Solution Properties and Inverse Modeling in Variational Imaging

Mila Nikolova

CMLA, ENS Cachan, CNRS, University Paris Saclay, France

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

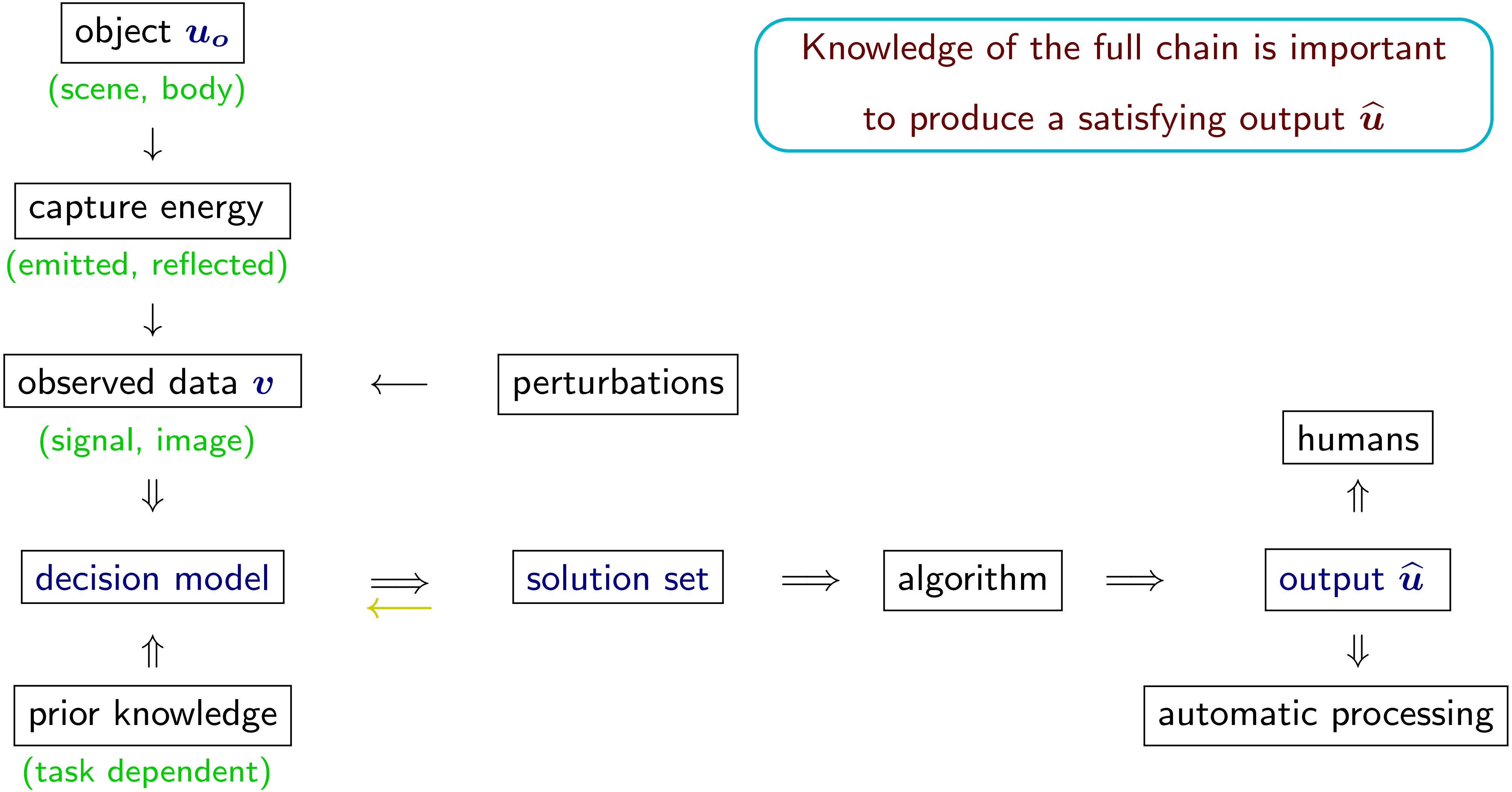
nikolova@cmla.ens-cachan.fr nikolova@noos.fr

Mathematical and Numerical Methods in Image Processing

Summer School, 25 July - 5 August 2016

Berlin Mathematical School

Knowledge of the full chain is important to produce a satisfying output $\widehat{\boldsymbol{u}}$



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

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

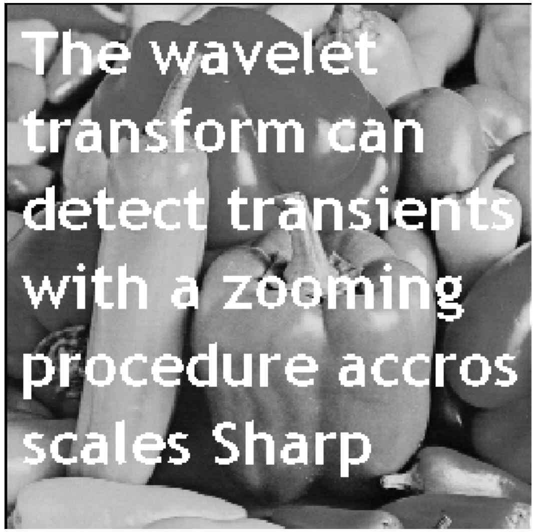
Processing tasks: $\widehat{\boldsymbol{u}} = \text{recover}(\boldsymbol{u_o})$ | $\widehat{\boldsymbol{u}} = \text{objects of interest}(\boldsymbol{u_o})$ | $\widehat{\boldsymbol{u}} = \text{classify}(\boldsymbol{u_o})$ | $(\cdots)$

