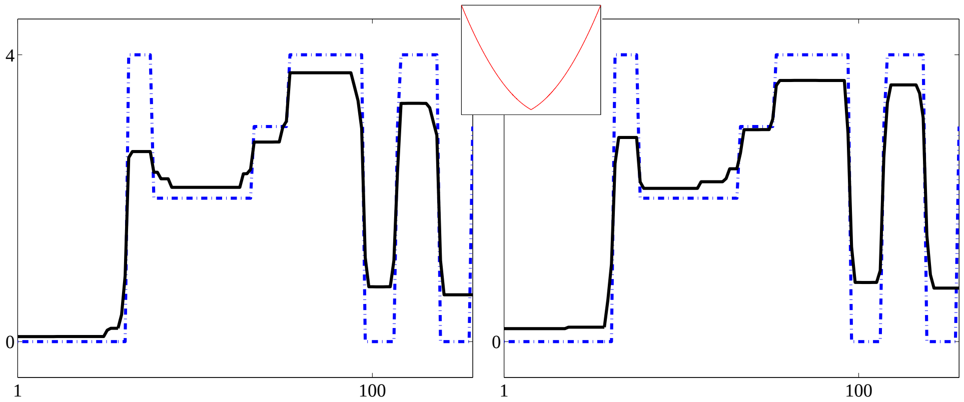


$$\mathcal{F}_v(u) = \|u - v\|^2$$
$$+\beta \sum \varphi(|u[i] - u[i-1]|)$$

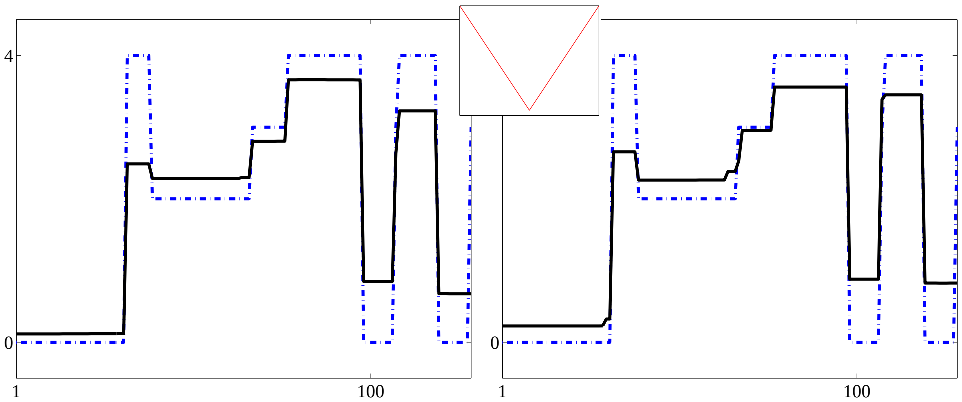




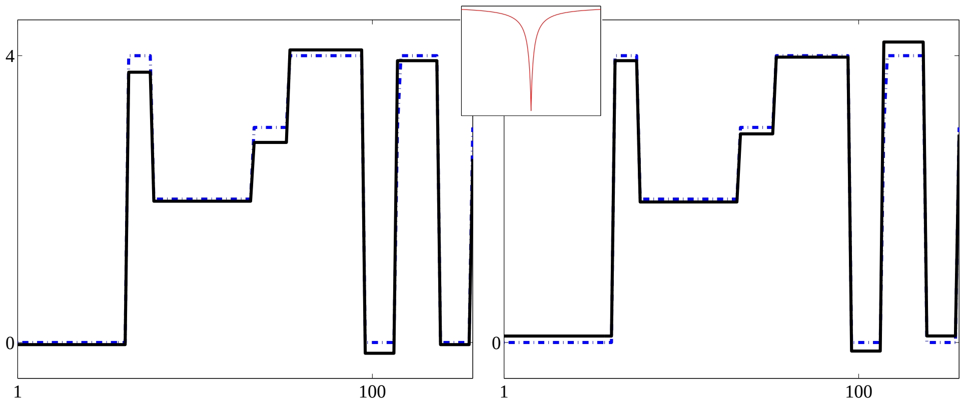
$\varphi(t) = \sqrt{\alpha + t^2}$, $\quad \varphi'(0) = 0$ **(smooth at 0)**



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

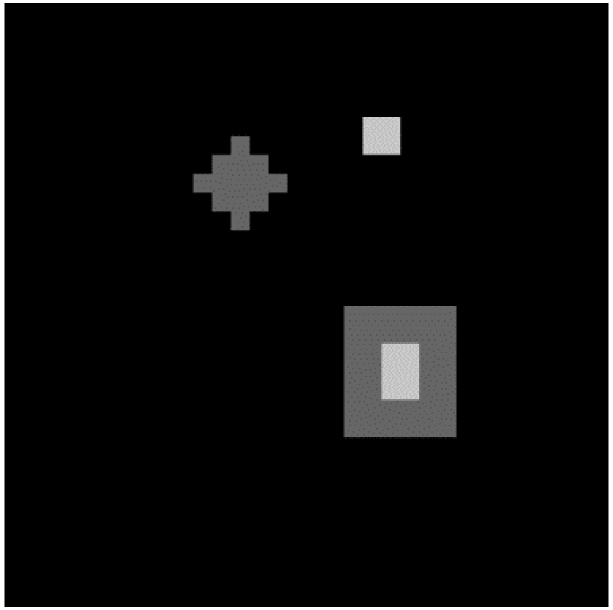


$\varphi(t) = |t|$, $\quad \varphi'(0^+) = 1$

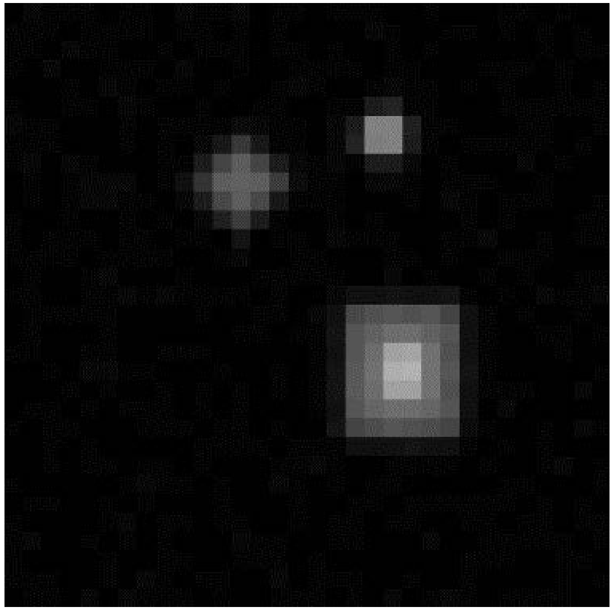


$\varphi(t) = \alpha|t|/(1 + \alpha|t|)$, $\quad \varphi'(0^+) = \alpha$

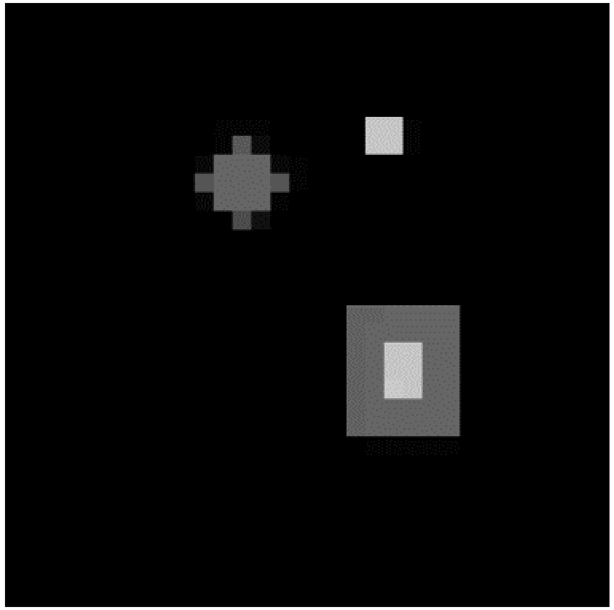
TV energy: $\mathcal{F}_v(u) = \|Au - v\|^2 + \beta \mathrm{TV}(u)$



Original



Data



Restored: TV energy

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

Questions to clarify the main property

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

Question **8** Give $\mathbf{Pr}(u_o = 0)$.

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

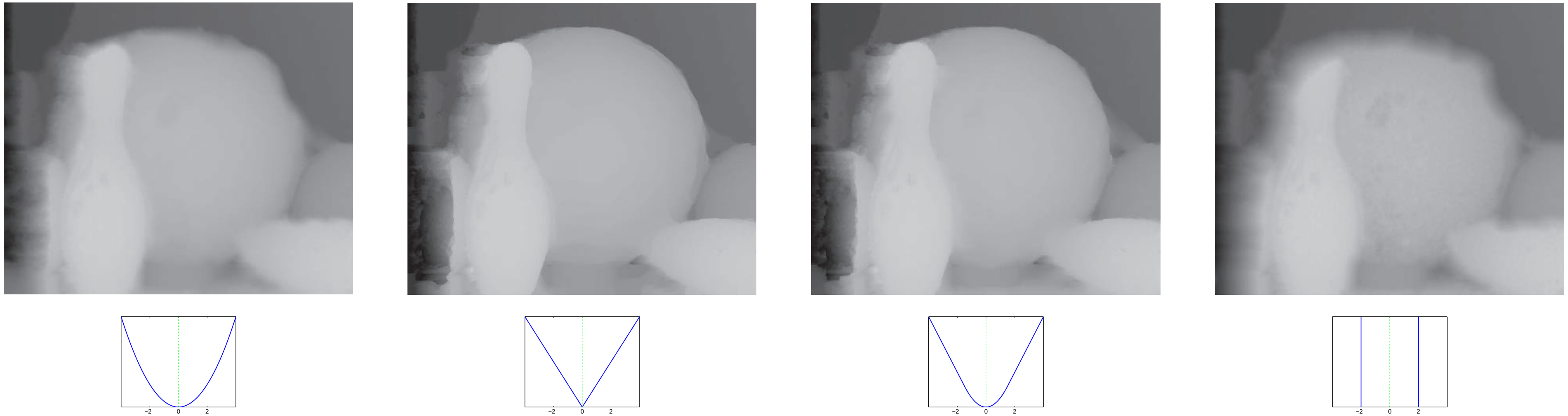
Useful reminder on p. 24.

Question **10** Determine the set $\{\nu \in \mathbb{R} \ : \ \mathcal{U}(\nu) = 0\}$. Comment the result.

Disparity estimation

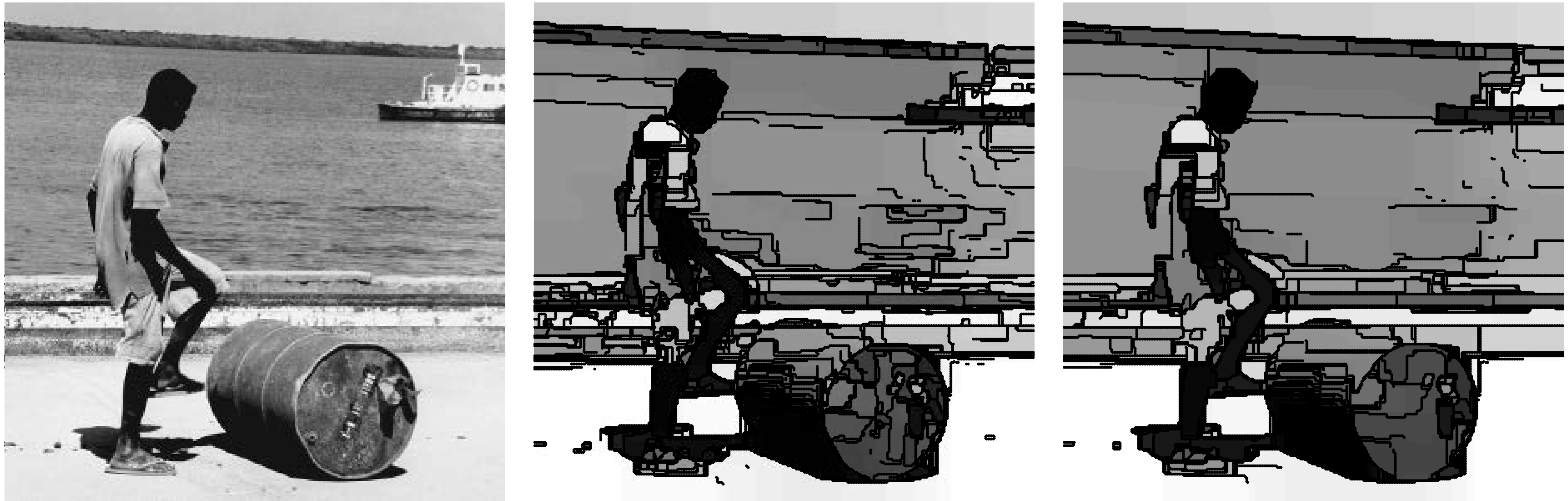
(a) Left input image (b) Right input image (c) True disparity

Figure 7. *Rectified stereo image pair and the ground truth disparity. Light gray pixels indicate structures near to the camera, and black pixels correspond to unknown disparity values.*



quadratic TV Huber Lipscitz

Image credits to the authors: Pock, Cremers, Bischof, and Chambolle "Global Solutions of Variational Models with Convex Regularization", SIIMS 3(4) 2010, pp. 1122-1145



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

Questions relevant to the Potts model (see p. 12)

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

Question **11** Compute the <u>global</u> 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.

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

Note: $\sum_{i=1}^{p} \varphi(u[i]) = \#\{i \ : \ u[i] \neq 0\} = \ell_0(u)$ is the counting norm.

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 **12** Compute each component $\mathcal{U}_i$

Question **13** 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 $i \not\in h$.

Summer School 2014: **Inverse Problem and Image Processing**

Tutorial: **Inverse modeling in inverse problems using optimization**

Outline

4 Minimizers relevant to non-smooth data-fidelity

General case

[Nikolova 02]

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

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

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

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

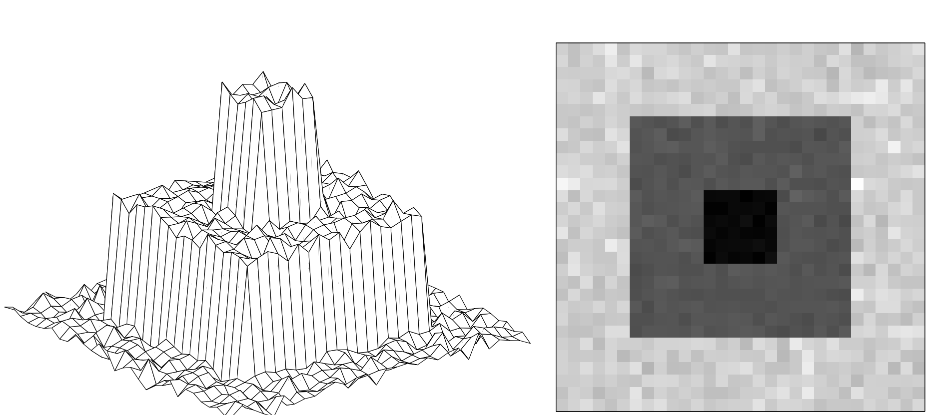
(Local) minimizers $\hat{u}$ of $\mathcal{F}_v$ achieve an **exact fit** to (noisy) data
$a_i\hat{u} = v[i]$ for a certain number of indexes i

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

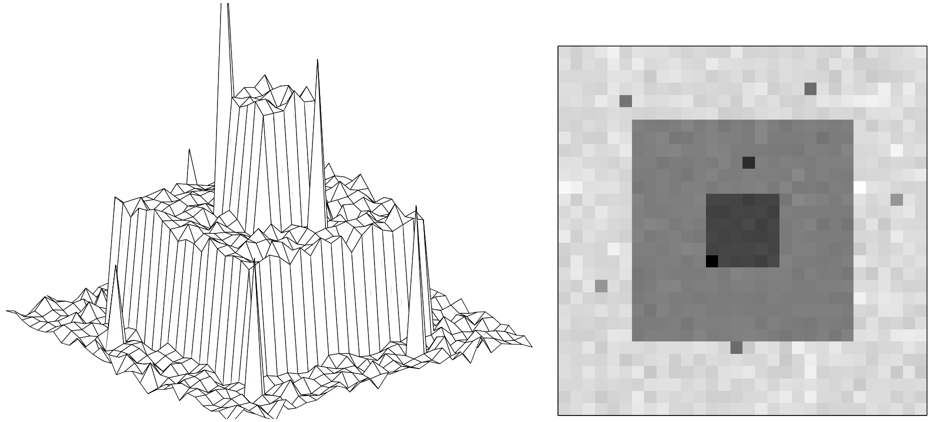
Question **14** Suggest cases when you would like that your minimizer obeys this property.

Question **15** Compute the minimizer of $\mathcal{F}_v(u) = |u - v| + \beta u^2$ for $u, v \in \mathbb{R}$ and $\beta > 0$.

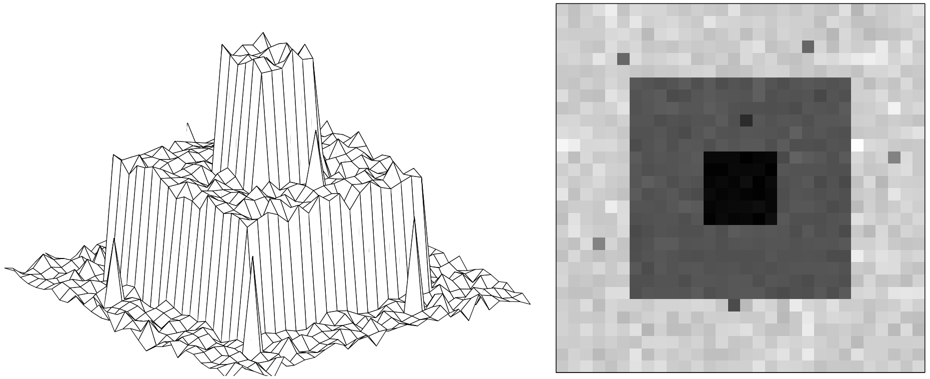
Question **16** Can you find a relationship between the properties of the minimizer when $\varphi'(0^+) > 0$ (chapter 3, p. 26) and when $\psi'(0^+) > 0$ (chapter 4, p. 35)