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

Editing

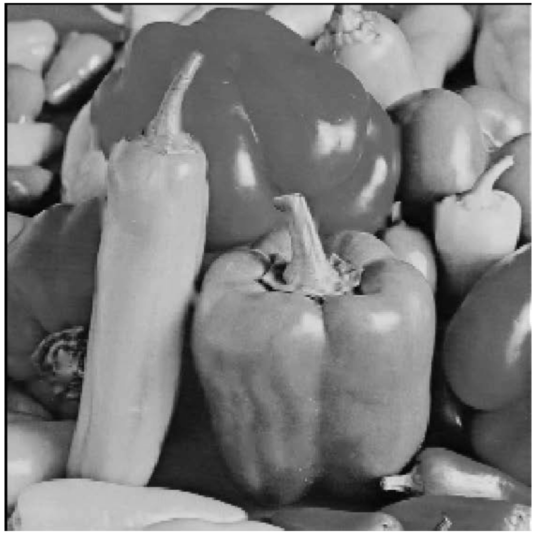
↓

[Pérez, Gangnet, Blake 04]



Inpainting

↓



[Chan, Steidl, Setzer 08]

Denoising

↓

[M. Lebrun, A. Buades and J.-M. Morel, 2113]

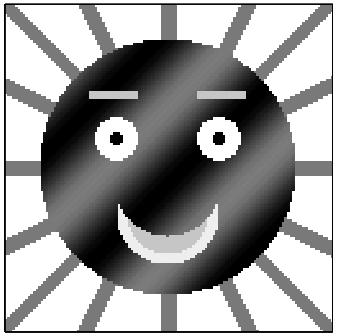
Image/signal processing tasks often require to solve **ill-posed inverse problems**

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

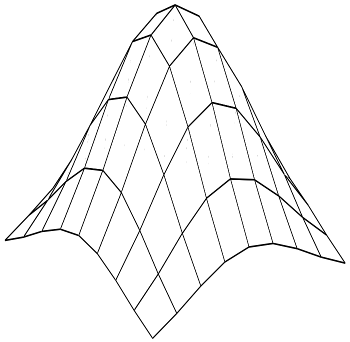
$\boldsymbol{A}$ is ill-conditioned $\equiv$ (nearly) noninvertible

Least-squares solution: $\widehat{\boldsymbol{u}} = \arg\min_{\boldsymbol{u}} \left\{ \|\boldsymbol{A}\boldsymbol{u} - \boldsymbol{v}\|^2 \right\}$

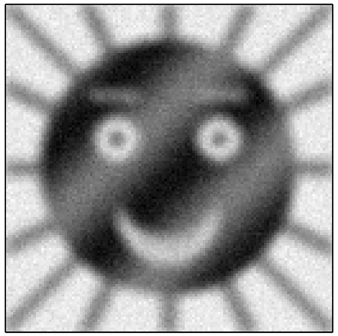
Tikhonov regularization: $\widehat{\boldsymbol{u}} := \arg\min_{\boldsymbol{u}} \left\{ \|\boldsymbol{A}\boldsymbol{u} - \boldsymbol{v}\|^2 + \boldsymbol{\beta} \sum_{\boldsymbol{i}} \|\boldsymbol{G_i}\boldsymbol{u}\|^2 \right\}$ for $\{\boldsymbol{G_i}\} \approx \boldsymbol{\nabla}$, $\boldsymbol{\beta} > \mathbf{0}$



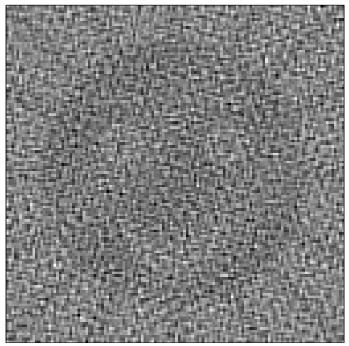
Original $\boldsymbol{u_o}$



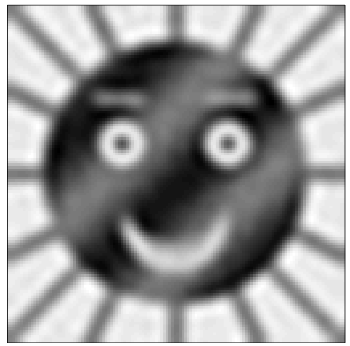
Blur $\boldsymbol{a}$



Data $\boldsymbol{v}$



$\widehat{\boldsymbol{u}}$: Least-squares



$\widehat{\boldsymbol{u}}$: Tikhonov

An ill-posed inverse problem

Example due to R.S.Wilson

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

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

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

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

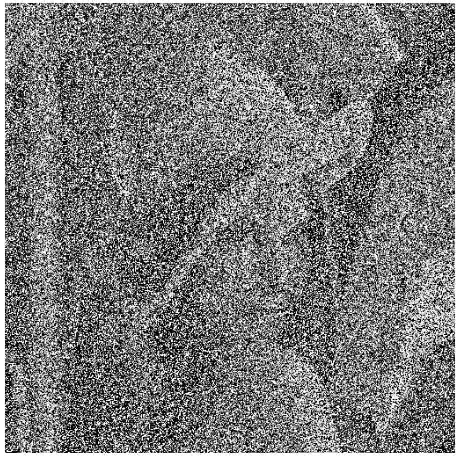
  $$\Rightarrow \quad \hat{u} = [\,9.2 \quad -12.6 \quad 4.5 \quad -1.1\,]^T$$