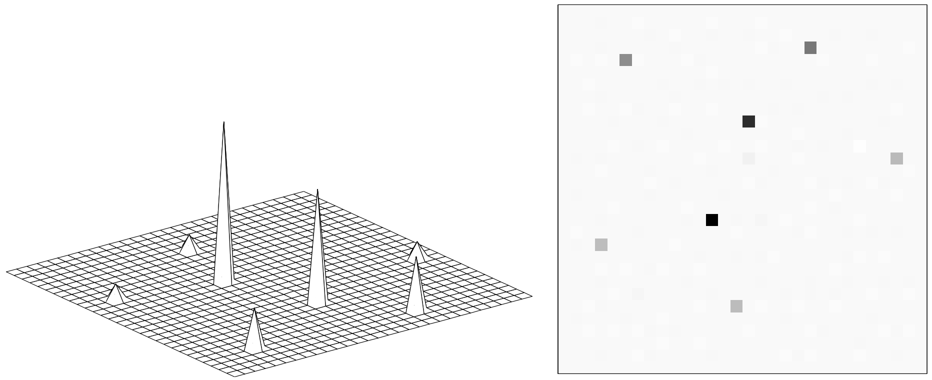
Original u_o



Data $v = u_o$+outliers

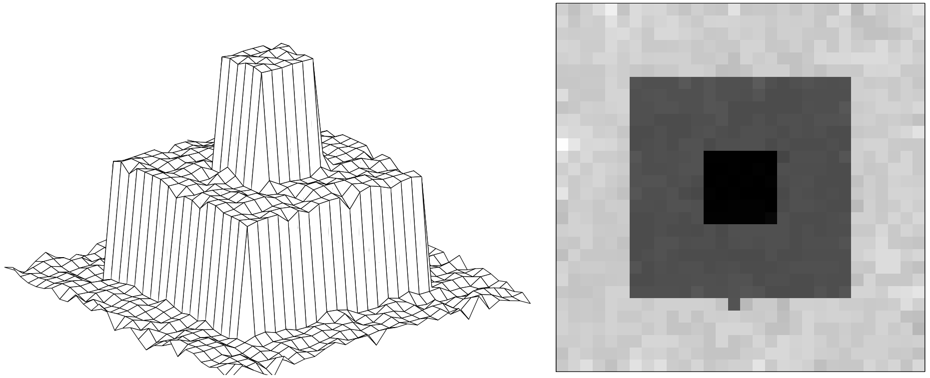


Restoration $\hat{u}$ for $\boldsymbol{\beta = 0.14}$

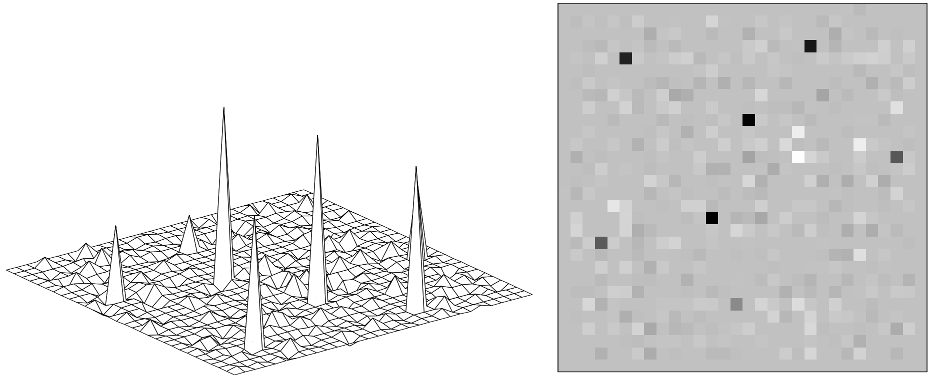


Residuals $v - \hat{u}$

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

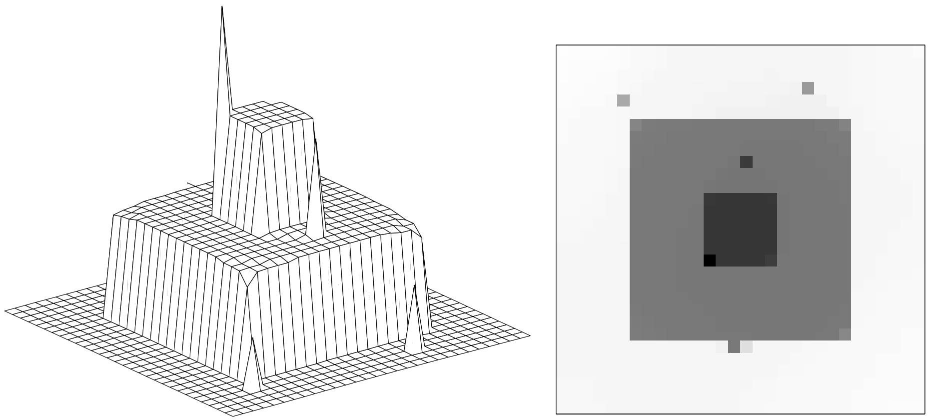


Restoration $\hat{u}$ for $\boldsymbol{\beta = 0.25}$

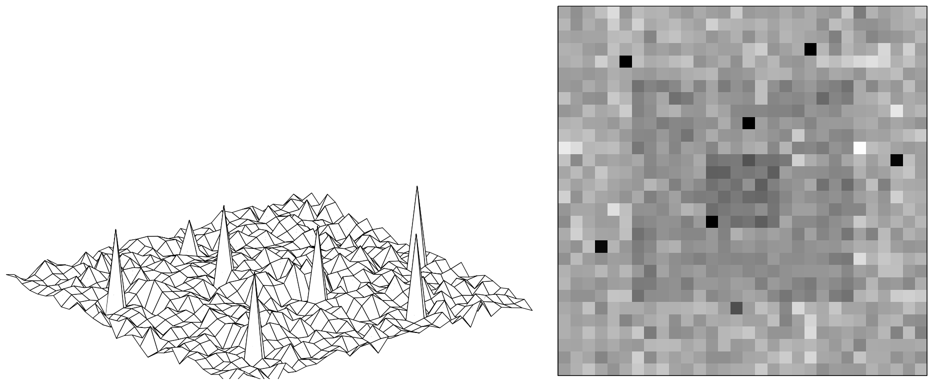


Residuals $v - \hat{u}$

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



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



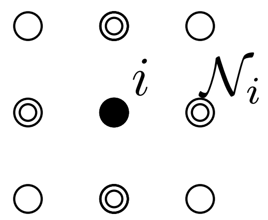
Residuals $v - \hat{u}$

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: $v \to \hat{h}^c(v) = \{i : \hat{u}[i] \neq v[i]\}$

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

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

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

Recovery of frame coefficients using ℓ_1 data-fitting [Durand, Nikolova 07]

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

  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]...

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

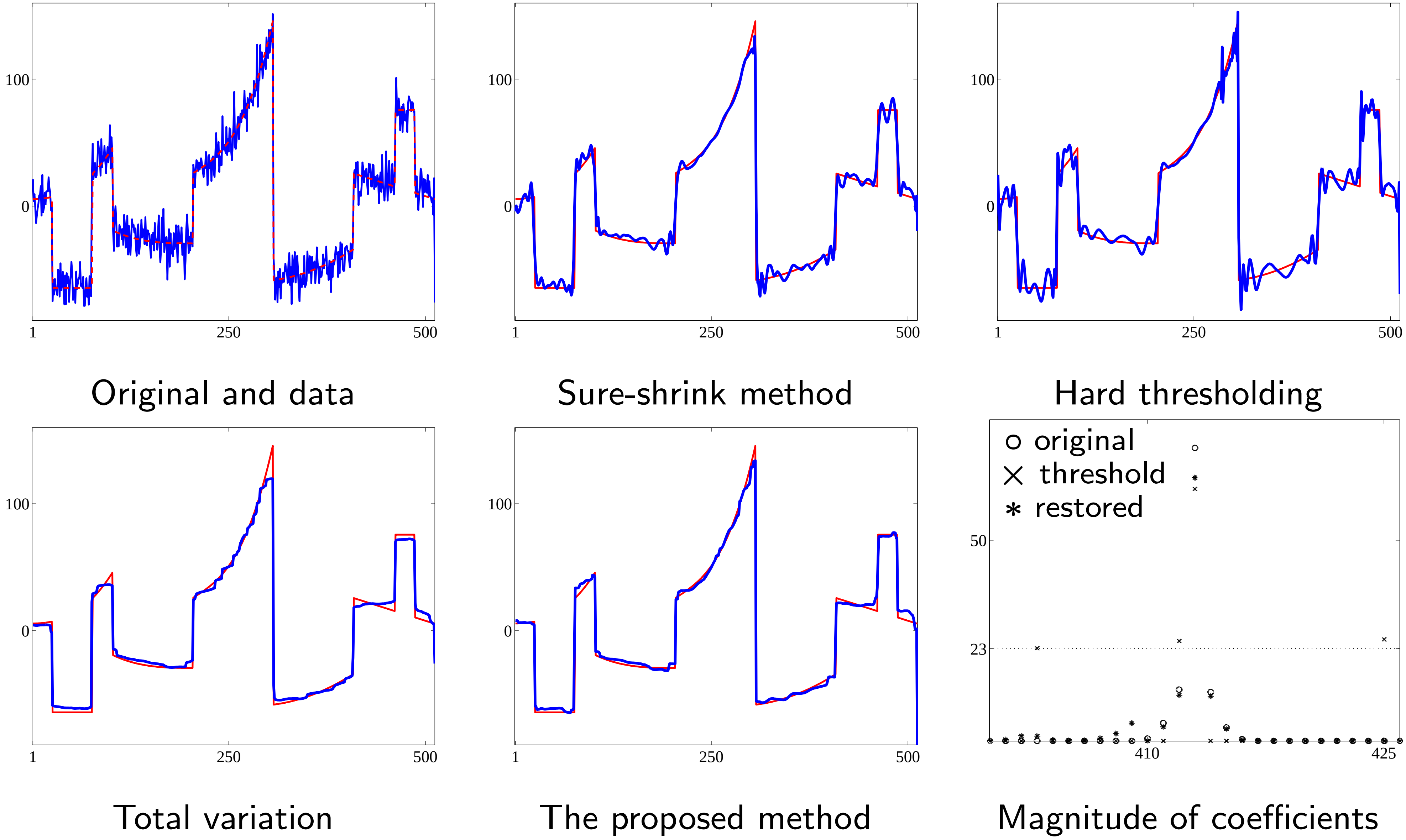
Keep $\hat{x}[i] = y_T[i]$	Restore $\hat{x}[i] \neq y_T[i]$
significant coefs: $y[i] \approx (Wu_o)[i]$	*outliers:* $\|y[i]\| \gg \|(Wu_o)[i]\|$ (frame-shaped artifacts)
thresholded coefs: $(Wu_o)[i] \approx 0$	edge coefs: $\|(Wu_o)[i]\| > \|y_T[i]\| = 0$ (“Gibbs” oscillations)

Then:

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

Question **17** Explain why the minimizers of $\mathcal{F}_y$ fulfill the desiderata.

Hint: “good” coefficients fitted exactly, “bad” coefficients corrected according to the prior.



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

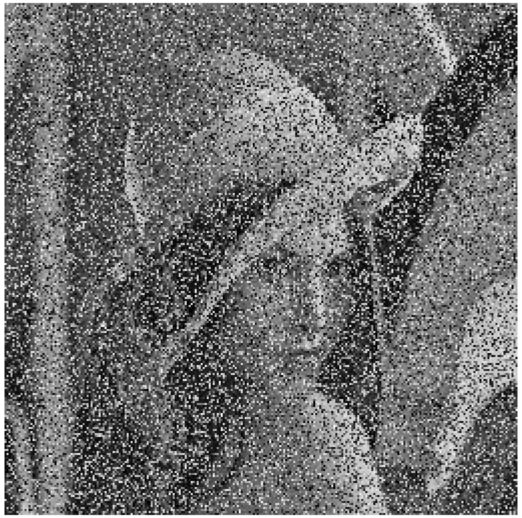
   (e.g. adaptive or center-weighted median)

   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 $i \in \hat{h}$

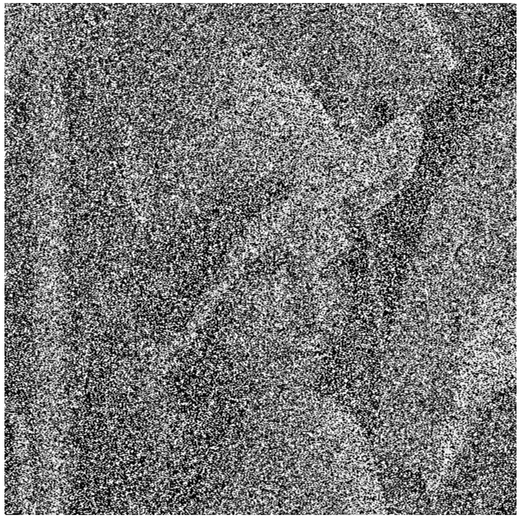


50% RV noise

ACWMF

DPVM

Our method



70 % SP noise(6.7dB)

Adapt.med.(25.8dB)

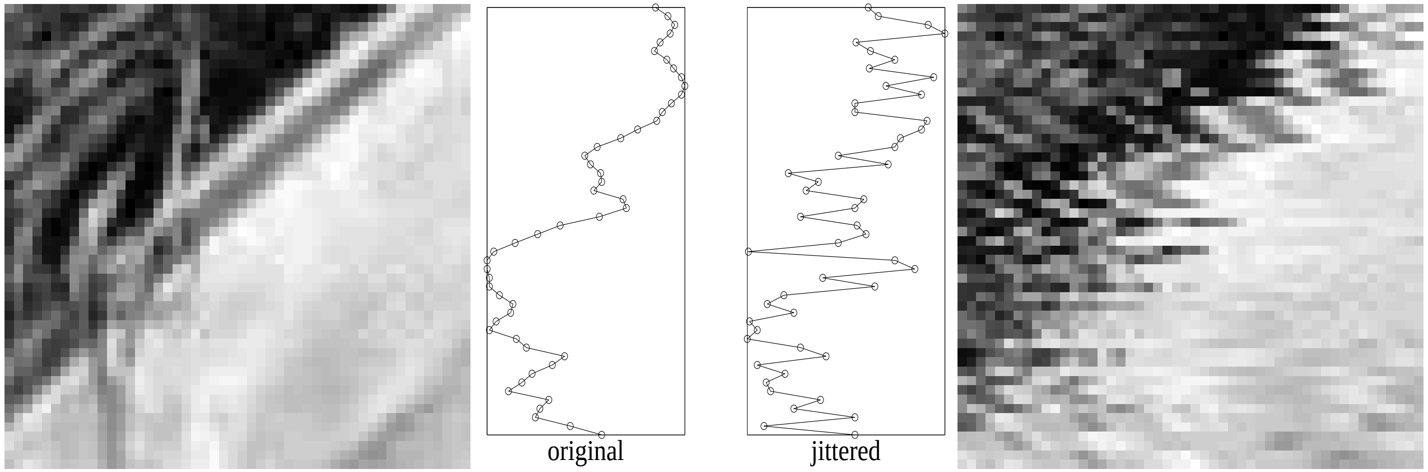
Our method(29.3dB)

Original Lena

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 $\boldsymbol{u_i[j]}$
- Data $\boldsymbol{v_i[j] = u_i[j + d_i]}$, $\boldsymbol{d_i}$ integer, $\boldsymbol{|d_i| \leqslant M}$, typically $M \leqslant 20$.
- Restore $\hat{\boldsymbol{u}} \equiv$ restore $\boldsymbol{\hat{d}_i,\ 1 \leqslant i \leqslant m}$



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 $\hat{u}_i$ 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 **18** Explain why the minimizers of $\mathcal{F}$ can solve the problem as stated.

Question **19** What changes if $\alpha = 1$ or if $\alpha = 0.5$?

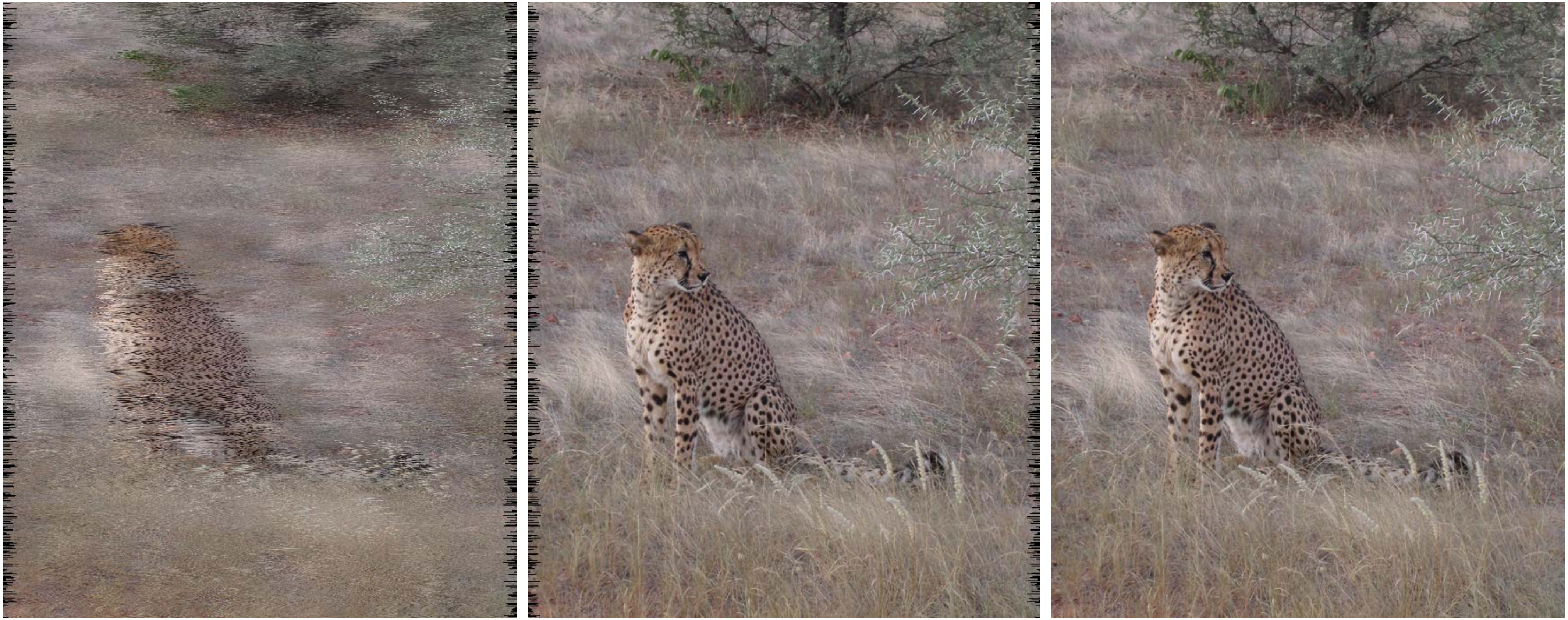
Question **20** Is it easy to solve the numerical problem?

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

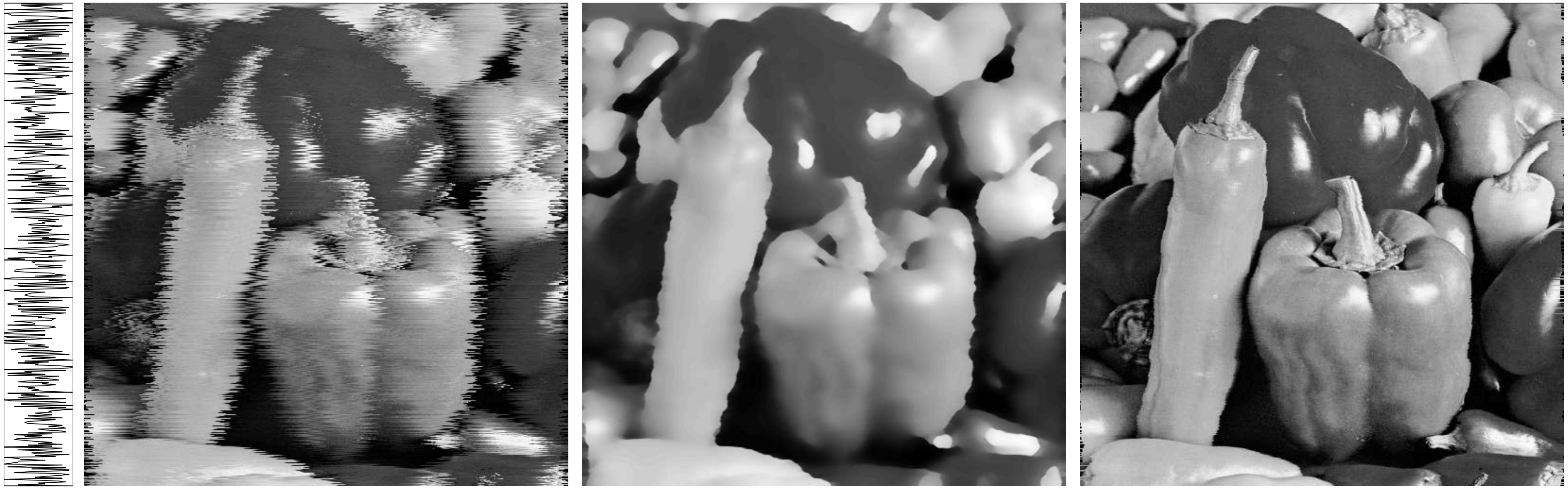
Jittered, $[-20, 20]$ | $\alpha = 1$ | Jitter: $6 \sin\left(\frac{n}{4}\right)$ | $\boldsymbol{\alpha = 1} \equiv$ Original

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

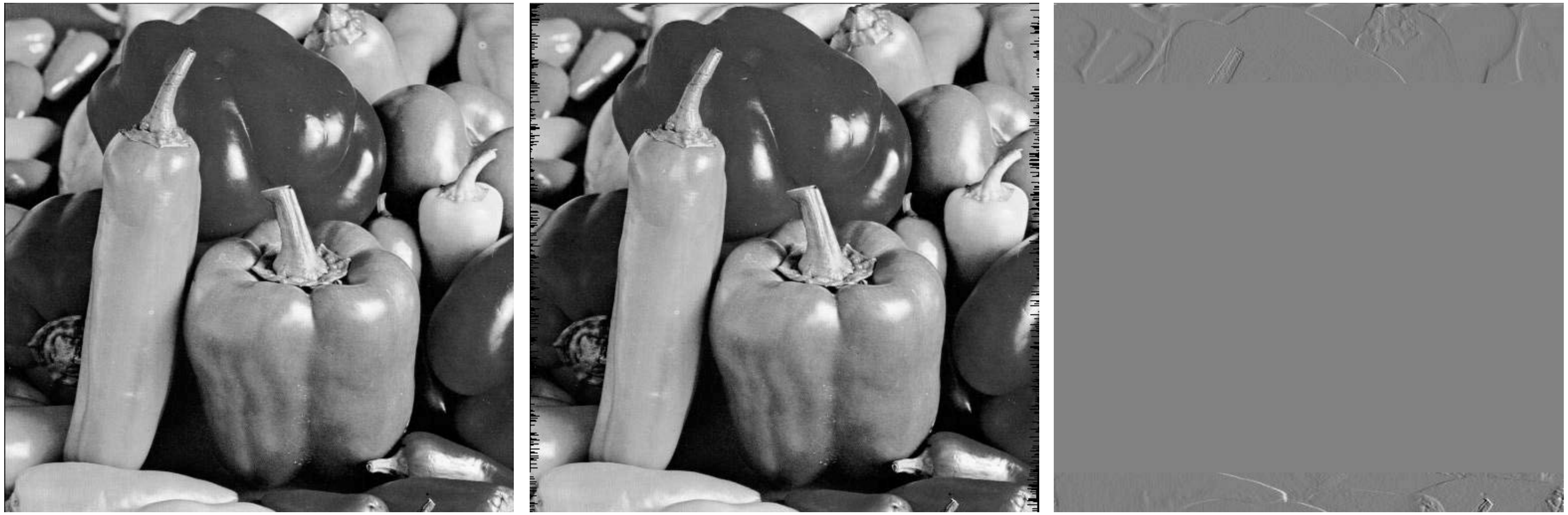
$(512{\times}512)$Jitter$M = 6$ $\boldsymbol{\alpha} \in \{\mathbf{1}, \frac{\mathbf{1}}{\mathbf{2}}\}$=Original Lena (256×256) Jitter $\{-6, .., 6\}$ $\boldsymbol{\alpha} \in \{\mathbf{1}, \frac{\mathbf{1}}{\mathbf{2}}\}$



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



Jitter　Jittered Image　Bayesian TV　Bake & Shake



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

[Kokaram98, Laborelli03, Shen04, Kang06, Scherzer11]

5. Comparison with Fully Smooth Energies

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

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

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

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

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

Question **21** What are the consequences if one approximates a nonsmooth energy by a smooth energy?