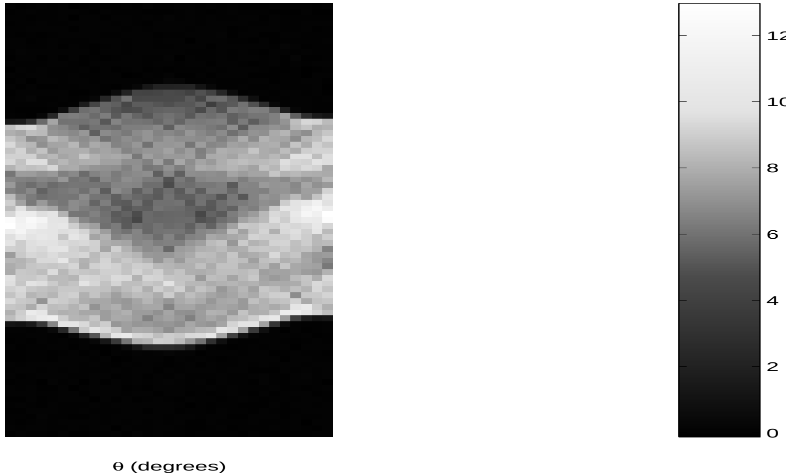
  Tikhonov regularization: $\hat{u} = \arg\min_{u \in \mathbb{R}^4} \mathcal{F}_v(u)$

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

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

Outline





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

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

Formulate your problem as the minimization (maximization) of a functional (an objective) $\mathcal{F}_v$ whose solution is the sought after signal/image

1 Variational regularization methods

u_o (unknown) $\quad v$ (data) $=$ Transform$(u_o)\bullet$ (perturbations)

solution $\hat{u}$

$$\hat{u} \begin{array}{l} \nearrow \text{ close to data production model } \Psi(u,v) \quad \text{(data-fidelity)} \\ \searrow \text{ coherent with priors and desiderata } \Phi(u) \quad \text{(prior – functional, constraint)} \end{array}$$

Combining models: $$\widehat{u} \in \arg\min_{u\in\Omega} \mathcal{F}_v(u) \qquad (\mathcal{P})$$

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

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

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

Data-fidelity models

Ψ (usually) models the production of data:

$$\boxed{\Psi = -\log\big(\mathsf{Likelihood}(\boldsymbol{v}|\boldsymbol{u})\big)}$$

Ψ involves a (linear) observation operator $\boldsymbol{A}$ (blur, projections, ...)– e.g. $v = Au_o + n$ (noise)

$(\mathcal{N})$	Gaussian noise ($n \sim \mathcal{N}(0, \sigma^2 I)$)	$\Rightarrow$	$\Psi(Au, v) = \frac{1}{2\sigma^2}\|Au - v\|_2^2$
$(\mathcal{L})$	Laplacian noise (centered, diversity b)	$\Rightarrow$	$\Psi(Au, v) = \frac{1}{b}\|Au - v\|_1$
(P)	Poisson observations	$\Rightarrow$	$\Psi(Au, v)) = \langle \mathbb{1}_q, Au\rangle - \langle v, \log(Au)\rangle, \quad Au > 0$
$(\mathcal{M})$	Multiplicative noise (K records)	$\Rightarrow$	$\Psi(Au, v)) = K\left\langle \mathbb{1}_q, \left(\log(Au) + \frac{v}{Au}\right)\right\rangle, \quad Au > 0$

Impulse noise: $\mathbb{P}(v_i = (Au_o)_i) = r, \quad \mathbb{P}(v_i = \gamma) = 1 - r$ where γ is random.

Remark 1.1 To deal with impulse noise, the Laplacian model $(\mathcal{L})$ is commonly used.

The information on $\boldsymbol{u_o}$ is implicitly contained in $\Psi(\cdot, v)$.
A good prior Φ is needed to extract the sought-after information ($\widehat{u}$) from the data (v).

Prior models, Regularizers

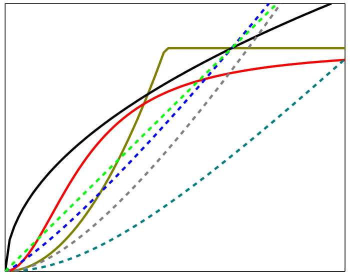
$\mathbf{\Phi}$ is a model for the sought-after $\widehat{\boldsymbol{u}}$, in restoration for the unknown $\boldsymbol{u}_o$

Ingredients: statistics, smoothness, edges, textures, special features, self-similarity...

- Bayesian approach: to model the interactions between samples
- Variational approach: PDE-based (anisotropic) to select good smoothers
- Among others...

Regularizers of the form $$\boxed{\Phi(u) = \sum_i \varphi_i(\|G_i u\|)}$$

$\boldsymbol{\varphi} : \mathbb{R}_+ \to \mathbb{R}_+$ potential function (PF), usually $\varphi_i = \varphi \ \forall\, i$. Examples $\Longrightarrow$
$\boldsymbol{\{G_i\}}$ — linear operators, ∇ is a discrete approximation of the gradient.



Some formulations

- Tikhonov $\Rightarrow$ $\{G_i\} \in \big\{I,\ \nabla,\ \nabla^2,\ \big(\nabla, \nabla^2\big)\ \big\}$, etc.
- Analysis $\Rightarrow$ $\{G_i\} = W$ for W a frame (e.g., a dictionary)
- Synthesis $\Rightarrow$ $A = BW$ and $\{G_i\} = I$ (here $u = W$(image) contains the coefficients)
- Hybrid $\Rightarrow$ $\{G_i\} = \nabla W^\dagger$ where $W^\dagger$ is a left inverse of W