Questions to clarify the theoretical results

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

Question **23** Determine the minimizer function $\mathcal{U}$.

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

Question **24** 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 **25** Characterize the set $\mathcal{L} = \{\nu \in \mathbb{R}^p \ : \ a_i \, \mathcal{U}(\nu) = \nu[i]\}$.

Outline

6 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) $\qquad$ $\varphi'(0^+) > 0$ (Φ is nonsmooth)

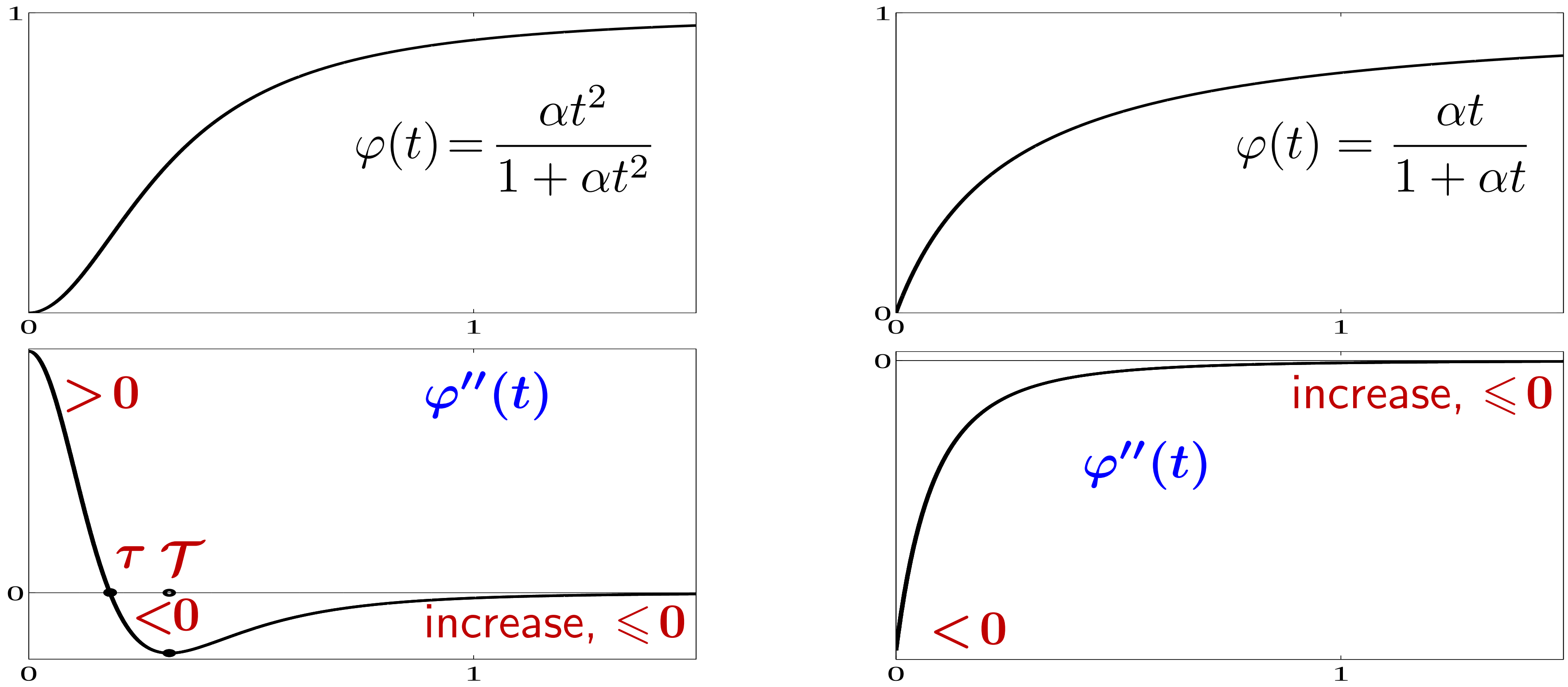
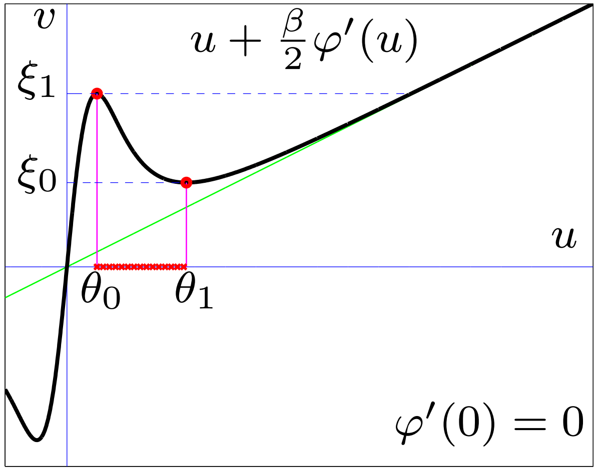
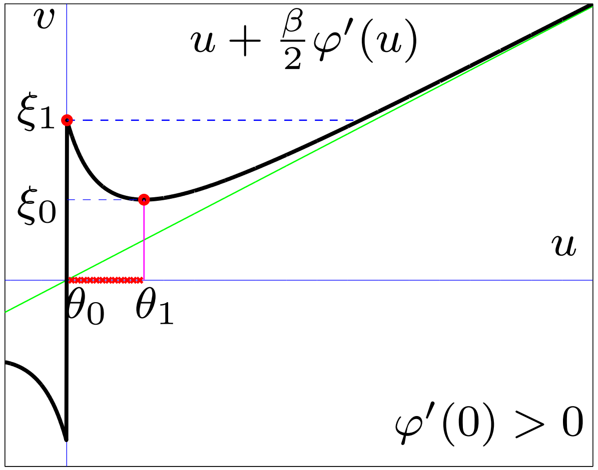


ILLUSTRATION ON $\mathbb{R}$

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





No local minimizer in (θ_0, θ_1)

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

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

strong smoothing

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

loose smoothing

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

$\qquad\qquad\qquad\ \ |v| \geqslant \xi \quad \Rightarrow \quad$ 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

Sharp edge property

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

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

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

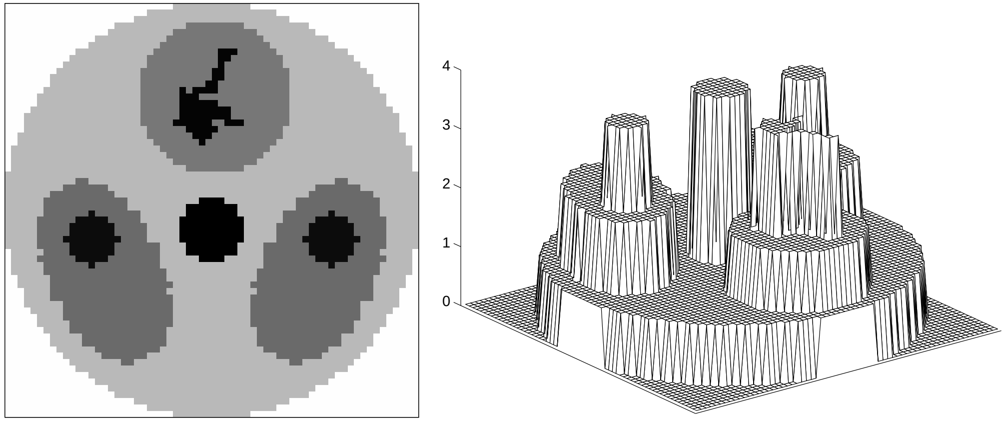
When β increases, then θ_0 decreases and θ_1 increases.

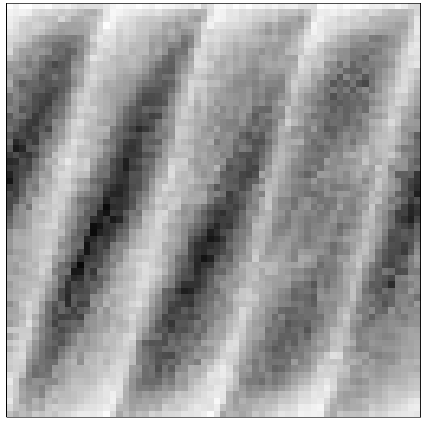
In particular

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

Question **26** Explain the prior model involved in $\mathcal{F}_v$ when φ is nonconvex with $\varphi'(0) = 0$ and with $\varphi'(0^+) > 0$.

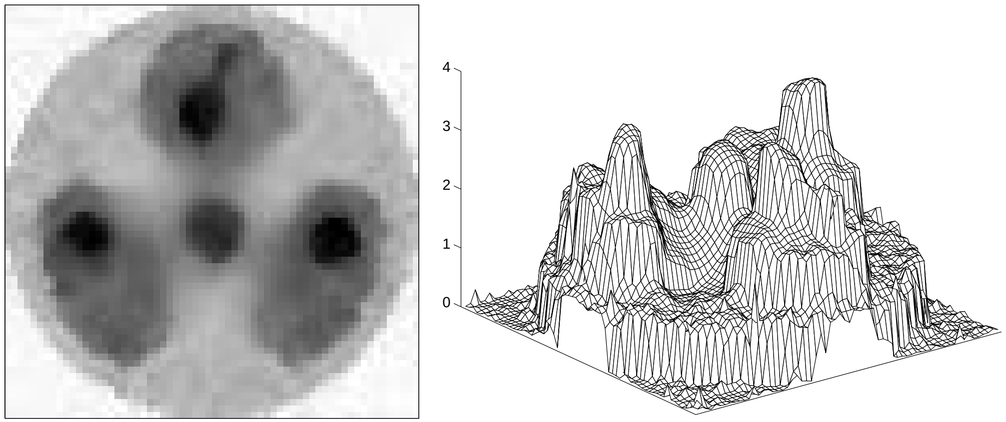
Image Reconstruction in Emission Tomography

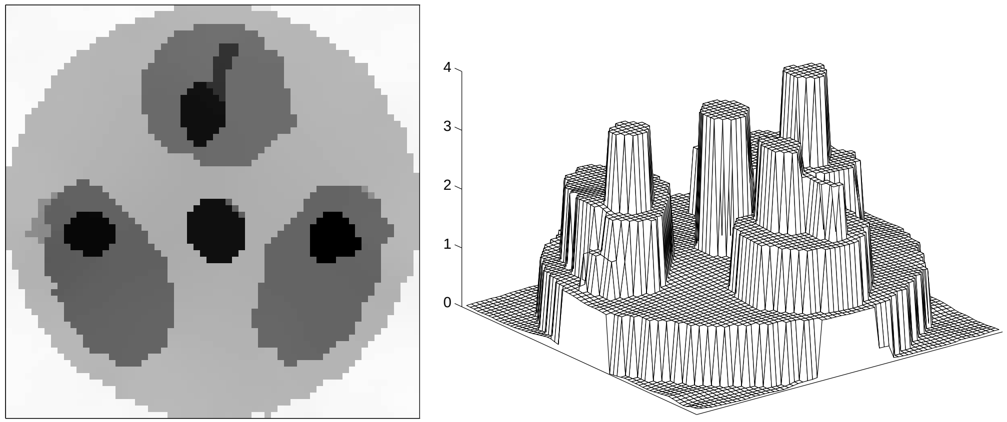




Original phantom

Emission tomography simulated data





φ is smooth (Huber function)

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

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

Selection for the global minimizer

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

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

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

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

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

Sketch of the results

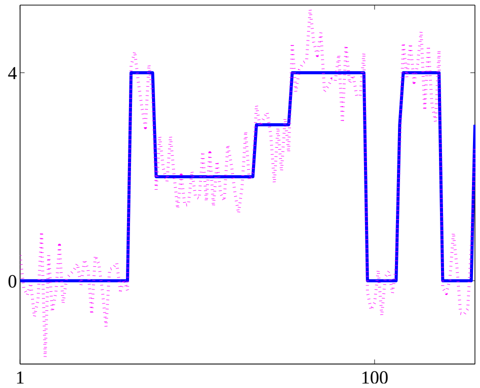
$\exists \; \xi_1 > 0$ such that $\xi > \xi_1 \quad \Rightarrow \quad \hat{u}$—perfect edges

Moreover:

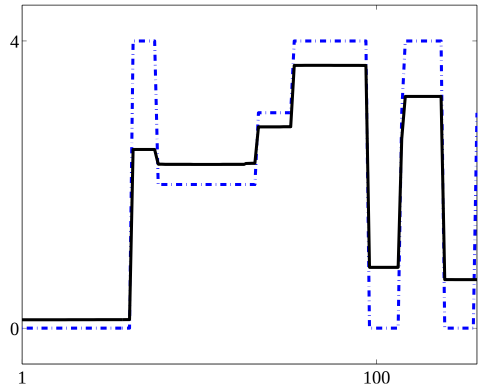
- Φ non smooth, then $\quad \xi > \xi_1 \quad \Rightarrow \quad \hat{u} = c \, u_o, \quad c < 1, \; \lim\limits_{\xi \to \infty} c = 1$
- $\varphi(t) = \eta, \; t \geqslant \eta$, then $\xi > \xi_1 \quad \Rightarrow \quad \hat{u} = u_o$

This holds true also for $\varphi(t) = \min\{\alpha t^2, \; 1\}$ and for $\varphi(t) = \begin{cases} 0 & \text{if} \quad t = 0 \\ 1 & \text{if} \quad t \neq 0 \end{cases}$

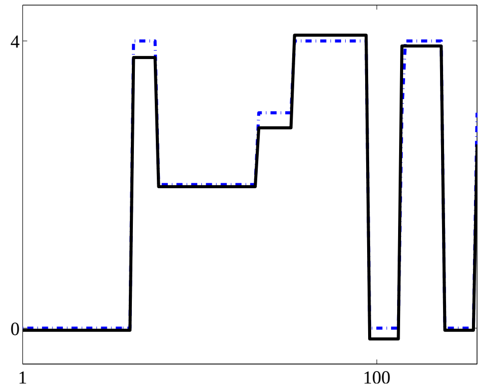
Comparison with Convex Edge-Preserving Regularization



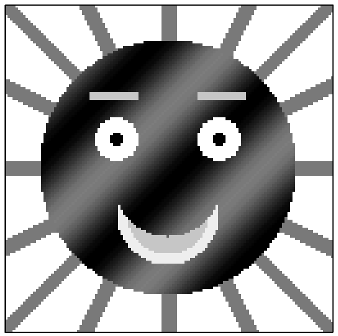
Data $v = u_o + n$



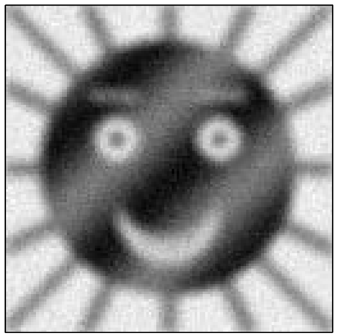
$\varphi(t) = |t|$



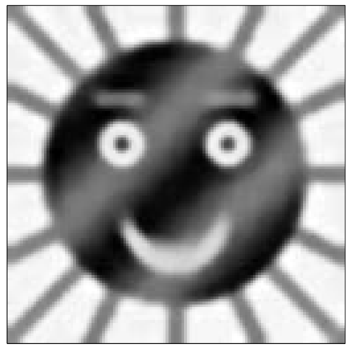
$\varphi(t) = \alpha|t|/(1 + \alpha|t|)$



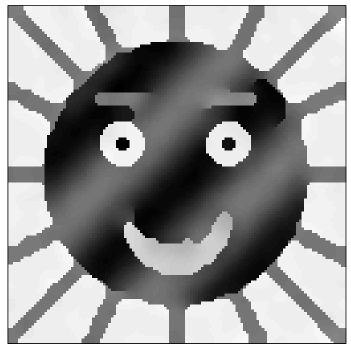
original



data

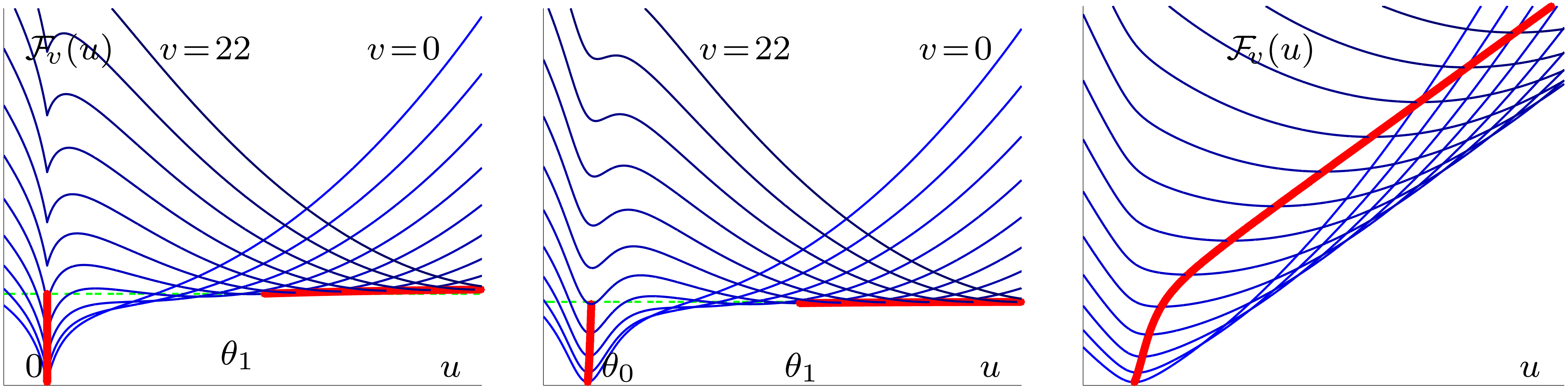


$\varphi(t) = |t|^{1.4}$



$\varphi(t) = \min\{\alpha t^2, 1\}$

Question **27** Why edges are sharper when φ is nonconvex?



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

$\mathcal{F}_v(u) = (u-v)^2 + \beta \frac{\alpha u^2}{(1+\alpha u^2)}$ global minimizer functions (●●)

$\mathcal{F}_v(u) = (u-v)^2 + \beta\sqrt{\alpha + u^2}$ unique minimizer function (●●)

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

Question **28** How to describe the global minimizer when v increases?

Summer School 2014: **Inverse Problem and Image Processing**

Tutorial: **Inverse modeling in inverse problems using optimization**

Outline

7. 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

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

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

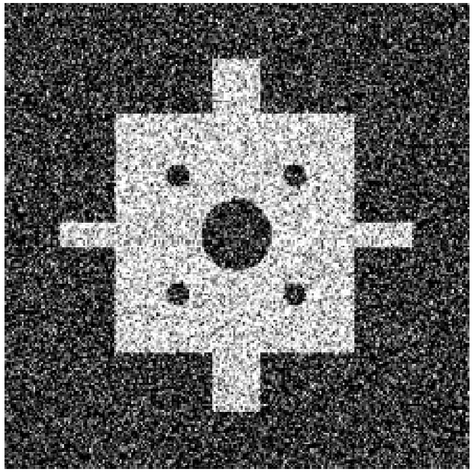
[T. Chan, S. Esedoglu, Nikolova 06]

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

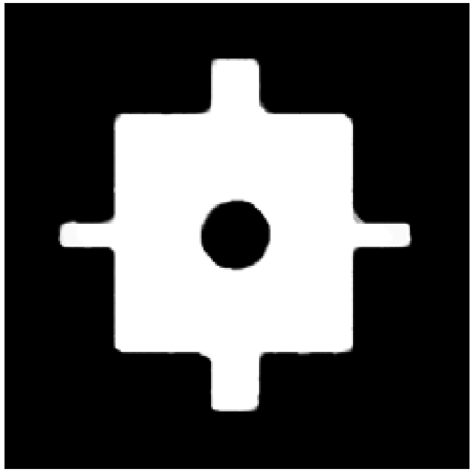
$$\hat{\Sigma} = \arg\min_{\Sigma} \left\{ \left\| \mathbb{1}_\Sigma - \mathbb{1}_\Omega \right\|_2^2 + \beta \mathrm{TV}(\mathbb{1}_\Sigma) \right\} \qquad \text{nonconvex problem} \qquad (\star)$$

usual techniques (curve evolution, level-sets) fail

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



Data

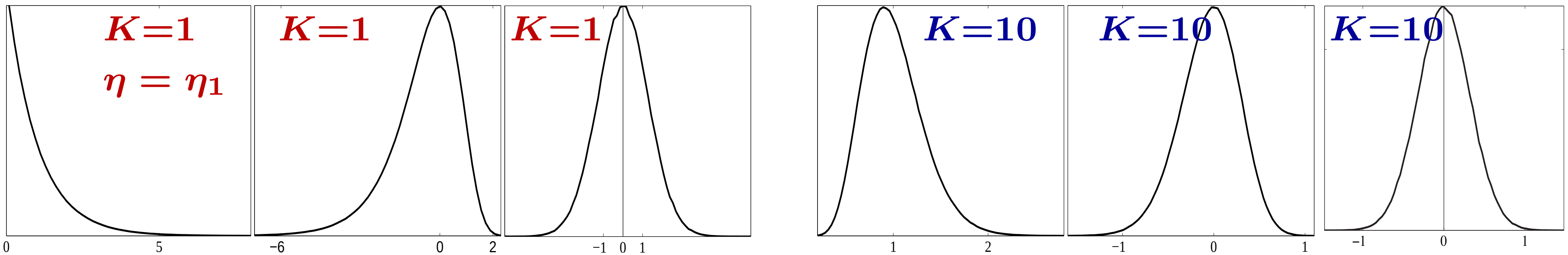


Restored

Multiplicative noise removal on Frame coefficients [Durand, Fadili, Nikolova 09]

Multiplicative noise arises in various active imaging systems e.g. synthetic aperture radar

- Original image: S_o
- One shot: $\Sigma_k = S_o \eta_k$
- Data: $\Sigma = \frac{1}{K}\sum_{k=1}^{K} \Sigma_k = S_o \frac{1}{K}\sum_{k=1}^{K} \eta_k = S_o\,\eta$ where $\mathrm{pdf}(\eta) =$ Gamma density
- Log-data: $v = \log \Sigma = \log S_o + \log \eta = u_0 + n$
- Frame Coefficients: $y = Wv = Wu_0 + Wn$ (W curvelets)



$\mathrm{pdf}(\eta) = \mathrm{pdf}(\eta_k)$ $\mathrm{pdf}(n)$ $\mathrm{pdf}\left(Wn\right)$ $\mathrm{pdf}(\eta)$ $\mathrm{pdf}(n)$ $\mathrm{pdf}\left(Wn\right)$

Question **29** Comment the noise distribution of Wn

- Hard Thresholding: $y_T[i] = \begin{cases} 0 & \text{if } \ |y[i]| \leqslant T, \\ y[i] & \text{otherwise} \end{cases} \qquad \forall i \in I, \quad T > 0$ (suboptimal).

$$I_1 = \{i \in I \ : \ |y[i]| > T\} \ \text{ and } \ I_0 = I \setminus I_1$$

- **Restored coefficients:** $\hat{x} = \arg\min_x \mathcal{F}_y(x)$ $\quad$ ($\ell_1 - \mathrm{TV}$ energy)

$$\mathcal{F}_y(x) = \lambda_0 \sum_{i \in I_0} \big|x[i]\big| + \lambda_1 \sum_{i \in I_1} \big|x[i] - y[i]\big| + \|\widetilde{W} x\|_{\mathbf{TV}}$$

$\hat{S} = B \exp\left(\widetilde{W} \hat{x}\right)$, where $\widetilde{W}$ left inverse, B bias correction

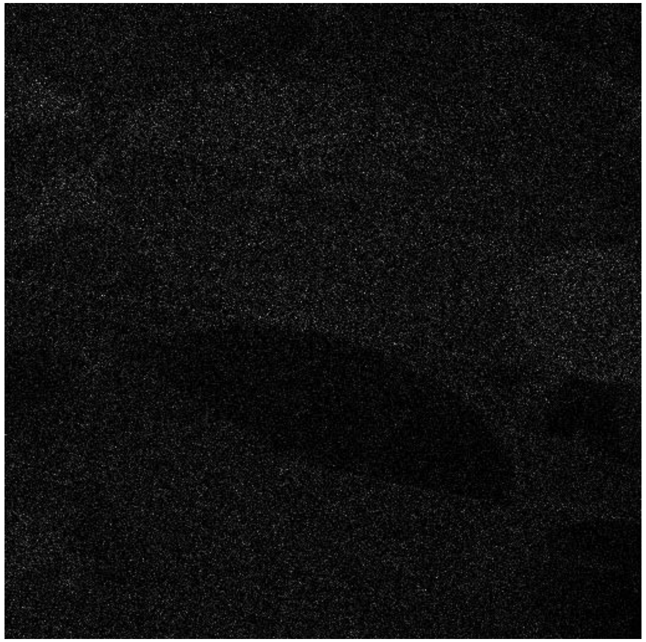
Question **30** Explain the job the minimizer $\hat{x}$ of $\mathcal{F}_y$ should do.

Some comparisons

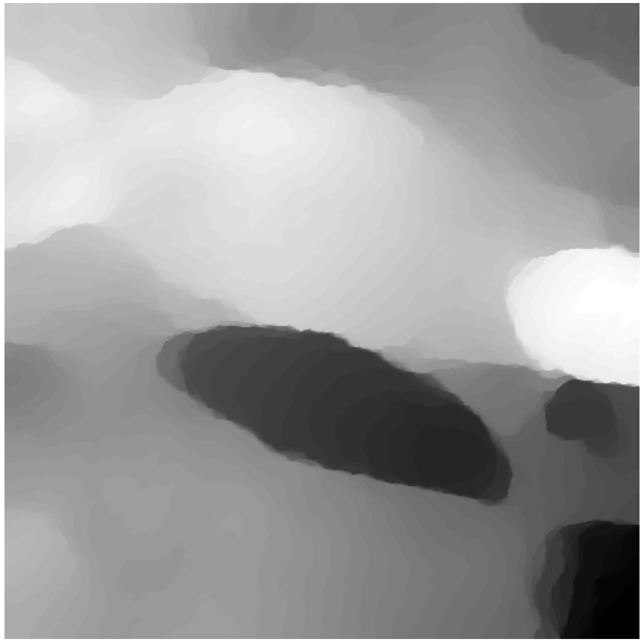
- **BS** [Chesneau,Fadili,Starck 08]: Block-Stein thresholds the curvelet coefficients, $\approx$ minimax(large class of images with additive noises), optimal threshold $\mathfrak{T} = 4.50524$
- **AA** [Aubert,Aujol 08]: $\Psi = -$ Log-Likelihood(Σ), $\Phi = \mathrm{TV}(\Sigma)$ $\quad$ (i.e. $\mathcal{F}_v \equiv$ MAP for Σ)
- **SO** [Shi,Osher 08]: relaxed inverse scale-space for $\mathcal{F}_v(u) = \|v - u\|_2^2 + \beta \mathrm{TV}(u)$ $\ \approx$ MAP(u)

  Stopping rule: $k^* = \max\{k \in \mathrm{N} : \mathrm{Var}(u^{(k)} - u_o) \geqslant \mathrm{Var}(n)\}$.

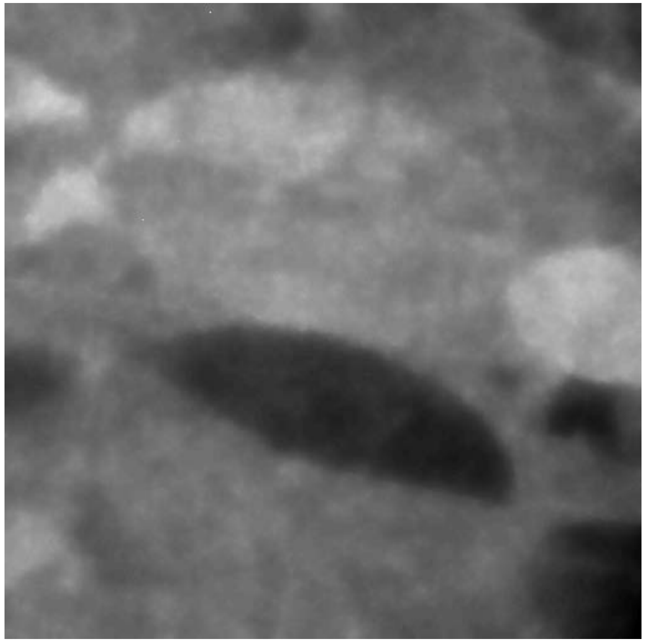
Monte-Carlo comparative experiment confirms the proposed method



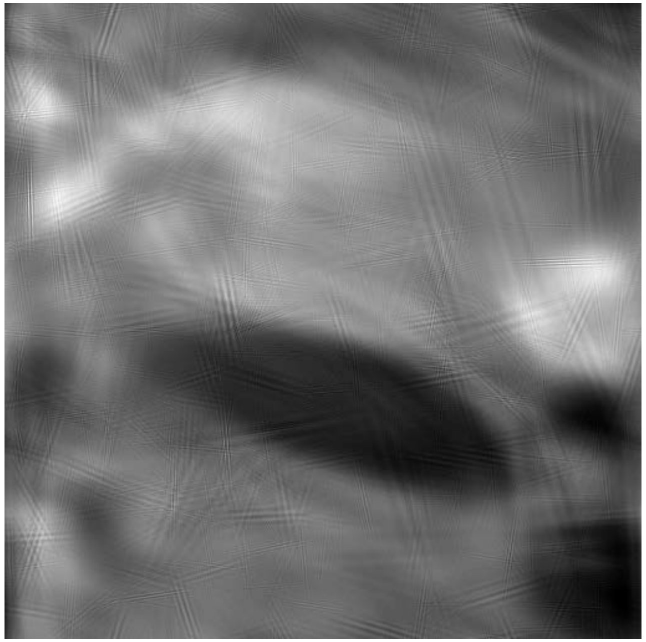
Noisy Fields $K = 1$ (512×512)



SO: PSNR=9.59, MAE=196

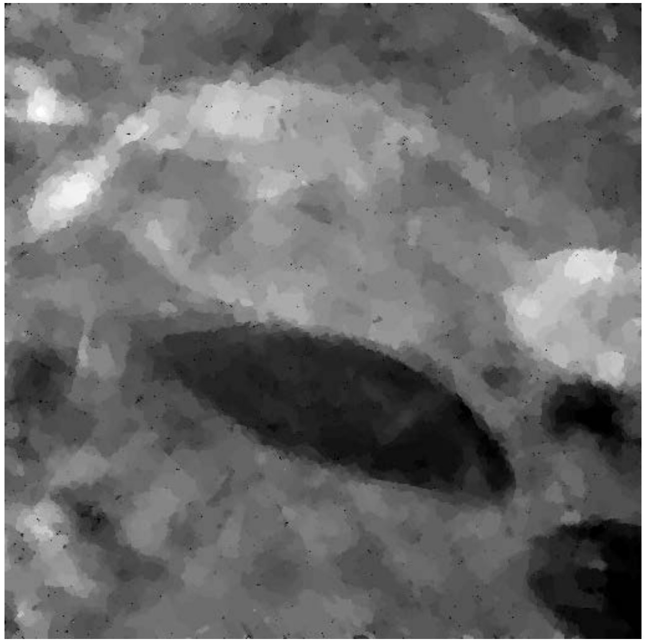


AA: PSNR=15.74, MAE=76.66

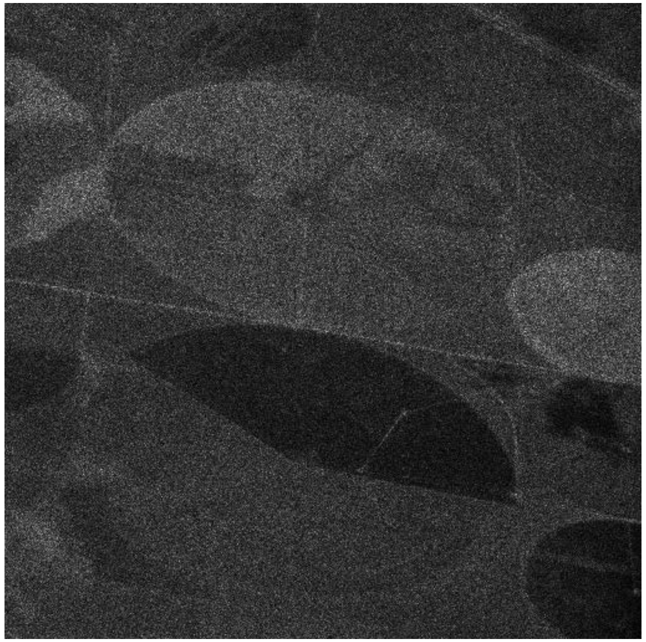


BS: PSNR=22.52, MAE=35.22

Fields (original)

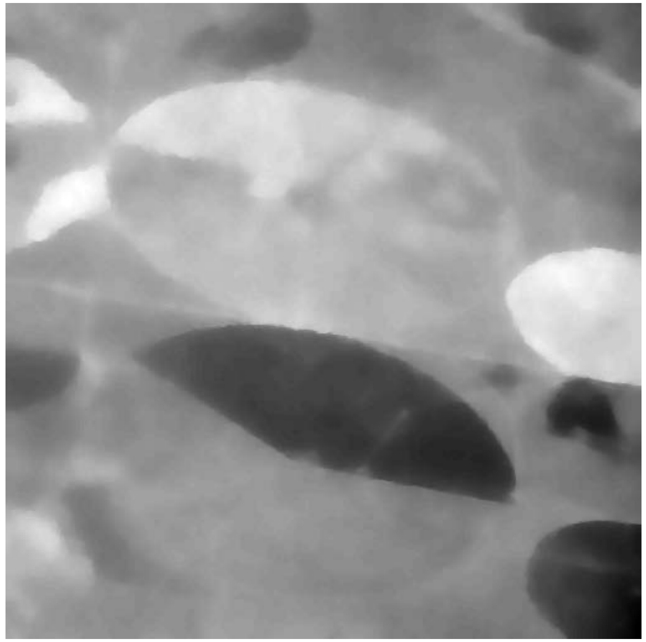


Our: PSNR=22.89, MAE=33.67

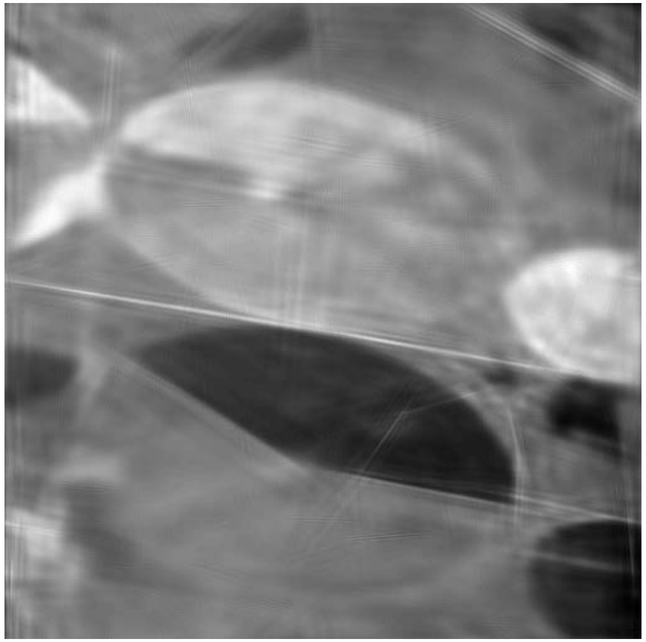


Noisy $K = 10$

SO: PSNR=25.36, MAE=25.14

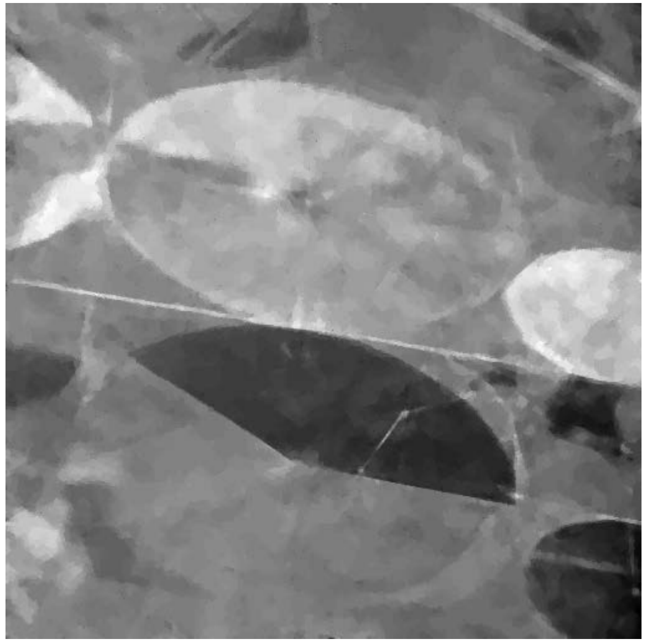


AA: PSNR=17.13, MAE=65.40



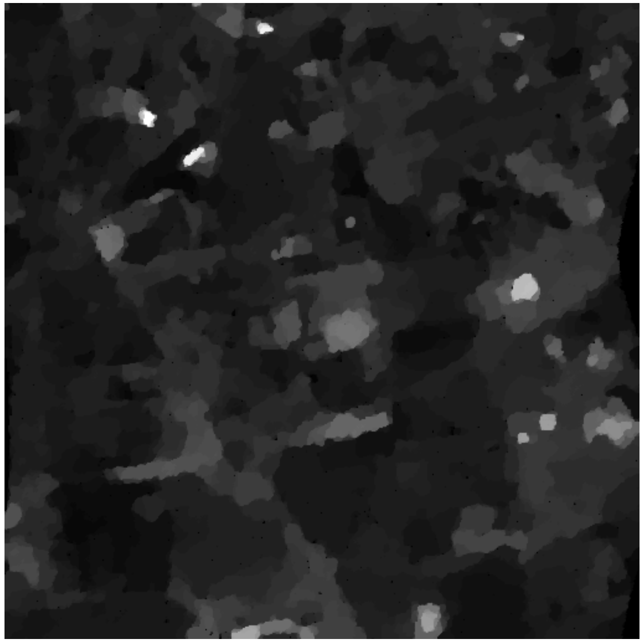
BS: PSNR=27.24, MAE=19.61

Fields (original)



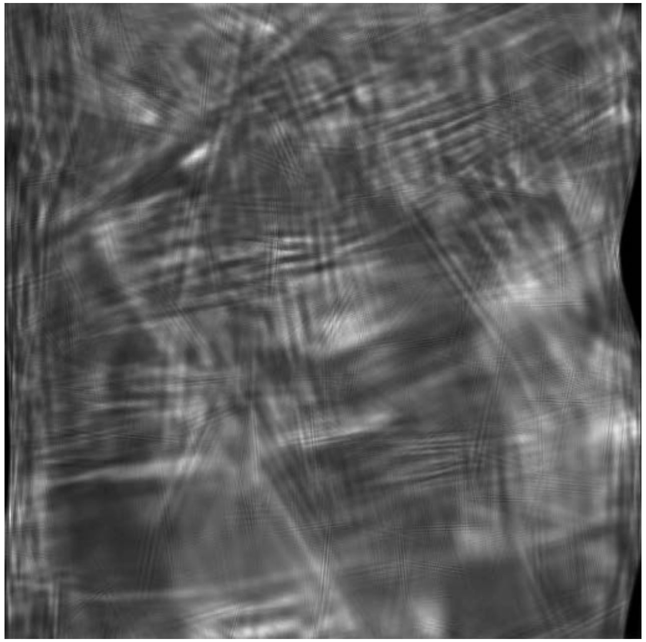
Our: PSNR=28.04, MAE=18.19

Noisy City $K = 1$ (512×512)



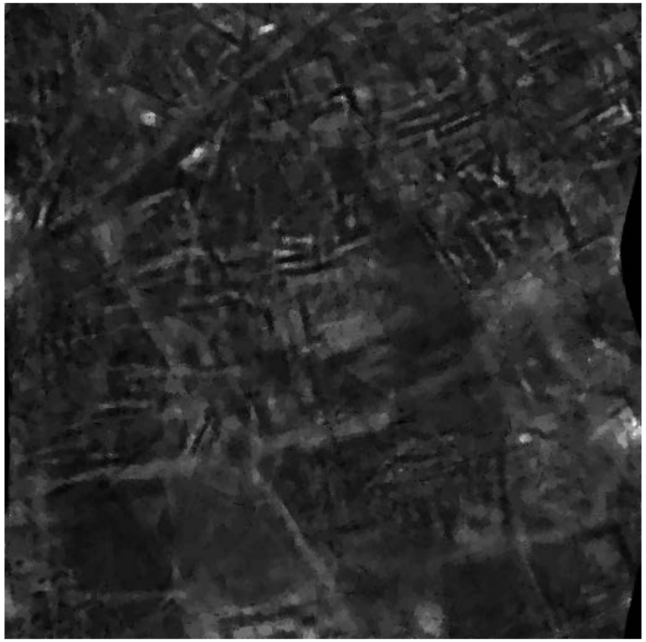
SO: PSNR=18.39, MAE=24.08

AA: PSNR=22.18, MAE=13.71



BS: PSNR=22.25, MAE=13.96

City (original)

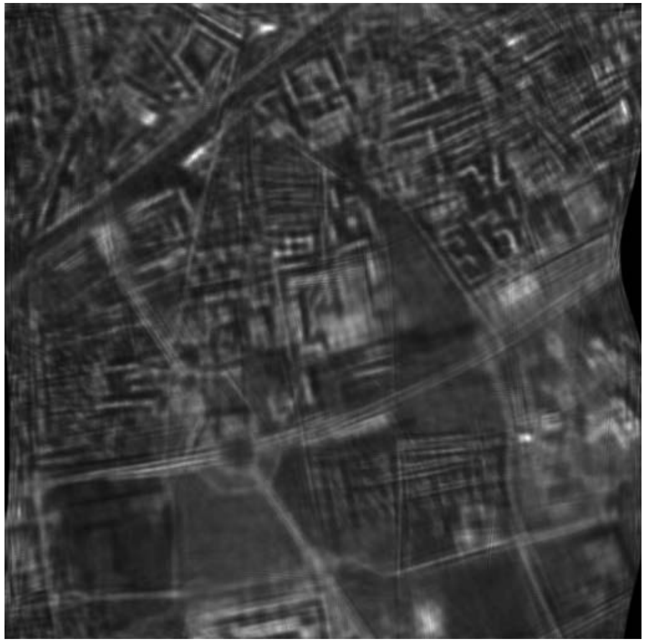


Our: PSNR=22.64, MAE=13.39

Noisy $K = 4$

SO: PSNR=24.40, MAE=10.76

AA: PSNR=24.55, MAE=10.06



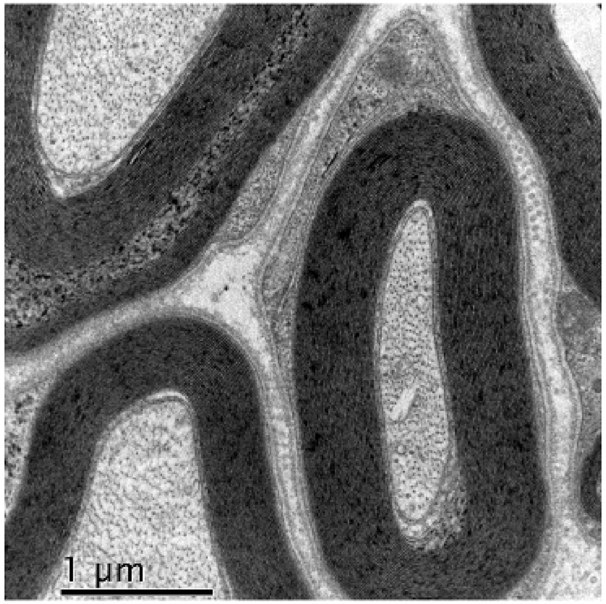
BS: PSNR=24.92, MAE=9.87

City (original)