Total Variation: $\mathrm{TV}(u) = \sum_i \|\,(\nabla u)_i\,\|_2$

Convex PFs	
$\varphi(\lvert t\rvert)$ is smooth at zero	$\varphi(\lvert t\rvert)$ is nonsmooth at zero
$\varphi(t) = t^\alpha,\ 1 < \alpha \leqslant 2$	$\varphi(t) = t$
$\varphi(t) = \sqrt{\alpha + t^2}$	
$\varphi(t) = \lvert t\rvert - \alpha \log\left(1 + \frac{\lvert t\rvert}{\alpha}\right)$	
$\varphi(t) = \begin{cases} t^2/(2\alpha) & \text{if} \quad \lvert t\rvert \leqslant \alpha, \\ \lvert t\rvert - \alpha/2 & \text{if} \quad \lvert t\rvert > \alpha \end{cases}$	

Nonconvex PFs	
$\varphi(\lvert t\rvert)$ is smooth at zero	$\varphi(\lvert t\rvert)$ is nonsmooth at zero
$\varphi(t) = \min\{\alpha t^2, 1\}$	$\varphi(t) = t^\alpha,\ 0 < \alpha < 1$
$\varphi(t) = \dfrac{\alpha t^2}{1 + \alpha t^2}$	$\varphi(t) = \dfrac{\alpha t}{1 + \alpha t}$
$\varphi(t) = \log(\alpha t^2 + 1)$	$\varphi(t) = \log(\alpha t + 1)$
$\varphi(t) = 1 - \exp(-\alpha t^2)$	$\varphi(t) = \begin{cases} 0 & \text{if} \quad t = 0 \\ 1 & \text{if} \quad t \neq 0 \end{cases}$

Commonly used PFs φ where $\alpha > 0$ is a parameter.

Some well known objective 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

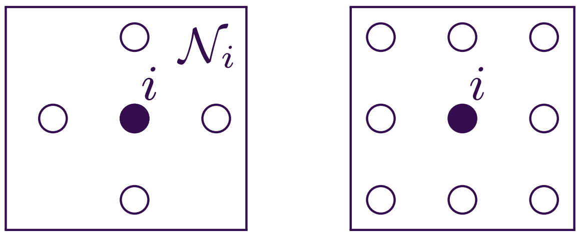
Potts model [Potts 52] (ℓ_0 semi-norm applied to differences):

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

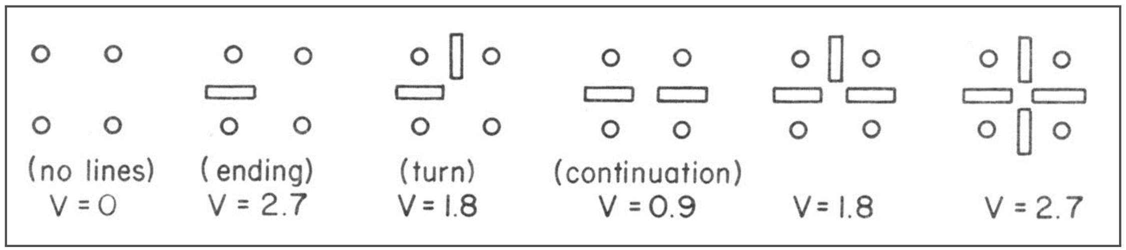
Markov Random Fields with Line Process [Geman, Geman 84]: $(\hat{u}, \hat{\ell}) = \arg\min_{u,\ell} \mathcal{F}_v(u, \ell)$