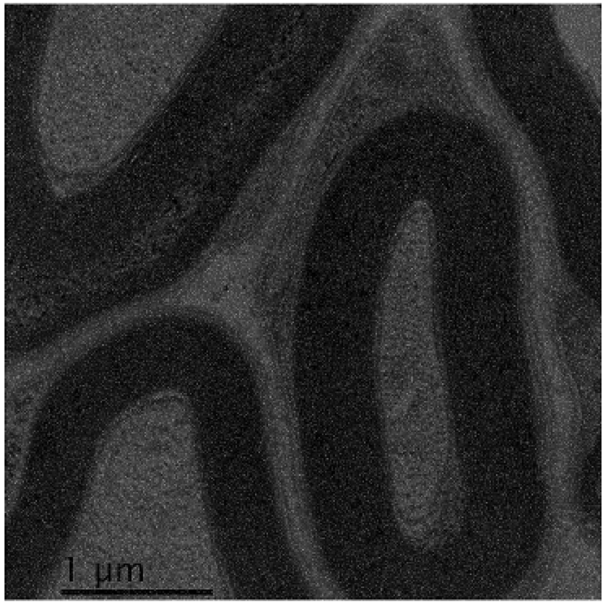
Our: PSNR=25.84, MAE=9.09

C. Clason, B. Jin, K. Kunisch
"Duality-based splitting for fast $\ell_1 - \mathrm{TV}$ image restoration", 2012,
http://math.uni-graz.at/optcon/projects/clason3/

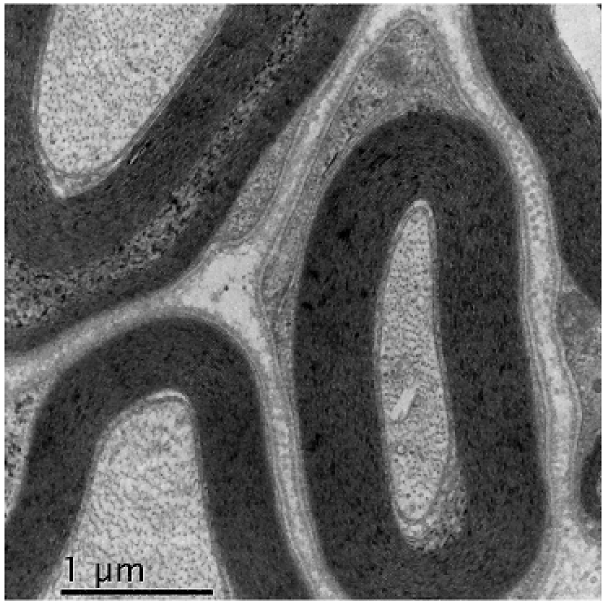
Scanning transmission electron microscopy (2048×2048 image)



true image



noisy image



restoration

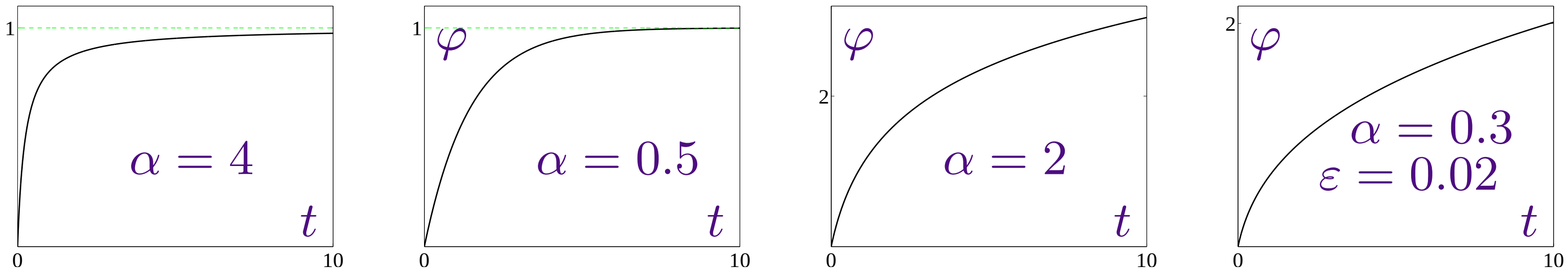
ℓ_1 data-fidelity with concave regularization [Nikolova, Ng, Tam 12]

$$\mathcal{F}_v(u) = \sum_{i\in I} \big|a_i u - v[i]\big| + \beta \sum_{j\in J} \varphi(\|G_j u\|_2), \quad \varphi'(0^+) > 0, \;\; \varphi''(t) < 0, \;\; \forall t \geqslant 0$$

$$I = \{1, \cdots, q\}, \quad J = \{1, \cdots, r\}$$

φ is strictly concave on $[0, +\infty)$.

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



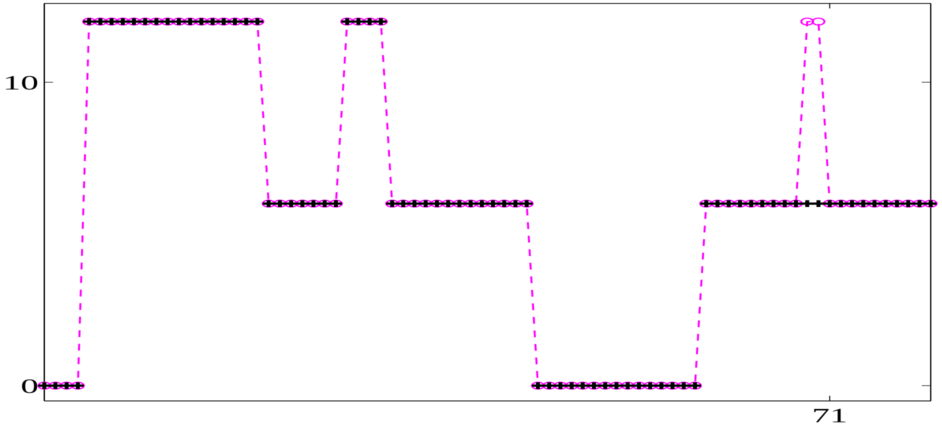
Motivation

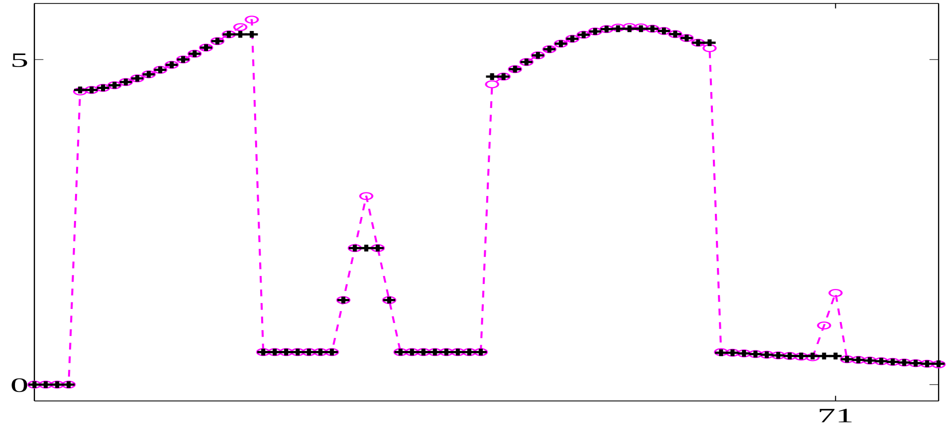
- This family of objective functions has never been considered before
- $\mathcal{F}_v$ can be seen as an extension of $L1 - \mathrm{TV}$
- $\hat{u}$—(local) minimizer of $\mathcal{F}_v$ $\overset{?}{\Longrightarrow}$ many i, j such that $a_i\hat{u} = v[i]$ and $G_j\hat{u} = 0$

Minimizers of $\mathcal{F}_v(u) = \|u - v\|_1 + \beta \sum_{i=1}^{p-1} \varphi(|u[i+1] - u[i]|)$

$\varphi(t) = \frac{\alpha t}{\alpha t + 1}$ for $\alpha = 4$

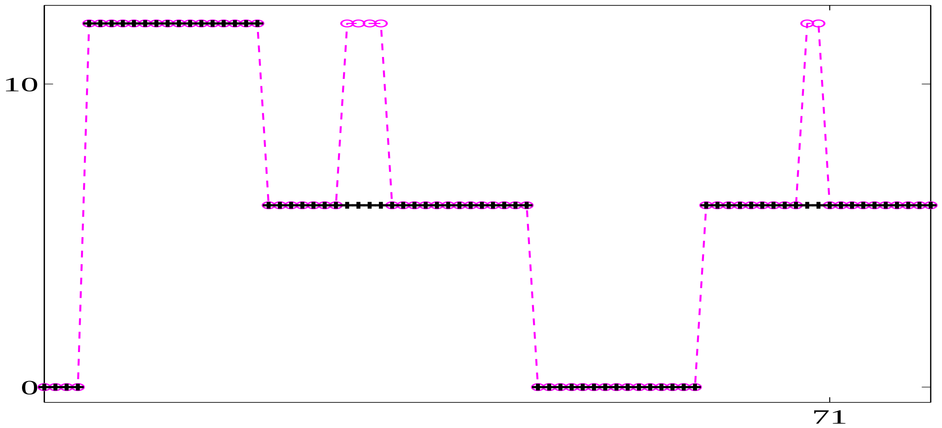
$\varphi(t) = \ln(\alpha t + 1)$ for $\alpha = 2$

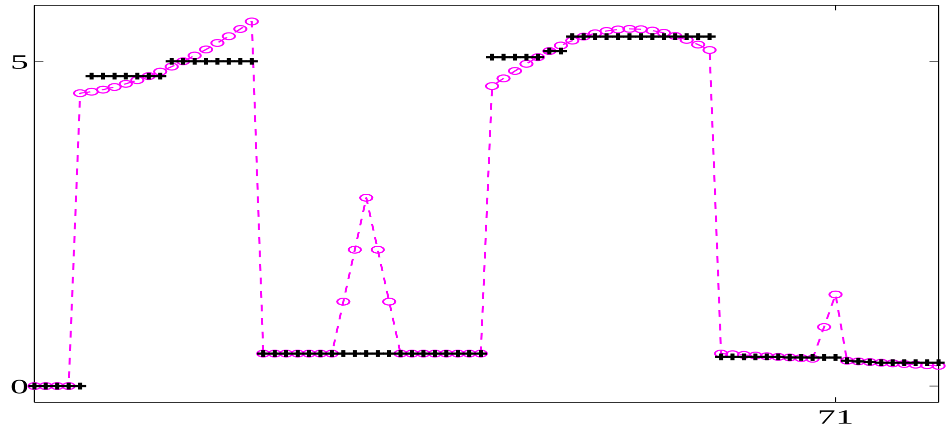




$\beta \in \{78, \cdots, 156\}$

$\beta \in 0.1 \times \{10, \cdots, 14\}$

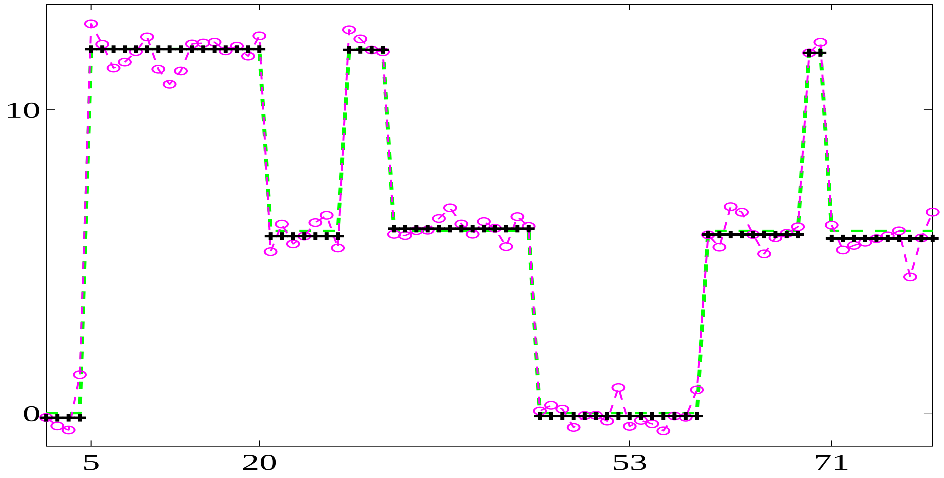




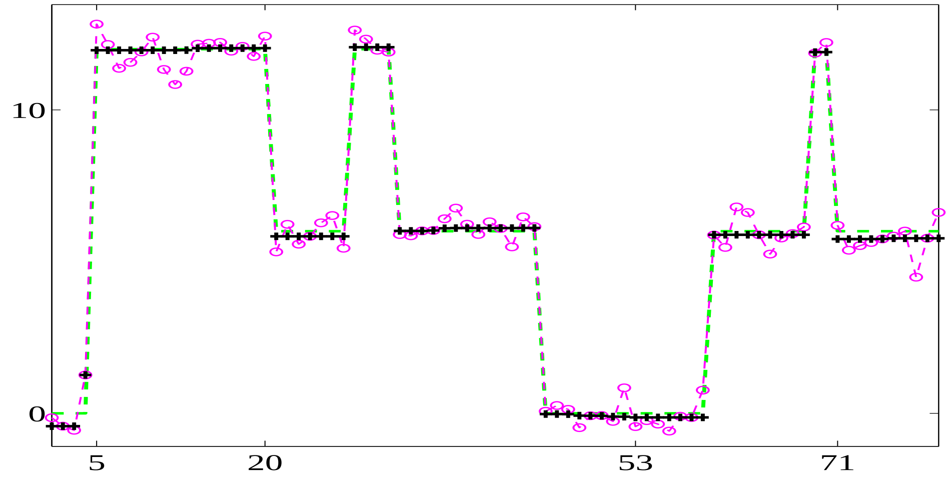
$\beta \in \{157, \cdots, 400\}$

$\beta \in 0.1 \times \{16, \cdots, 30\}$

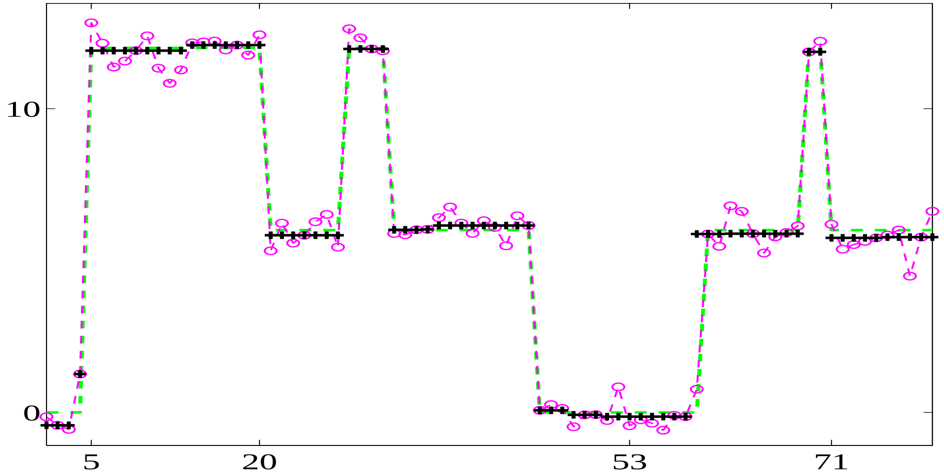
Data samples (ooo), Minimizer samples $\hat{u}[i]$ (+++).



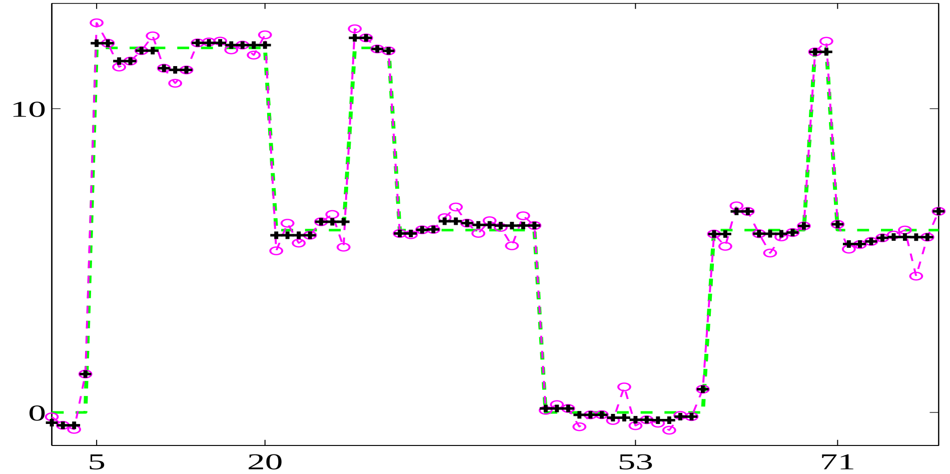
(a) $\varphi(t) = \frac{\alpha\, t}{\alpha\, t+1}$, $\alpha = 4$, $\beta = 3$



(b) $\varphi(t) = 1 - \alpha^t$, $\alpha = 0.1$, $\beta = 2.5$



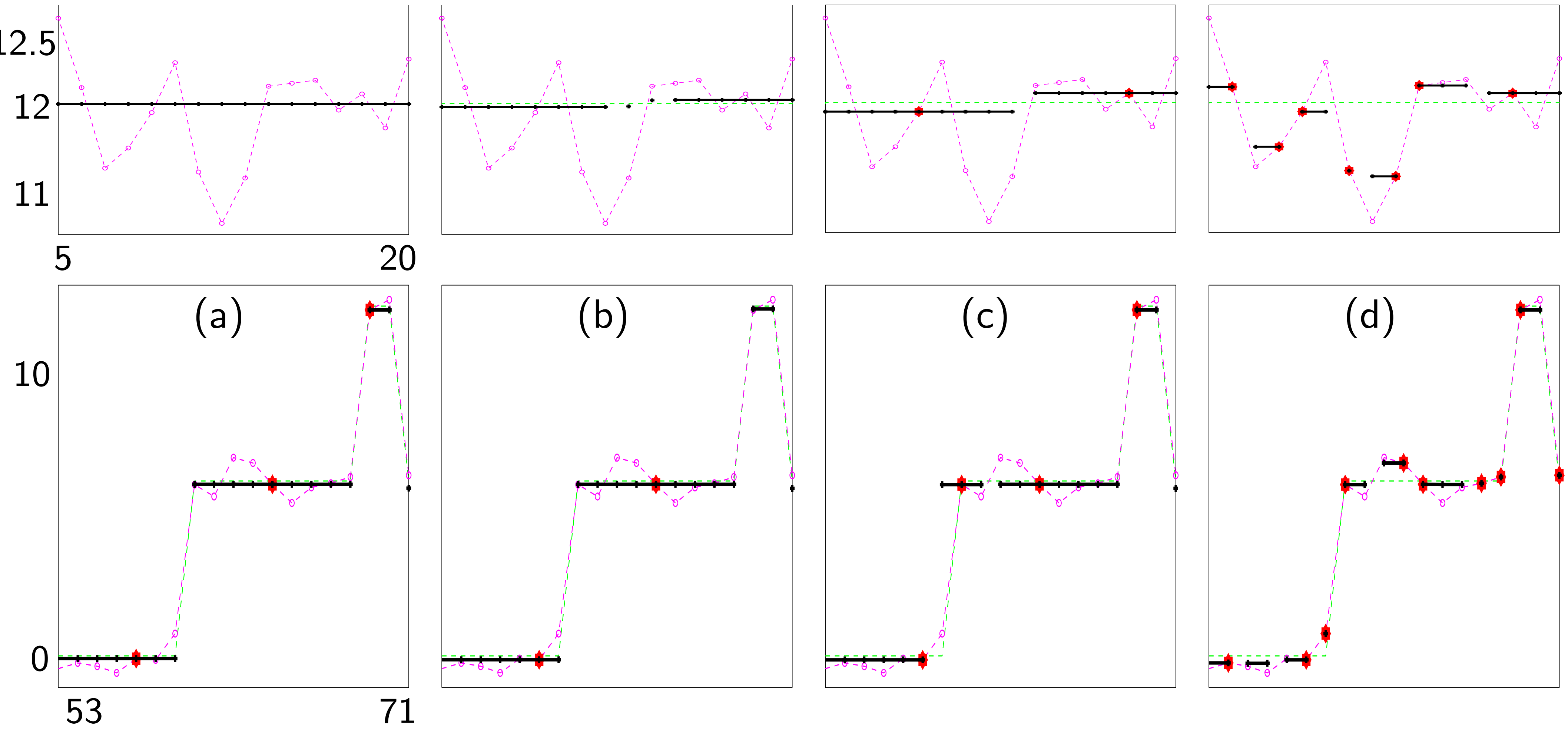
(c) $\varphi(t) = \ln(\alpha t + 1)$, $\alpha = 2$, $\beta = 1.3$



(d) $\varphi(t) = (t + 0.1)^\alpha$, $\alpha = 0.5$, $\beta = 1.4$

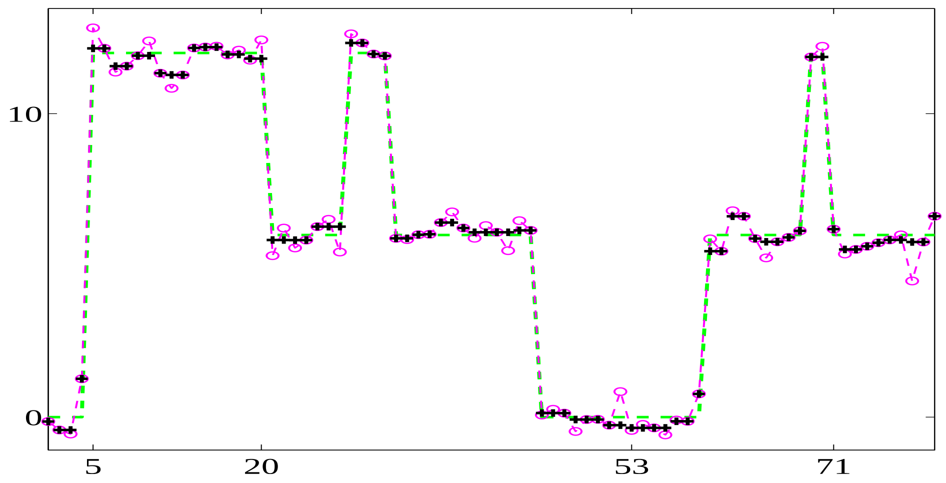
Denoising: Data samples (ooo) are corrupted with Gaussian noise. Minimizer samples $\hat{u}[i]$ (+++). Original (– – –). $\boldsymbol{\beta}$—the largest value so that the gate at 71 survives.

Zooms



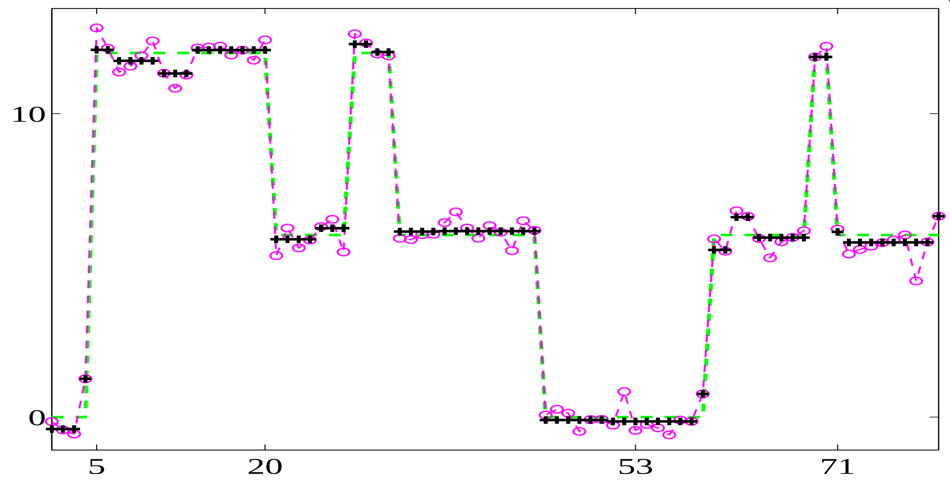
Constant pieces—solid black line.

Data points $v[i]$ fitted exactly by the minimizer $\hat{u}$ (♦).



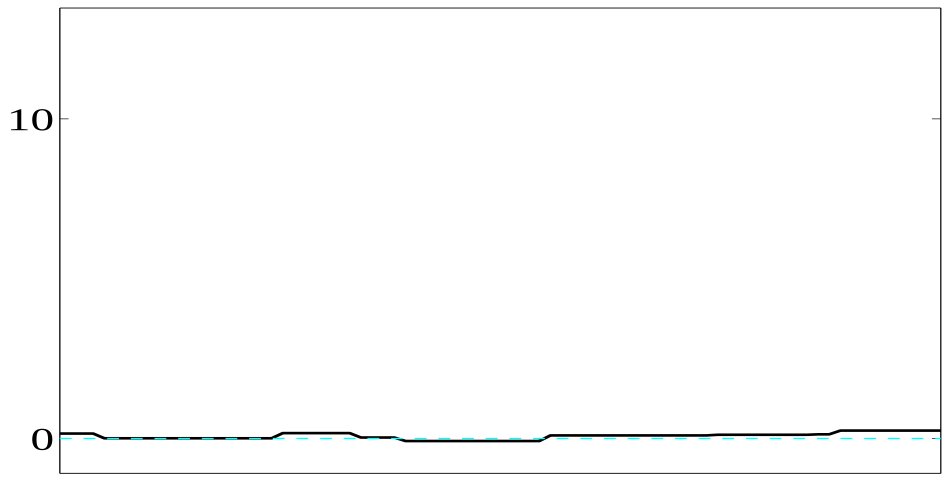
$\varphi(t)=t,\ \beta=0.8 \quad (\ell_1-\mathrm{TV})$

the convex relaxation of $\mathcal{F}_v$



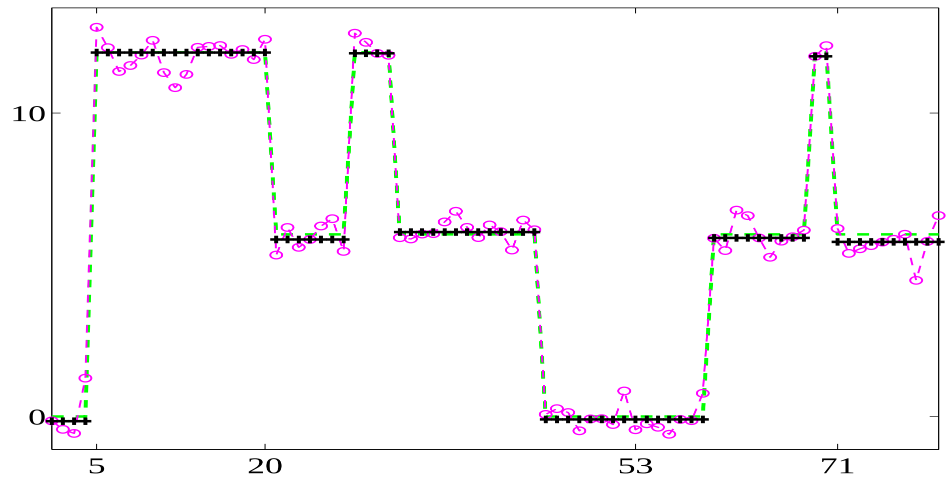
the minimizer for $\varphi(t)=\frac{\alpha\ t}{\alpha\ t+1}$, $\alpha=4$, $\beta=3$

closest to $(\ell_1-\mathrm{TV})$



error for $\varphi(t)=\frac{\alpha\ t}{\alpha\ t+1}$, $\alpha=4$, $\beta=3$

$\|\text{original}-\hat{\boldsymbol{u}}\|_\infty=0.24$



$\varphi(t)=\frac{\alpha\ t}{\alpha\ t+1}$, $\alpha=4$, $\beta=3$

original $\in[0,12]$, data $v\in[-0.6,12.9]$

On the figures, $\hat{u}$ are global minimizers of $\mathcal{F}_v$ (Viterbi algorithm)

Question **31** Can you sketch the main properties of the minimizers of $\mathcal{F}_v$?

Question **32** What seems being the role of the asymptotic of φ?

Numerical evidence:

critical values $\beta_1, \cdots, \beta_n$ such that

- $\beta \in [\beta_i, \beta_{i+1}) \Rightarrow$ the minimizer remains unchanged
- $\beta \geqslant \beta_{i+1} \quad \Rightarrow$ the minimizer is simplified

Result proven (under conditions) for the minimizers of $L_1 - \mathrm{TV}$ in [Chan, Esedoglu 2005]

Given $v \in \mathbb{R}$ consider the function

$$\mathcal{F}_v(u) = |u - v| + \beta\varphi(|u|) \quad \text{for} \quad \varphi(u) = \frac{\alpha u}{1 + \alpha u} \qquad u \in \mathbb{R}, \qquad \beta > 0$$

Question **33** Does $\mathcal{F}_v$ have a global minimizer for any v? Explain.

Question **34** Determine $\varphi''(u)$ for $u \in \mathbb{R} \setminus \{0\}$.

Question **35** Show that $\forall\, v \in \mathbb{R}$, any minimizer $\hat{u}$ of $\mathcal{F}_v$ obeys $\hat{u} \in \{0, v\}$.

Question **36** Can you extend this result to the other φ on p. 71?

- $\mathcal{F}_v$ does have global minimizers, for any $\{a_i\}$, for any v and for any $\beta > 0$.

- Let $\hat{u}$ be a (local) minimizer of $\mathcal{F}_v$. Set

$$\begin{aligned} \widehat{I}_0 &= \{i \in I \ : \ a_i \hat{u} = v[i]\} \\ \widehat{J}_0 &= \{j \in J \ : \ G_j \hat{u} = 0\} \end{aligned}$$

$\hat{u}$ is the **unique** point solving the liner system

$$\begin{cases} a_i \hat{u} = v[i] & \forall i \in \widehat{I}_0 \\ G_j \hat{u} = 0 & \forall j \in \widehat{J}_0 \end{cases}$$

Each pixel of a (local) minimizer $\hat{u}$ of $\mathcal{F}_v$ is involved in (at least) one equation $a_i \hat{u} = v[i]$, or in (at least) one equation $G_j \hat{u} = 0$, or in both types of equations.

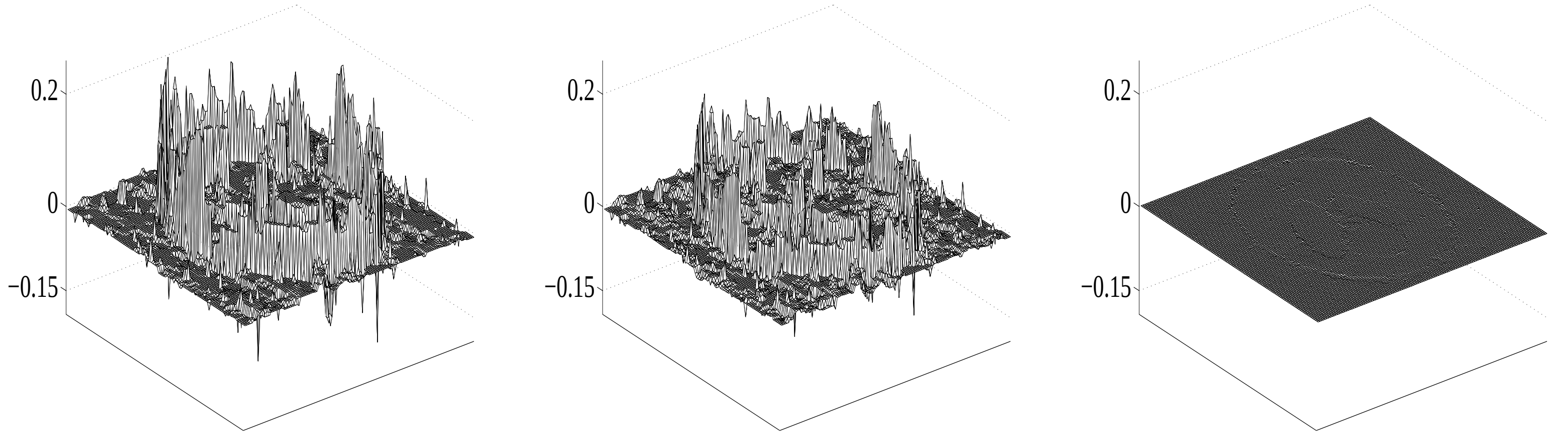
- "Contrast invariance" of (local) minimizers

- The matrix with rows $\left(a_i, \forall i \in \widehat{I}_0, \ \ G_j, \forall j \in \widehat{J}_0\right)$ has **full column rank**

- All (local) minimizers of $\mathcal{F}_v$ are **strict**

MR Image Reconstruction from Highly Undersampled Data

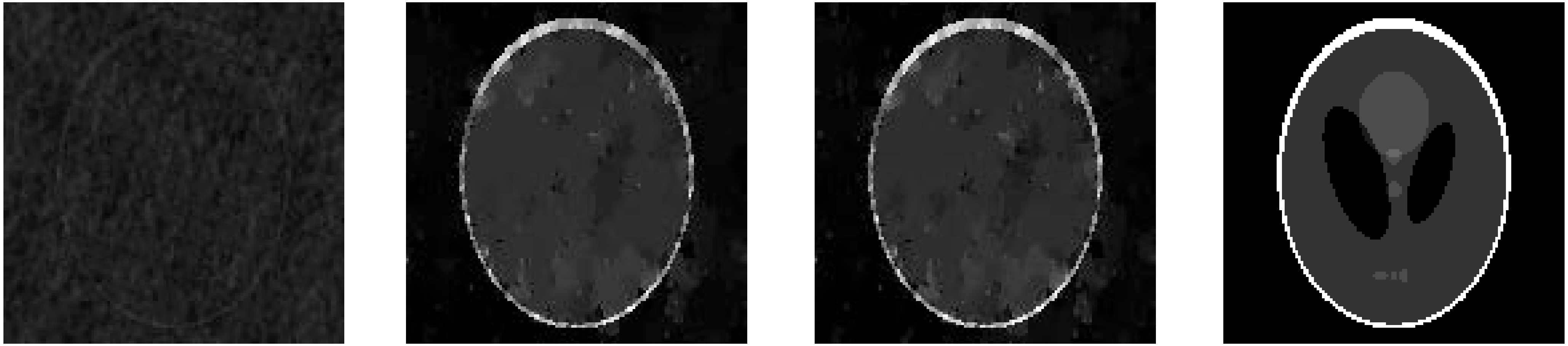
0-filling Fourier | $\|\cdot\|_2^2$+TV | $\|\cdot\|_1$+TV | Our method



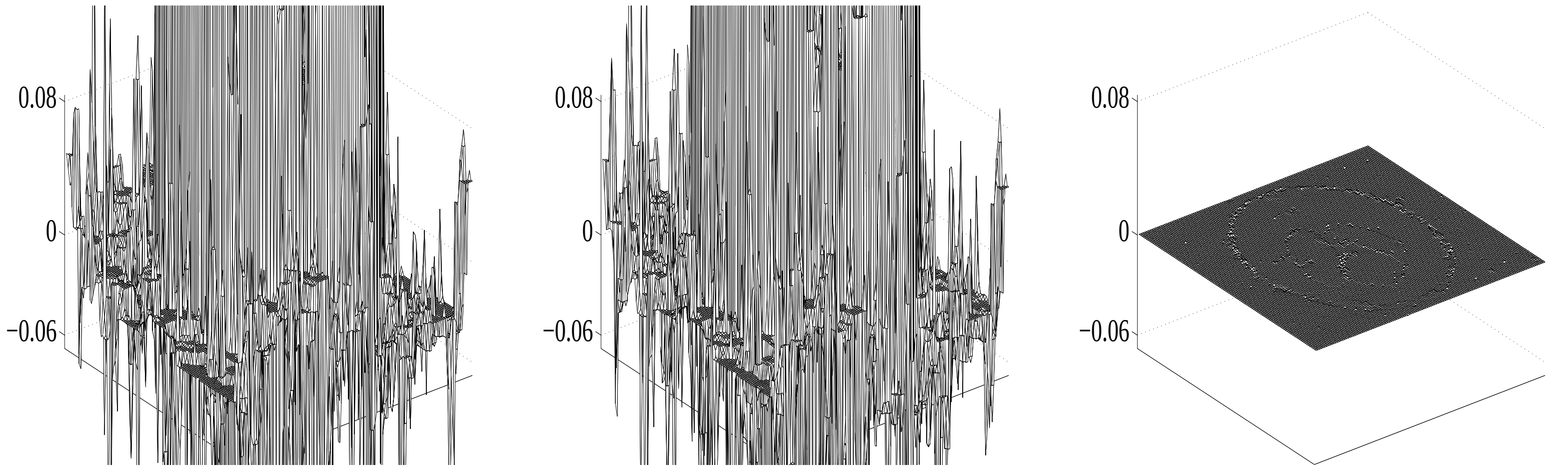
Reconstructed images from **7% noisy** randomly selected samples in the k-space.

Our method for $\varphi(t) = \dfrac{\alpha t}{\alpha t + 1}$.

MR Image Reconstruction from Highly Undersampled Data



0-filling Fourier	$\|\cdot\|_2^2$+TV	$\|\cdot\|_1$+TV	Our method



Reconstructed images from **5% noisy** randomly selected samples in the k-space.

Our method for $\varphi(t) = \dfrac{\alpha t}{\alpha t + 1}$.

Cartoon

Observed

ℓ_1-TV

Our method, $\varphi(t) = \frac{\alpha t}{\alpha t + 1}$

Summer School 2014: **Inverse Problem and Image Processing**

Tutorial: **Inverse modeling in inverse problems using optimization**

Outline

8. Fully smoothed ℓ_1 – TV

$$\boxed{\begin{array}{c}\boldsymbol{\mathcal{F}_v(u) = \Psi(u,v) + \beta\Phi(u), \quad \beta > 0}\\ \boldsymbol{\Psi(u,v) = \sum_{i=1}^{p}\psi(u[i]-v[i])} \quad \text{and} \quad \boldsymbol{\Phi(u) = \sum_i \varphi(|G_i u|)}\end{array}}$$

$\psi(\cdot) := \psi(\cdot, \alpha_1)$

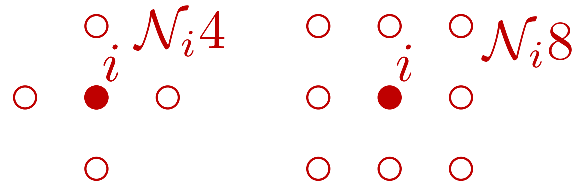
$\varphi(\cdot) := \varphi(\cdot, \alpha_2)$

$(\alpha_1, \alpha_2) > 0$

$\{G_i \in \mathbb{R}^{1\times p}\}$ – forward discretization:

$\mathcal{N}4$ Only vertical and horizontal differences;

$\mathcal{N}8$ Diagonal differences are added.



(ψ, φ) belong to the *family of functions* $\theta(\cdot, \alpha) : \mathbb{R} \to \mathbb{R}$ satisfying

H 1 For any $\alpha > 0$ fixed, $\theta(\cdot, \alpha)$ is $\mathcal{C}^{s\geqslant 2}$-continuous, even and $\theta''(t, \alpha) > 0$, $\forall\, t \in \mathbb{R}$.

H 2 For any $\alpha > 0$ fixed, $|\theta'(t, \alpha)| < 1$ and for $t > 0$ fixed, it is strictly decreasing in $\alpha > 0$

$$\alpha > 0 \quad \Rightarrow \quad \lim_{t\to\infty} \theta'(t, \alpha) = 1$$

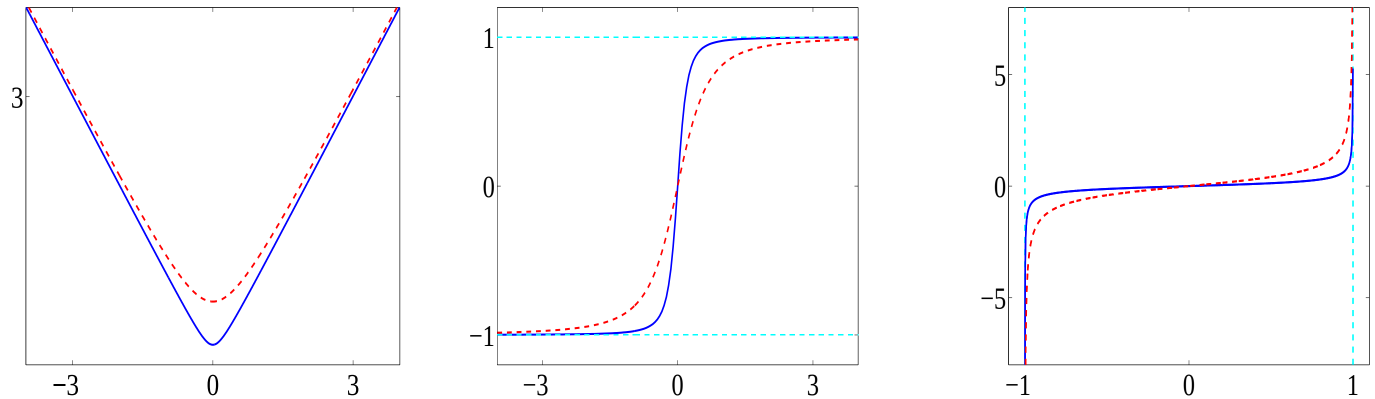
$$t \in \mathbb{R} \quad \Rightarrow \quad \lim_{\alpha\to 0} \theta'(t, \alpha) = 1 \quad \text{and} \quad \lim_{\alpha\to\infty} \theta'(t, \alpha) = 0\ .$$

$$\theta'(t, \alpha) := \frac{d}{dt}\theta(t, \alpha)$$

$\Rightarrow$ $\mathcal{F}_v$ is a fully smoothed ℓ_1 – TV energy.