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

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



some possible neighbors $\mathcal{N}_i$



line model

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

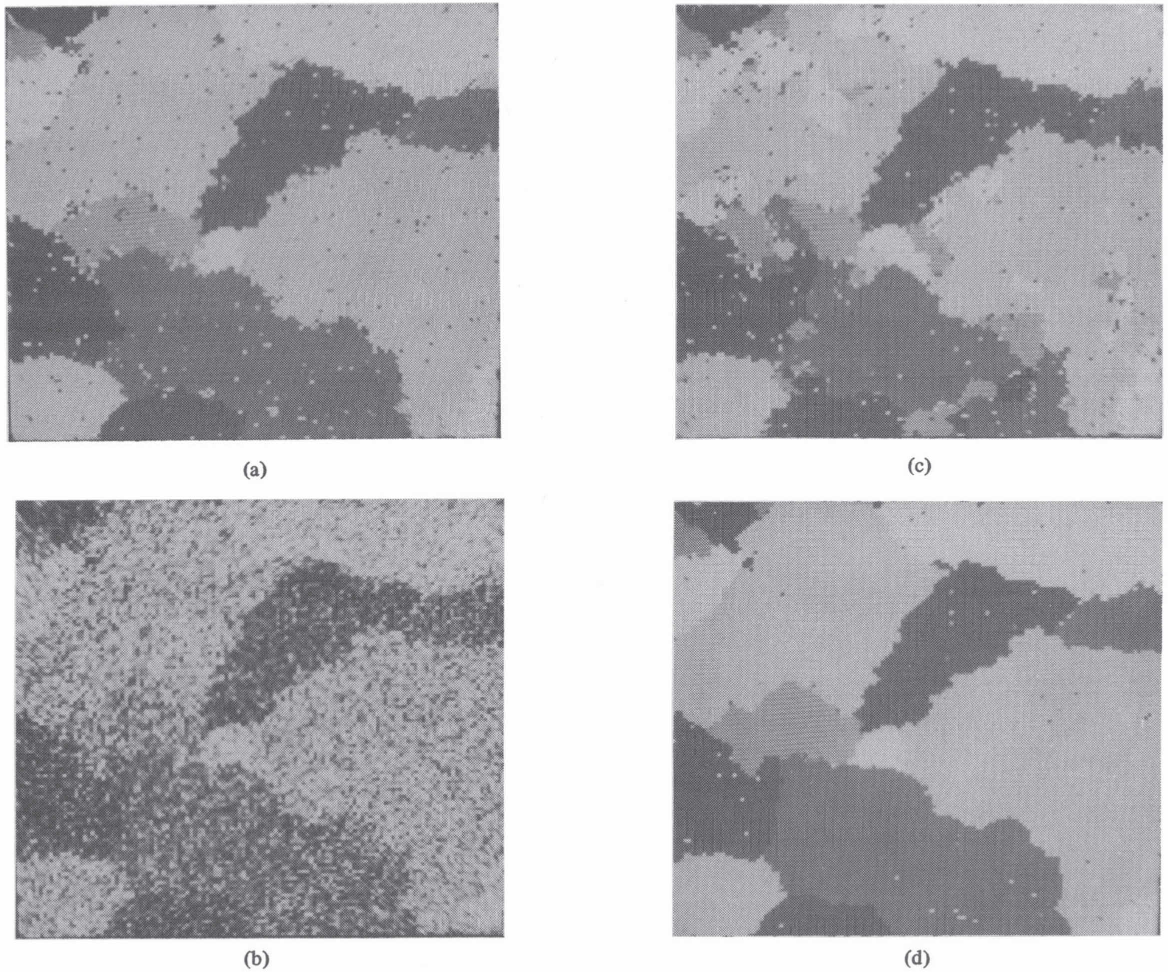


Fig. 2. (a) Original image: Sample from MRF. (b) Degraded image: Additive noise. (c) Restoration: 25 iterations. (d) Restoration: 300 iterations.

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]

PDE's framework

$\Phi(\mathbf{u})$

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

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

$$\mathbf{TV}(u) = \sup \left\{ \int_\Omega u \,\mathrm{div} w \, dx \;\middle|\; w \in \mathcal{C}_c^1(\Omega),\ \|w\|_\infty \leqslant 1 \right\} \approx \int \|\nabla u\|_2 \, dx \approx \sum_i \|G_i u\|$$

Edge-preserving functions φ [Charbonnier, Blanc-Féraud, Aubert, Barlaud 97] $\lim_{t\to\infty} \frac{\varphi'(t)}{t} = 0$

Total Generalized Variation (TGV) [Bredies,Kunish, Pock 2010]:

$$\mathbf{TGV}_\alpha^k(u) = \sup \left\{ \int_\Omega u \,\mathrm{div}^k w \, dx \;\middle|\; w \in \mathcal{C}_c^k(\Omega, \mathrm{Sym}^k(\mathbb{R}^d)),\ \|\mathrm{div}^l w\|_\infty \leqslant \alpha_l,\ l = 0, \ldots, k-1 \right\}$$

Minimizer approach

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

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

2 Analysing the optimal solutions

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

Strong results

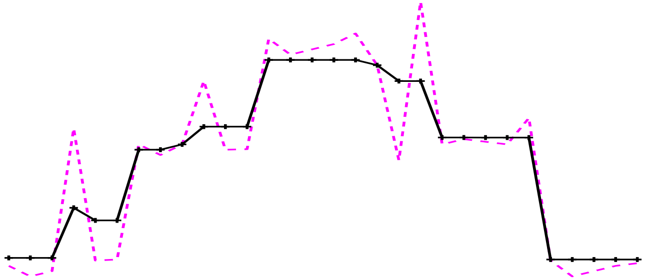
$\Longrightarrow$ **tools for "inverse" modelling**

The knowledge on the optimal solution for different families of Ψ and Φ gives us tools how to design new variational problems whose solutions exhibit predictable features

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

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

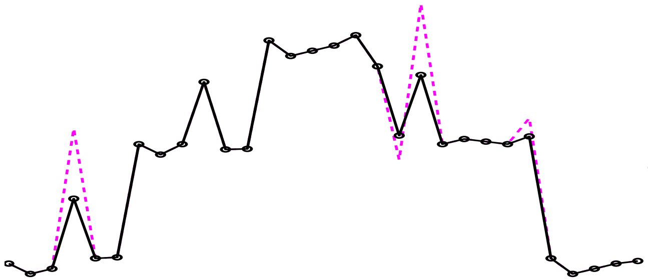
Illustration: the role of the smoothness of $\mathcal{F}_v$



STAIR-CASING

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

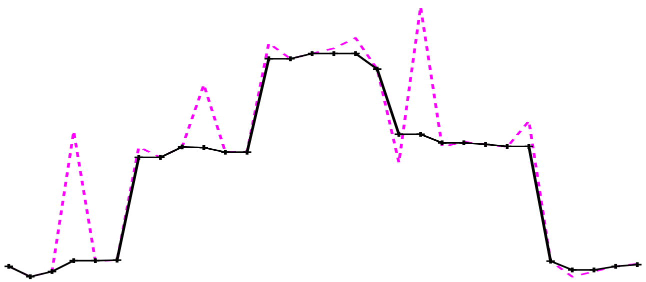
smooth **non-smooth**



EXACT DATA-FIT

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

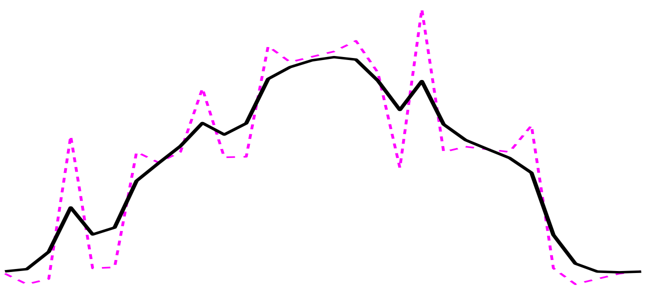
non-smooth **smooth**



BOTH EFFECTS

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

non-smooth **non-smooth**



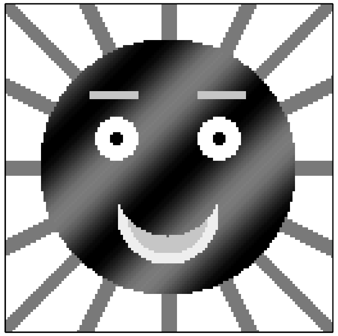
$$\mathcal{F}_v(u) = \sum_{i=1}^{p}(u_i - v_i)^2 + \beta \sum_{i=1}^{p-1} (u_i - u_{i+1})^2$$

smooth **smooth**

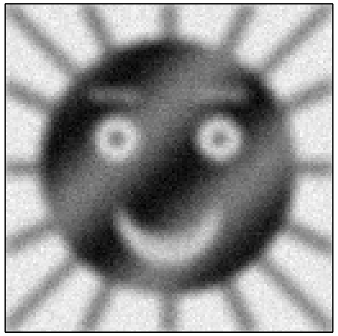
Data (— — —), Minimizer (—)