	θ	θ'
f1	$\sqrt{t^2+\alpha}$	$\dfrac{t}{\sqrt{t^2+\alpha}}$
f2	$\alpha \log\left(\cosh\left(\dfrac{t}{\alpha}\right)\right)$	$\tanh\left(\dfrac{t}{\alpha}\right)$
f3	$\lvert t\rvert - \alpha \log\left(1+\dfrac{\lvert t\rvert}{\alpha}\right)$	$\dfrac{t}{\alpha+\lvert t\rvert}$

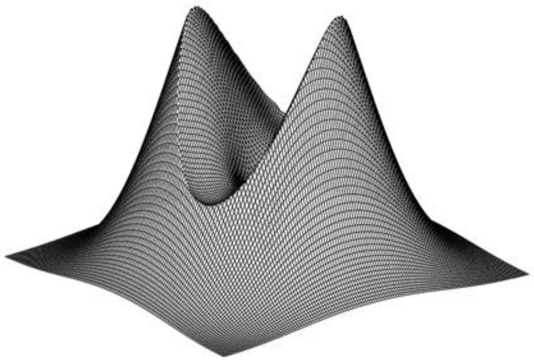
Choices for $\theta(\cdot,\alpha)$ obeying H1 and H2. When $\alpha \searrow 0$, $\theta(\cdot,\alpha)$ becomes stiff near the origin.



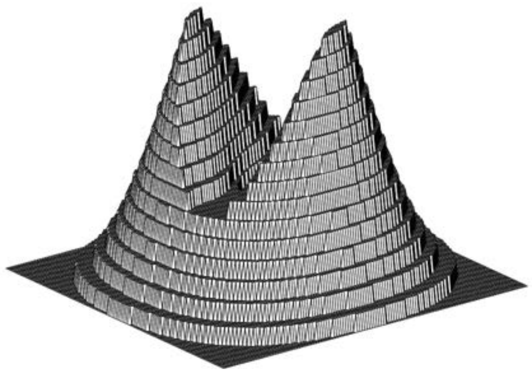
$\theta(t) = \sqrt{t^2+\alpha}$ $\quad$ $\theta'(t) = \frac{t}{\sqrt{t^2+\alpha}}$ $\quad$ $(\theta')^{-1}(y) = y\sqrt{\frac{\alpha}{1-y^2}}$

Plots of f1 for $\boldsymbol{\alpha = 0.05}$ (——) and for $\boldsymbol{\alpha = 0.5}$ (– – –).

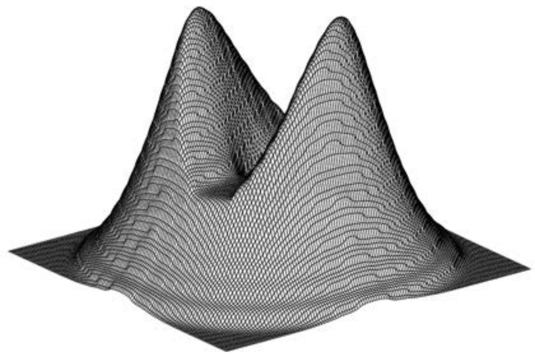
The minimizers $\hat{u}$ of $\mathcal{F}_v$ can decrease the quantization noise



Real-valued original



v quantized on $\{0, \cdots, 15\}$



Restored $\hat{u}$

[Nikolova, Wen, R. Chan 12]

- For any $\beta > 0$, $\mathcal{F}_v(\mathbb{R}^p)$ has a unique minimizer function $\mathcal{U} : \mathbb{R}^p \to \mathbb{R}^p$ which is $\mathcal{C}^{s-1}$.

Define $\mathcal{G} := \bigcup_{i=1}^{p} \bigcup_{j=1}^{p} \Big\{ g \in \mathbb{R}^{1\times p} \ : \ g[i] = -g[j] = 1, \ i \neq j, \ g[k] = 0 \text{ if } k \notin \{i,j\} \Big\}$

All difference operators G_i belong to $\mathcal{G}$.

$$N_{\mathcal{G}} := \bigcup_{g\in\mathcal{G}} \Big\{ v \in \mathbb{R}^p \ : \ g\,\mathcal{U}(v) = 0 \Big\} \quad \text{and} \quad N_I := \bigcup_{i=1}^{p} \bigcup_{j=1}^{p} \Big\{ v \in \mathbb{R}^p \ : \ \mathcal{U}_i(v) = v[j] \Big\}$$

Question **37** How to interpret the sets $N_{\mathcal{G}}$ and N_I?

- The sets $N_{\mathcal{G}}$ and N_I are closed in $\mathbb{R}^p$ and obey

$$\mathbb{L}^p(N_{\mathcal{G}}) = 0 \quad \text{and} \quad \mathbb{L}^p(N_I) = 0$$

The property is true for any $\beta > 0$ and $(\alpha_1, \alpha_2) > 0$.

- $\mathbb{R}^p \setminus (N_{\mathcal{G}} \cup N_I)$ is open and dense in $\mathbb{R}^p$.

  The elements of $(N_{\mathcal{G}} \cup N_I)$ are highly exceptional in $\mathbb{R}^p$.

- The minimizers $\hat{u}$ of $\mathcal{F}_v$ generically satisfy $\hat{u}[i] \neq \hat{u}[j]$ for any (i,j) such that $i \neq j$ and $\hat{u}[i] \neq v[j]$ for any (i,j).

The minimizers $\hat{u}$ of $\mathcal{F}_v$ have pixel values that are different from each other and different from any data pixel.

Question **38** Describe the consequences if $\ell_1 - \mathrm{TV}$ is approximated by a smooth function like $\mathcal{F}_v$.

Recall the illustration on p. 24 and the results in section 3 (p. 26) and section 4 (p. 35).

Further... [Bauss, Nikolova, Steidl 13]

- For any $\alpha_1 > 0$ fixed, there is an inverse function $\boldsymbol{(\psi')^{-1}(\cdot, \alpha_1)} : (-1, 1) \to \mathbb{R}$ which is odd, $\mathcal{C}^{s-1}$ and strictly increasing.

  $\alpha_1 \mapsto (\psi')^{-1}(y, \alpha_1)$ is also strictly increasing on $(0, +\infty)$, for any $y \in (0, 1)$.

- Set $\boldsymbol{\eta := \|G\|_1}$. Then

$$\boldsymbol{\beta\eta < 1 \quad \Rightarrow \quad \|\hat{u} - v\|_\infty \leqslant (\psi')^{-1}(\beta\eta, \alpha_1) \quad \forall\, v \in \mathbb{R}^p}$$

- Also, $\boldsymbol{\|\hat{u} - v\|_\infty \nearrow (\psi')^{-1}(\beta\eta, \alpha_1)}$ as $\boldsymbol{\alpha_2 \searrow 0}$.

Full control on the bound $\boldsymbol{\|\hat{u} - v\|_\infty}$.

Question **39** Can you suggest applications where the properties of $\mathcal{F}_v$ are important?

Exact histogram specification

- v – input digital gray value $m \times n$ image / stored as an $p := mn$ vector
- $v[i] \in \{0, \cdots, L-1\} \quad \forall\, i \in \{1, \cdots, p\}$ 8-bit image $\Rightarrow \quad L = 256$
- Histogram of v: $H_v[k] = \frac{1}{p}\#\big\{v[i] = k \;:\; i \in \{1, \cdots, p\}\big\} \quad \forall\, k \in \{0, \cdots, L-1\}$
- Target histogram: $\zeta = (\zeta[1], \cdots, \zeta[L])$
- Goal of histogram specification (HS): convert v into $\hat{u}$ so that $H_{\hat{u}} = \zeta$

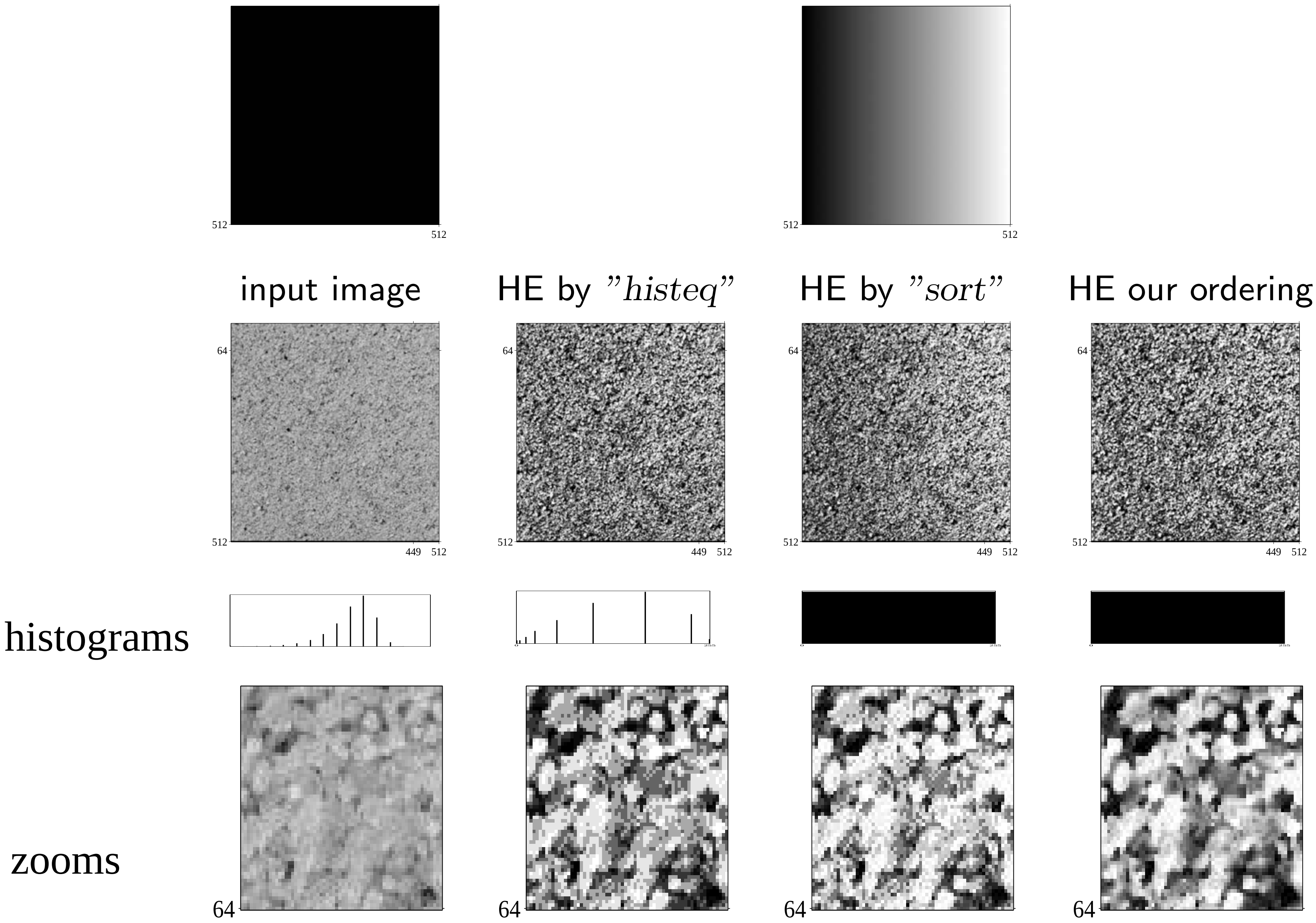
  order the pixels in v: $i \prec j$ if $v[i] < v[j]$

$$\underbrace{i_1 \prec i_2 \prec \cdots \prec i_{\zeta[1]}}_{\zeta[1]} \prec \cdots \prec \underbrace{i_{p-\zeta[L]+1} \prec \cdots \prec i_p}_{\zeta[L-1]}$$

- Ill-posed problem for digital (quantized) images since $p \gg L$
- An issue: obtain a meaningful total strict ordering of all pixels in v

Histogram equalization is a particular case of HS where $\zeta[k] = p/L \quad \forall\, k \in \{0, \cdots L-1\}$

Histogram Equalization (HE) using Matlab and our ordering



Modern sorting algorithms

For any pixel $v[i]$, extract K auxiliary information, $a_k[i]$, $k \in \{1, \cdots, K$, from v. Set $a_0 := v$. Then

$$i \prec j \quad \text{if} \quad v[i] \leqslant v[j] \ \text{ and } \ a_k[i] < a_k[j] \ \text{ for some } \ k \in \{0, \cdots, K\}.$$

Local Mean Algorithm (LM) [Coltuc, Bolon, Chassery 06]

– If two pixels are equal and their local mean is the same, take a larger neighborhood.

– The procedure smooths edges and sorting often fails.

Wavelet Approach (WA) [Wan, Shi 07]

– Use wavelet coefficients from different subbands to order the pixels.

– Heavy and high level of failure.

Specialized variational approach (SVA) [Nikolova, Wen and R. Chan 12]

– Minimize $\mathcal{F}_v$ for a parameter choice yielding $\|\hat{u} - v\|_\infty \lessapprox 0.1$.

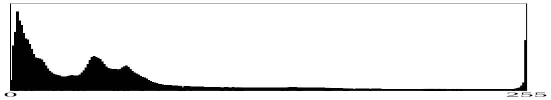
– Almost no failure, faithful order and fast algorithm. [Nikolova 13]

Some results using $\mathcal{F}_v$ for color image enhancement

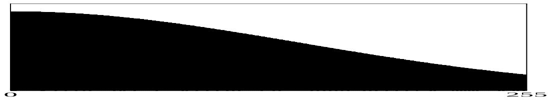
New fast histogram based color enhancement algorithm. [Nikolova, Steidl 14]

- [NS 14] M. Nikolova and G. Steidl, "Fast Hue and Range Preserving Histogram Specification: Theory and New Algorithms for Color Image Enhancement", *IEEE Trans. Image Process.*, to appear.
- [HYL 11] J. H. Han, S. Yang, and B. U. Lee, A novel 3-D color histogram equalization method with uniform 1-D gray scale histogram, IEEE Trans. Image Process., vol. 20, no. 2, pp. 506-512, Feb. 2011.
- [BCPR 07] M. Bertalmío, V. Caselles, E. Provenzi, and A. Rizzi, "Perceptual color correction through variational techniques", *IEEE Trans. Image Process.*, vol. 16, no. 4, pp. 1058–1072, Apr. 2007.
- [APBC 09] R. Palma-Amestoy, E. Provenzi, M. Bertalmío, and V. Caselles, "A perceptually inspired variational framework for color enhancement", *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 31, no. 3, pp. 458–474, 2009.
- [ACE G 12] P. Gertreuer, "Automatic color enhancement (ACE) and its fast implementation", *Image Processing On Line*, DOI: 10.5201/ipol.2012.g-ace, vol. 2012

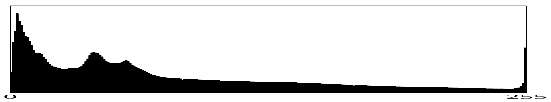
club (1800×3200)



Hist.-based [NS 14]

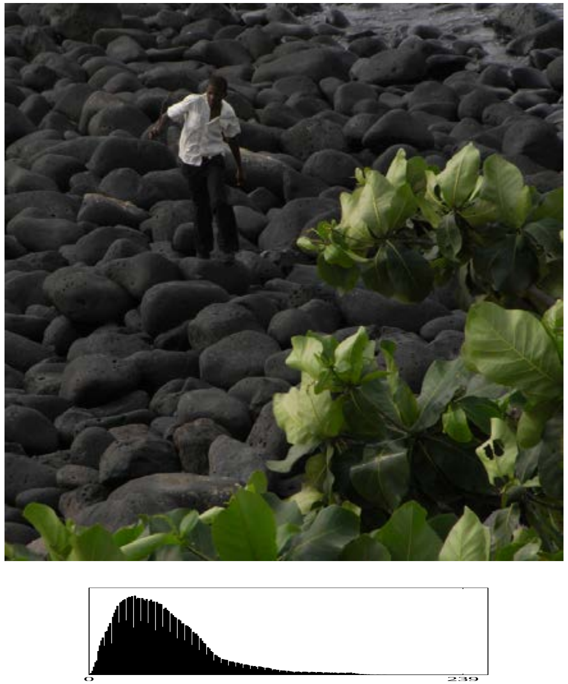


Hist.-based [NS 14]



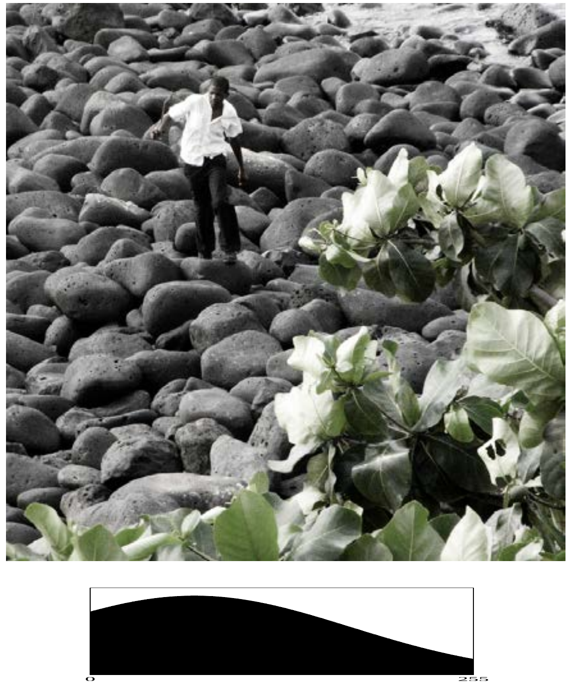
Perceptual [APBC 09]

boy-on-stones (800×800)



Hist.-based [NS 14]

Hist.-based [HYL 11]

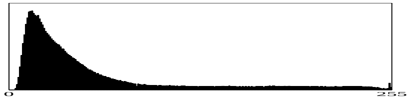


Perceptual [BCPR 07]

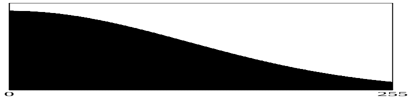
Perceptual [APBC 09]

ACE [G 12]

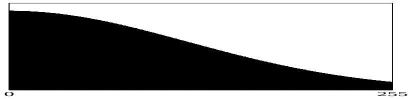
orchid (768×1024)



Hist.-based [NS 14]



Hist.-based [HYL 11]



Perceptual [APBC 09]

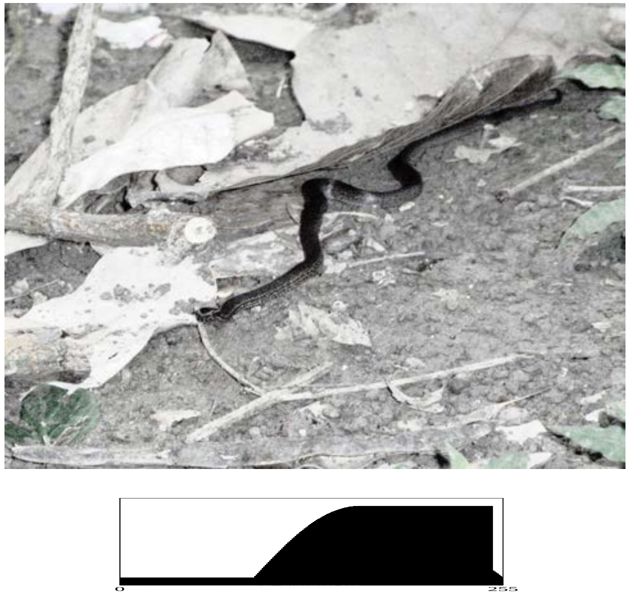
Perceptual [BCPR 07]

ACE [G 12]

Input “orchid” with a bad flashlight effect.

snake (1000×1000)

Hist.-based [HYL 11]



Hist.-based [NS 14]

Goal – enhance the snake.