We shall explain why and how to use

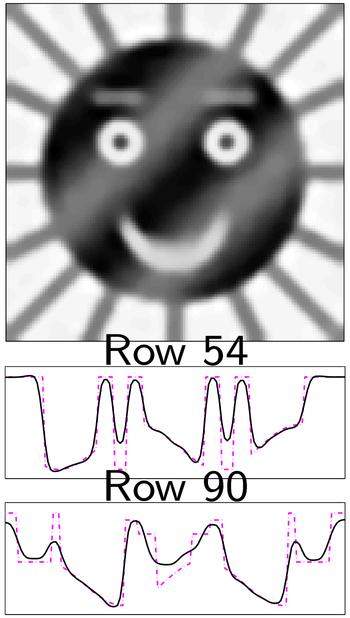
Original u_o



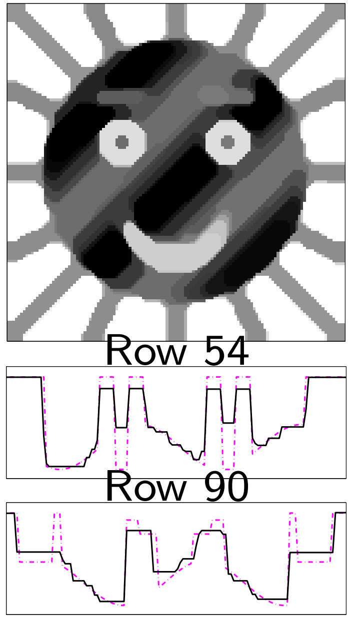
Data $v = a * u_o + n$



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



$\varphi(t) = |t|$

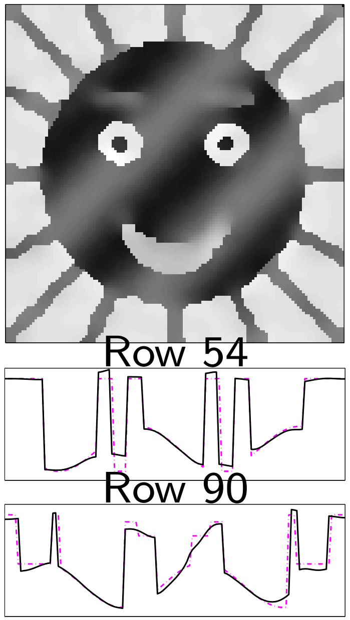


φ convex

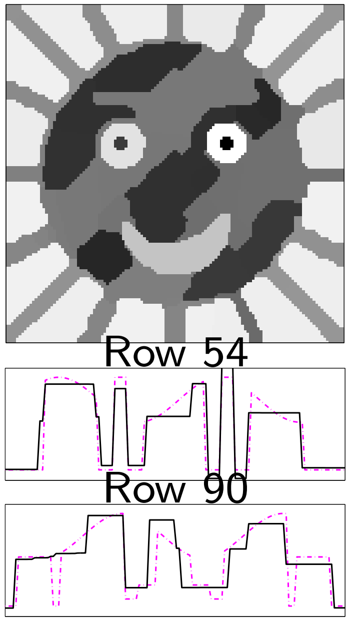
$$\mathcal{F}_v(u) = \|Au - v\|^2 + \beta \sum_i \varphi((\nabla u)[i])$$

φ smooth at 0 **φ nonsmooth at 0**

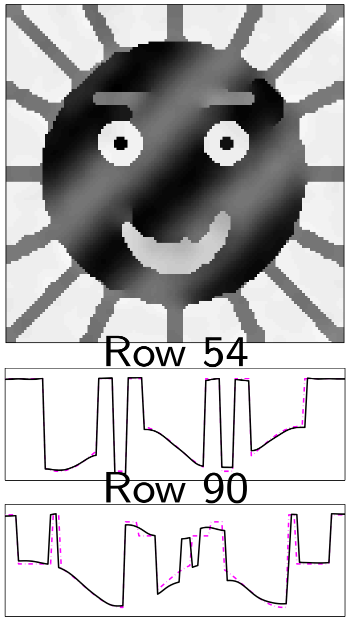
$\varphi(t) = \alpha t^2/(1+\alpha t^2)$