Summer School 2014: **Inverse Problem and Image Processing**

Tutorial: **Inverse modeling in inverse problems using optimization**

Outline

9 Inverse modeling and Bayesian MAP

[Nikolova 07]

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 $V = v$:

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

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:

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

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

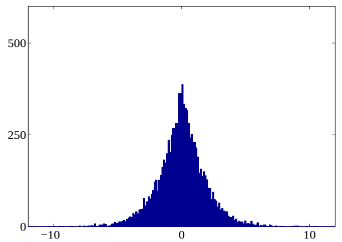
Example: MAP shrinkage *[Simoncelli99, Belge-Kilmer00, Antoniadis02]*

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

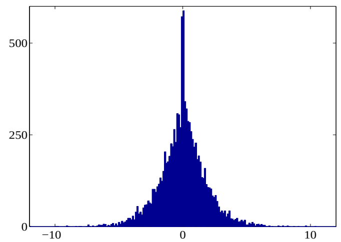
  Experiments have shown that $\boldsymbol{\alpha \in (0,1)}$ for many real-world images
- MAP restoration $\Leftrightarrow$ $\boldsymbol{\hat{x}[i] = \arg\min_{t \in \mathbb{R}} \big((t - y[i])^2 + \lambda |t|^{\alpha}\big), \ \forall i}$

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

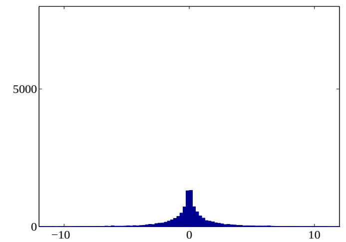
(1) sample $x \sim f_X$ and $n \sim \mathcal{N}(0, \sigma^2)$, (2) form $y = x + n$, (3) compute the **true** MAP $\hat{x}$



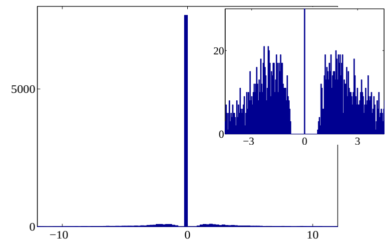
GG,$\alpha = 1.2, \lambda = \frac{1}{2}$



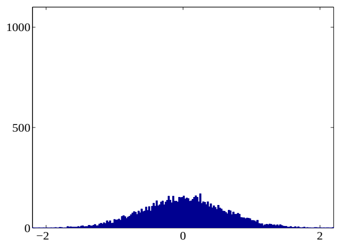
The true MAP $\hat{x}$



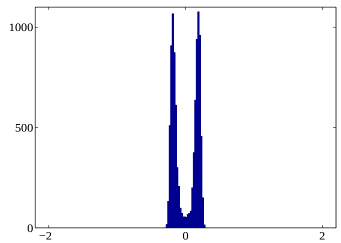
GG, $\alpha = \frac{1}{2}$, $\lambda = 2$



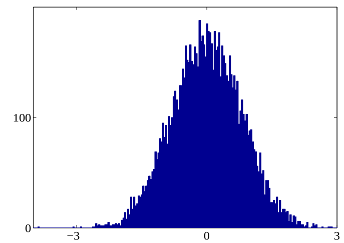
True MAP $\hat{x}$



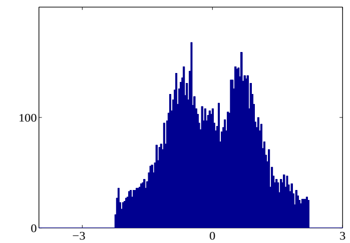
Noise $\mathcal{N}(0, \sigma^2)$



Noise estimate



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



Noise estimate

Theoretical explanations

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

The analytical results on $\boldsymbol{\hat{u}} = \arg\min\limits_{\boldsymbol{u}} \boldsymbol{\mathcal{F}_v(u)}$ =MAP yield:

- $\boldsymbol{f_U}$ continuous and non-smooth at 0, $\boldsymbol{\varphi'(0^+) > 0}$ Ch. 3, p. 26

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

  The effective prior: $G_i \hat{u} = 0$ for many i. (e.g. locally constant images)

- $\boldsymbol{f_N}$ continuous and nonsmooth at 0, $\boldsymbol{\psi'(0^+) > 0}$ Ch. 4, p. 35

  $v \in \mathcal{O}_{\hat{h}} \Rightarrow \left[a_i\, \hat{u} = v_i, \ \forall i \in \hat{h}\right] \Rightarrow \mathbf{Pr}\Big(\boldsymbol{a_i\, \hat{u} = v_i, \forall i \in \hat{h}}\Big) \geqslant \mathbf{Pr}(\boldsymbol{V \in \mathcal{O}_{\hat{h}}}) > \mathbf{0}$

  The effective model: there are uncorrupted data entries.

- $\boldsymbol{-\ln f_U}$ (resp., φ) continuous and nonconvex $\Rightarrow \mathbf{Pr}\Big(\boldsymbol{\theta_0 < \|G_i \hat{U}\| < \theta_1}\Big) = \mathbf{0}, \ \forall \boldsymbol{i}$

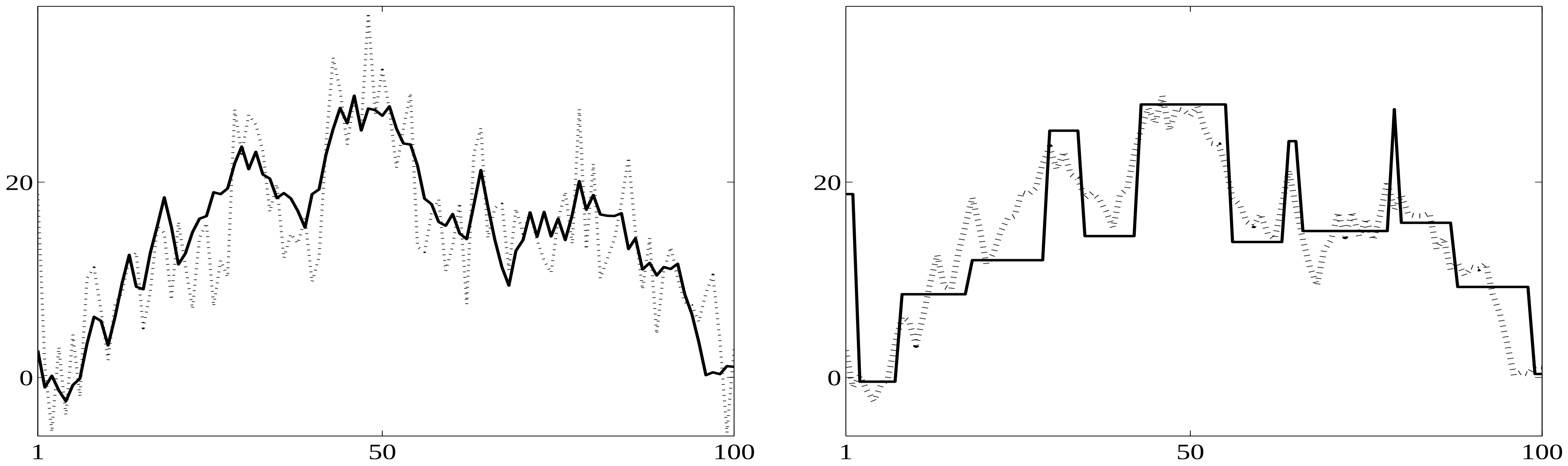
  The effective prior: edges.

- $\boldsymbol{-\ln f_U}$ nonconvex, nonsmooth at 0, $\boldsymbol{\varphi'(0^+) > 0}$ and $\boldsymbol{\varphi'' \leqslant 0}$ Ch. 6, p. 54

  $\Rightarrow \ \mathbf{Pr}\Big(\boldsymbol{\|G_i \hat{u}\| = 0}\Big) > \mathbf{0}$ and $\mathbf{Pr}\Big(\boldsymbol{0 < \|G_i \hat{u}\| < \theta_1}\Big) = \mathbf{0}$

Illustration

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



Original u_o (—) by f 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

Knowledge on the features of the minimizers enables new energies yielding appropriate solutions to be conceived

`` We're in Act I of a digital revolution.''

Jay Cassidy (film editor at Mathematical Technologies Inc.)

Thank you!

10 Some References

10 Some References

1. Alliney S (1992) Digital filters as absolute norm regularizers. IEEE Trans Signal Process SP-40:1548–1562
2. Ambrosio L, Fusco N, Pallara D (2000) Functions of bounded variation and free discontinuity Problems. Oxford Mathematical Monographs, Oxford University Press
3. Antoniadis A, Fan J (2001) Regularization of wavelet approximations. J Acoust Soc Am 96: 939–967
4. Aubert G, Kornprobst P (2006) Mathematical problems in image processing, 2nd edn. Springer, Berlin
5. Aubert G, Aujol J.-F. (2008) A variational approach to remove multiplicative noise, SIAM J. on Appl. Maths., 68:925-946
6. Aujol J-F, Gilboa G, Chan T, Osher S (2006) Structure-texture image decomposition - modeling, algorithms, and parameter selection. Int J Comput Vis 67:111–136
7. Auslender A and Teboulle M. (2003) Asymptotic Cones and Functions in Optimization and Variational Inequalities. Springer, New York
8. Attouch, H., Bolte, J. and Svaiter, B. F. (2013) Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forwardbackward splitting, and regularized GaussSeidel methods. Math. Program. 137:91-129
9. Bar L, Brook A, Sochen N, Kiryati N (2007) Deblurring of color images corrupted by impulsive noise. IEEE Trans Image Process 16:1101–1111
10. Bar L, Kiryati N, Sochen N (2006) Image deblurring in the presence of impulsive noise, International. J Comput Vision 70:279–298

11. Bar L, Sochen N, Kiryati N (2005) Image deblurring in the presence of salt-and-pepper noise. In Proceeding of 5th international conference on scale space and PDE methods in computer vision, ser LNCS, vol 3439, pp 107–118
12. Bae E, Yuan J, Tai X.-C. (2011) Global minimization for continuous multiphase partitioning problems using a dual approach, Int. J. Comput. Vis., 92:112–129.
13. Baus F, Nikolova M, Steidl G. (2014) Fully smoothed ℓ_1-TV models: Bounds for the minimizers and parameter choice, J Math Imaging Vis 48:295-307
14. Belge M, Kilmer M, Miller E (2000) Wavelet domain image restoration with adaptive edge-preserving regularization. IEEE Trans Image Process 9:597–608
15. Besag JE (1974) Spatial interaction and the statistical analysis of lattice systems (with discussion). J Roy Stat Soc B 36:192–236
16. Besag JE (1989) Digital image processing : towards Bayesian image analysis. J Appl Stat 16:395–407
17. Black M, Rangarajan A (1996) On the unification of line processes, outlier rejection, and robust statistics with applications to early vision. Int J Comput Vis 19:57–91
18. Blake A, Zisserman A (1987) Visual reconstruction. MIT Press, Cambridge
19. Bobichon Y, Bijaoui A (1997) Regularized multiresolution methods for astronomical image enhancement. Exp Astron 7:239–255
20. Bolte, J., Sabach, S. and Teboulle, M. (2013) Proximal alternating linearized minimization for nonconvex and nonsmooth problems, Math. Program. Ser. A, online
21. Bouman C, Sauer K (1993) A generalized Gaussian image model for edge-preserving map estimation. IEEE Trans Image Process 2:296–310

22. Bouman C, Sauer K (1996) A unified approach to statistical tomography using coordinate descent optimization. IEEE Trans Image Process 5:480–492
23. Bredies K, Kunich K, Pock T (2010) Total generalized variation. SIAM J Imaging Sci 3(3): 480–491.
24. Bredies K, and Holler, M. (2014) Regularization of linear inverse problems with total generalized variation. J Inverse and Ill-posed Problems DOI: 10.1515/jip-2013-0068.
25. Candès EJ, Donoho D, Ying L (2005) Fast discrete curvelet transforms. SIAM Multiscale Model Simul 5:861–899
26. Candès EJ, Guo F (2002) New multiscale transforms, minimum total variation synthesis. Applications to edge-preserving image reconstruction. Signal Process 82:1519–1543
27. Catte F, Coll T, Lions PL, Morel JM (1992) Image selective smoothing and edge detection by nonlinear diffusion (I). SIAM J Num Anal 29:182–193
28. Chambolle A (2004) An Algorithm for Total Variation Minimization and Application. J. Math Imaging Vis 20:89–98
29. Chan T, Esedoglu S, Nikolova M. (2006) Algorithms for finding global minimizers of image segmentation and denoising models. SIAM J. Appl. Math. 66:1632-1648
30. Chan T, Esedoglu S (2005) Aspects of total variation regularized L^1 function approximation. SIAM J Appl Math 65:1817–1837
31. Chan TF, Wong CK (1998) Total variation blind deconvolution. IEEE Trans Image Process 7:370–375
32. Charbonnier P, Blanc-Féraud L, Aubert G, Barlaud M (1997) Deterministic edge-preserving regularization in computed imaging. IEEE Trans Image Process 6:298–311
33. Chellapa R, Jain A (1993) Markov random fields: theory and application. Academic, Boston

34. Chen X., M. Ng, Zhang C. (2012) Non-Lipschitz ℓ_p-Regularization and Box Constrained Model for Image Restoration. IEEE Trans Image Process 21:4709–4721.

35. Chesneau C, Fadili J, Starck J-L (2010) Stein block thresholding for image denoising. Appl. Comput. Harmon. Anal. 28:67-88

36. Ciarlet PG (1989) Introduction to numerical linear algebra and optimization. Cambridge University Press, Cambridge

37. Coifman RR, Sowa A (2000) Combining the calculus of variations and wavelets for image enhancement. Appl Comput Harmon Anal 9:1–18

38. Do MN, Vetterli M (2005) The contourlet transform: an efficient directional multiresolution image representation. IEEE Trans Image Process 15:1916–1933

39. Dobson D, Santosa F (1996) Recovery of blocky images from noisy and blurred data. SIAM J Appl Math 56:1181–1199

40. Donoho DL, Johnstone IM (1994) Ideal spatial adaptation by wavelet shrinkage. Biometrika 81:425–455

41. Donoho DL, Johnstone IM (1995) Adapting to unknown smoothness via wavelet shrinkage. J Acoust Soc Am 90:1200–1224

42. Dontchev AL, Zollezi T (1993) Well-posed optimization problems. Springer, New York

43. Durand S, Froment J (2003) Reconstruction of wavelet coefficients using total variation minimization. SIAM J Sci Comput 24:1754–1767

44. Durand S, Nikolova M (2006) Stability of minimizers of regularized least squares objective functions I: study of the local behavior. Appl Math Optim 53:185–208, II: Study of the global behaviour. Appl Math Optim 53:259–277

45. Durand S, Nikolova M (2007) Denoising of frame coefficients using ℓ^1 data-fidelity term and edge-preserving regularization. SIAM J Multiscale Model Simulat 6:547–576
46. Durand S, Fadili J, Nikolova M. (2010) Multiplicative noise removal using L1 fidelity on frame coefficients. J. Math Imaging and Vision, 36:201-226
47. Duval V, Aujol J-F, Gousseau Y (2009) The TVL1 model: a geometric point of view. SIAM J Multiscale Model Simulat 8:154–189
48. Ekeland I, Temam R (1976) Convex analysis and variational problems. North-Holland/SIAM, Amsterdam
49. Fessler F (1996) Mean and variance of implicitly defined biased estimators (such as penalized maximum likelihood): applications to tomography. IEEE Trans Image Process 5:493–506
50. Fiacco A, McCormic G (1990) Nonlinear programming. Classics in applied mathematics. SIAM, Philadelphia
51. Geman D (1990) Random fields and inverse problems in imaging, vol 1427, École d'Été de Probabilités de Saint-Flour XVIII - 1988, Springer, Lecture notes in mathematics, pp 117–193
52. Geman D, Reynolds G (1992) Constrained restoration and recovery of discontinuities. IEEE Trans Pattern Anal Mach Intell PAMI-14:367–383
53. Geman D, Yang C (1995) Nonlinear image recovery with half-quadratic regularization. IEEE Trans Image Process IP-4:932–946
54. Geman S, Geman D (1984) Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. IEEE Trans Pattern Anal Mach Intell PAMI-6:721–741
55. Green PJ (1990) Bayesian reconstructions from emission tomography data using a modified EM algorithm. IEEE Trans Med Imaging MI-9:84–93

56. Lustig M, Donoho D, Santos JM and Pauly LM (2008) Compressed Sensing MRI: a look how CS can improve our current imaging techniques: IEEE Signal Proc. Magazine. 25:72–82.
57. Haddad A, Meyer Y (2007) Variational methods in image processing, in "Perspective in Nonlinear Partial Differential equations in Honor of Haïm Brezis," Contemp Math (AMS) 446:273–295
58. Hiriart-Urruty J-B, Lemaréchal C (1996) Convex analysis and minimization algorithms, vols I, II. Springer, Berlin
59. Hofmann B (1986) Regularization for applied inverse and ill posed problems. Teubner, Leipzig
60. Kak A, Slaney M (1987) Principles of computerized tomographic imaging. IEEE Press, New York
61. Keren D, Werman M (1993) Probabilistic analysis of regularization. IEEE Trans Pattern Anal Mach Intell PAMI-15:982–995
62. Li S (1995) Markov random field modeling in computer vision, 1st edn. Springer, New York
63. Li SZ (1995) On discontinuity-adaptive smoothness priors in computer vision. IEEE Trans Pattern Anal Mach Intell PAMI-17:576–586
64. Luisier F, Blu T (2008) SURE-LET multichannel image denoising: interscale orthonormal wavelet thresholding. IEEE Trans Image Process 17:482–492
65. Malgouyres F (2002) Minimizing the total variation under a general convex constraint for image restoration. IEEE Trans Image Process 11:1450–1456
66. Morel J-M, Solimini S (1995) Variational methods in image segmentation. Birkhäuser, Basel
67. Morozov VA (1993) Regularization methods for ill posed problems. CRC Press, Boca Raton
68. Moulin P, Liu J (1999) Analysis of multiresolution image denoising schemes using generalized Gaussian and complexity priors. IEEE Trans Image Process 45:909–919

69. Moulin P, Liu J (2000) Statistical imaging and complexity regularization. IEEE Trans Inf Theory 46:1762–1777
70. Mumford D, Shah J (1989) Optimal approximations by piecewise smooth functions and associated variational problems. Commun Pure Appl Math 42:577–684
71. Nashed M, Scherzer O (1998) Least squares and bounded variation regularization with nondifferentiable functional. Numer Funct Anal Optim 19:873–901
72. Nikolova M (2000) Local strong homogeneity of a regularized estimator. SIAM J Appl Math 61:633–658
73. Nikolova M (2000) Thresholding implied by truncated quadratic regularization. IEEE Trans Image Process 48:3437–3450
74. Nikolova M (2002) Minimizers of cost-functions involving nonsmooth data-fidelity terms. Application to the processing of outliers. SIAM J Num Anal 40:965–994
75. Nikolova M (2004) A variational approach to remove outliers and impulse noise. J Math Imaging Vis 20:99–120
76. Nikolova M (2004) Weakly constrained minimization. Application to the estimation of images and signals involving constant regions. J Math Imaging Vis 21:155–175
77. Nikolova M (2005) Analysis of the recovery of edges in images and signals by minimizing nonconvex regularized least-squares. SIAM J Multiscale Model Simulat 4:960–991
78. Nikolova M (2007) Analytical bounds on the minimizers of (nonconvex) regularized least- squares. AIMS J Inverse Probl Imag 1:661–677
79. Nikolova M (2007) Model distortions in Bayesian MAP reconstruction, AIMS J Inverse Probl Imag, 1:399–422

80. Nikolova M, Ng M, Tam CP (2013) On ℓ_1 Data Fitting and Concave Regularization for Image Recovery. SIAM J. Sci. Comput 35:397-430
81. Nikolova M (2009) One-iteration dejittering of digital video images, J Vis Commun Image R, 20:254–274
82. Nikolova M (2013) Description of the minimizers of least squares regularized with ℓ_0 norm. Uniqueness of the global minimizer. SIAM J. Imaging Sci., 6: 904–937
83. Nikolova M, Wen YW, Chan R (2013) Exact histogram specification for digital images using a variational approach. J Math Imaging Vis 46: 309-325
84. Nikolova M, Steidl G (2014) Fast Hue and Range Preserving Histogram Specification: Theory and New Algorithms for Color Image Enhancement. IEEE Trans Image Process (to appear)
85. Papadakis N., Yildizoglu R., Aujol J-F. and Caselles V. (2013) High-dimension multi-label problems: convex or non convex relaxation? SIAM J. Imaging Sci. 6:2603–2639
86. Perona P, Malik J (1990) Scale-space and edge detection using anisotropic diffusion. IEEE Trans Pattern Anal Mach Intell PAMI-12:629–639
87. Potts R. B. (1952) Some generalized order-disorder transformations, Proc. Cambridge Philos. Soc., 48:106–109
88. Rockafellar RT, Wets JB (1997) Variational analysis. Springer, New York
89. Rudin L, Osher S, Fatemi C (1992) Nonlinear total variation based noise removal algorithm. Physica 60 D:259–268
90. Robini, M. and Magnin, I. (2010) Optimization by stochastic continuation. SIAM J. Imaging Sci., 3:1096-1121

91. Robini, M. and Reissman, P.-J. (2013) From simulated annealing to stochastic continuation: a new trend in combinatorial optimization. J. Global Optim 56:185–215
92. Sauer K, Bouman C (1993) A local update strategy for iterative reconstruction from projections. IEEE Trans Signal Process SP-41:534–548
93. Scherzer O, Grasmair M, Grossauer H, Haltmeier M, Lenzen F (2009) Variational problems in imaging. Springer, New York
94. Tautenhahn U (1994) Error estimates for regularized solutions of non-linear ill posed problems. Inverse Probl 10:485–500
95. Tikhonov A, Arsenin V (1977) Solutions of ill posed problems, Winston, Washington
96. Vogel C (2002) Computational methods for inverse problems. Frontiers in applied mathematics series, vol 23. SIAM, New York
97. Welk M, Steidl G, Weickert J (2008) Locally analytic schemes: a link between diffusion filtering and wavelet shrinkage. Appl Comput Harmon Anal 24:195–224
98. Winkler G (2006) Image analysis, random fields and Markov chain Monte Carlo methods. A mathematical introduction. Applications of mathematics, 2nd edn, vol 27. Stochastic models and applied probability. Springer, Berlin
99. Yuan J, Bae E, Tai X.-C. (2010) A study on continuous max-flow and min-cut approaches. Comp Vis and Pattern Recognition: 2217-2224
100. Yuan J, Bae E, Tai X.-C., Boykov Y. (2010) A continuous max-flow approach to Potts model. Computer VisionECCV 2010:379-392.