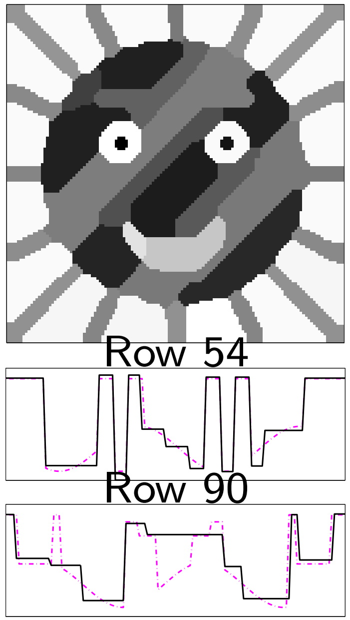
$\varphi(t) = \alpha|t|/(1+\alpha|t|)$



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

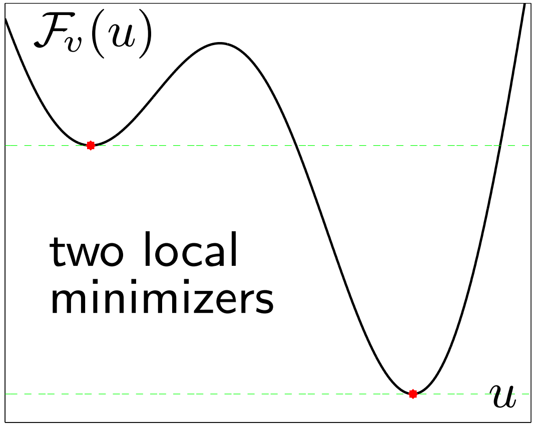


$\varphi(t) = 1 - \mathbb{1}_{(t=0)}$

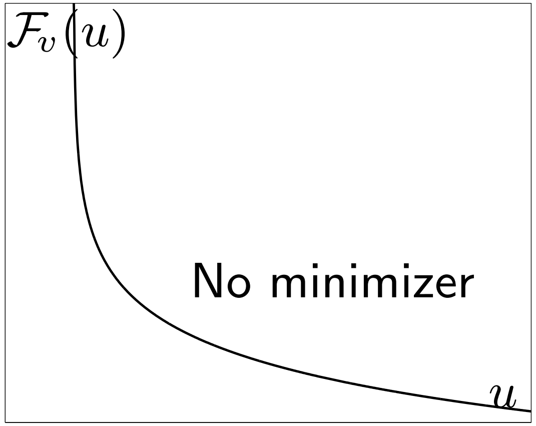


nonconvex

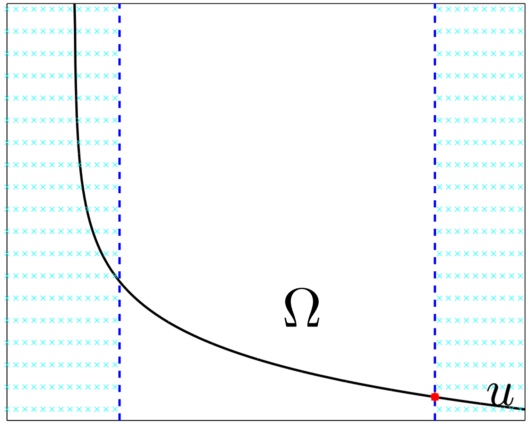
Optimization problems



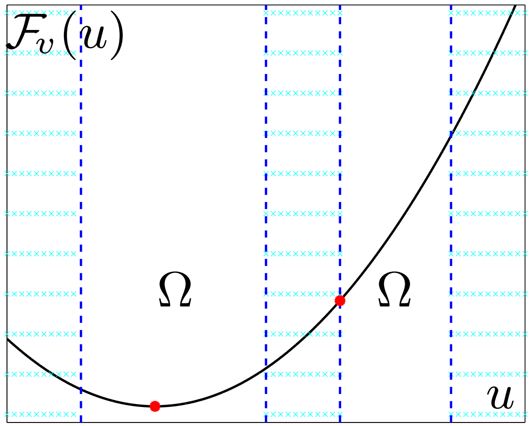
$\mathcal{F}_v$ nonconvex



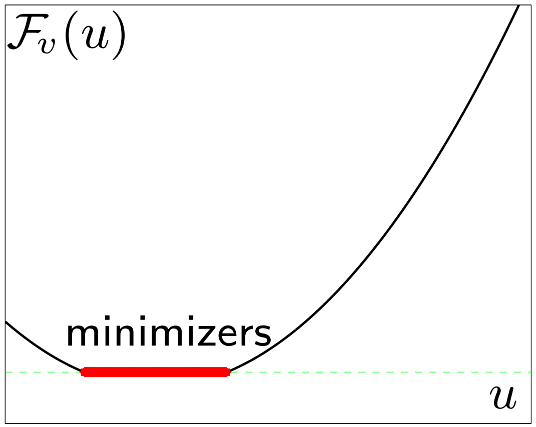
$\mathcal{F}_v$ convex non coercive $\Omega = \mathbb{R}$



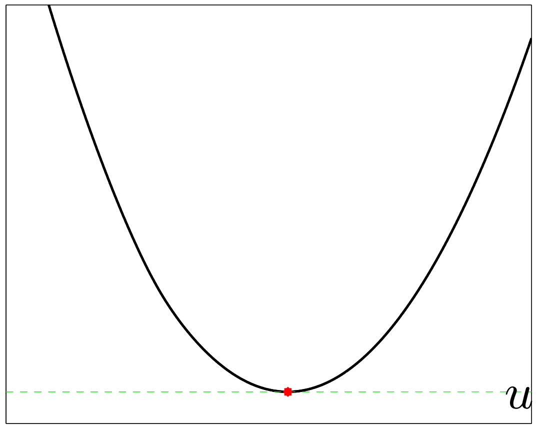
$\mathcal{F}_v$ convex non coercive Ω compact



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



$\mathcal{F}_v$ non strictly convex



$\mathcal{F}_v$ strictly convex coercive

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

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

  If $\mathcal{F}_v$ is coercive or if $\mathcal{F}_v$ continuous and Ω compact then $\widehat{U} \neq \varnothing$
  It in addition $\mathcal{F}_v$ is strictly convex, then $\widehat{U} = \{\widehat{u}\}$

  Otherwise – check:

  If there is λ finite such that $\{u \in \mathbb{R}^p \ |\mathcal{F}_v(u) \leqslant \lambda\}$ is bounded then $\widehat{U} \neq \varnothing$

  If $\mathcal{F}_v$ is asymptotically level stable then $\widehat{U} \neq \varnothing$ [10]

- Nonconvex problems

  Their optimal solutions often exhibit very desirable features
  Computing a global minimizer is seldom possible but progress [11, 12]
  Convex relaxation methods can sometimes do the job [13, 14, 15, 16]
  Nowadays – convergent algorithms for nonconvex problems [17, 18]

Definition 2 1 Let $f : \mathbb{R}^n \to \mathbb{R}$ and $S \subseteq \mathbb{R}^n$. Consider the problem $\min\{f(u) \mid u \in S\}$.

- $\widehat{u}$ is a *strict* minimizer if there is a neighborhood $\mathcal{O} \subset S$, $\widehat{u} \in \mathcal{O}$ so that $f(u) > f(\widehat{u})$ $\forall u \in \mathcal{O} \setminus \{\widehat{u}\}$.
- $\widehat{u}$ is an **isolated** (local) minimizer if $\widehat{u}$ is the only minimizer in an open subset $\mathcal{O}' \subset \mathcal{O}$ [19]

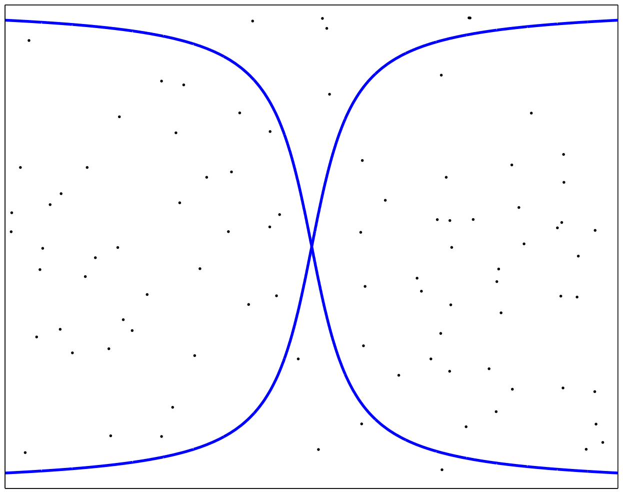
On the assessment of properties and assumptions

Definitions 2.2 and 2.3

A property (an assumption) is called generic on $\mathbb{R}^q$ if it holds on a dense open subset of $\mathbb{R}^q$. I.e. it can fail on a set N such that $N \subseteq N' \subset \mathbb{R}^q$ where N' is closed in $\mathbb{R}^q$ and its Lebesgue measure in $\mathbb{R}^q$ is $\mathbb{L}^q(N') = 0$.

A property holds almost everywhere (i.e. with probability one) in $\mathbb{R}^q$ if it fails only on a set N with $\mathbb{L}^q(N) = 0$. Its closure $\overline{N}$ in $\mathbb{R}^q$ can have $\mathbb{L}^q(\overline{N}) > 0$ in which case $\mathbb{R}^q \setminus N$ does not contain open subsets. E.g., $N = \{x \in [0,1] \mid x \text{ is rational}\}$ then $\mathbb{L}^1(N) = 0$ and $\mathbb{L}^1(\overline{N}) = 1$.

property is generic $\overset{\Rightarrow}{\not\Leftarrow}$ property holds with probability one



$$N := \{(s,t) \ : \ t = \pm \arctan(s)\}$$

N is closed in $\mathbb{R}^2$ and $\mathbb{L}^2(N) = 0$

Non-smooth functions

Rademacher's theorem: If $f_v : \mathbb{R}^n \to \mathbb{R}$ is Lipschitz continuous, then f_v is differentiable almost everywhere in $\mathbb{R}^n$. [20, 21]

A *kink* is a point u where $\nabla f v(u)$ is not defined (in the usual sense).

The (one-sided) directional derivative of f at $u \in \mathbb{R}^n$ along the direction of $d \in \mathbb{R}^n$ reads as

$$\delta f(u)(d) := \lim_{t \searrow 0} \frac{f(u+td) - f(u)}{t}$$

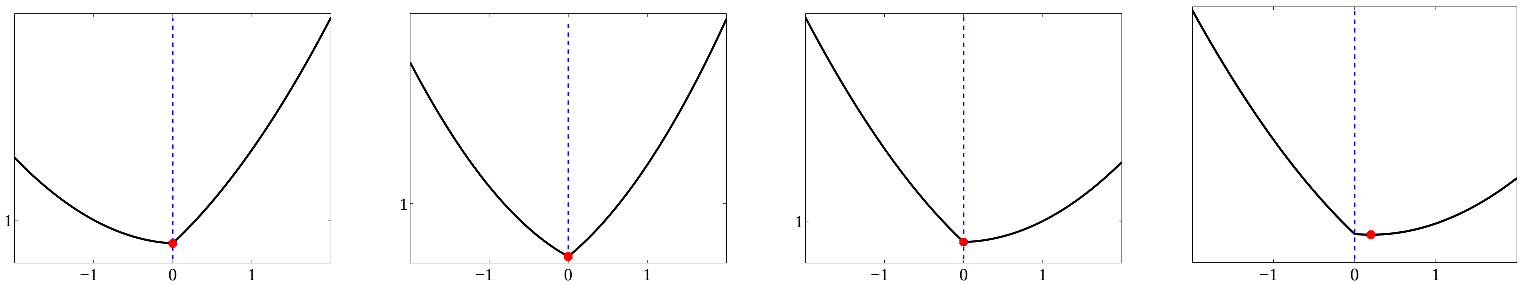
$\delta f(u)(d)$ is the right-hand side derivative. The left-hand side derivative is $-\delta f(u)(-d)$.

At a kink: $\delta f(u)(d) \neq -\delta f(u)(-d)$.

Directional derivatives are simple to use for nonconvex functions.

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

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



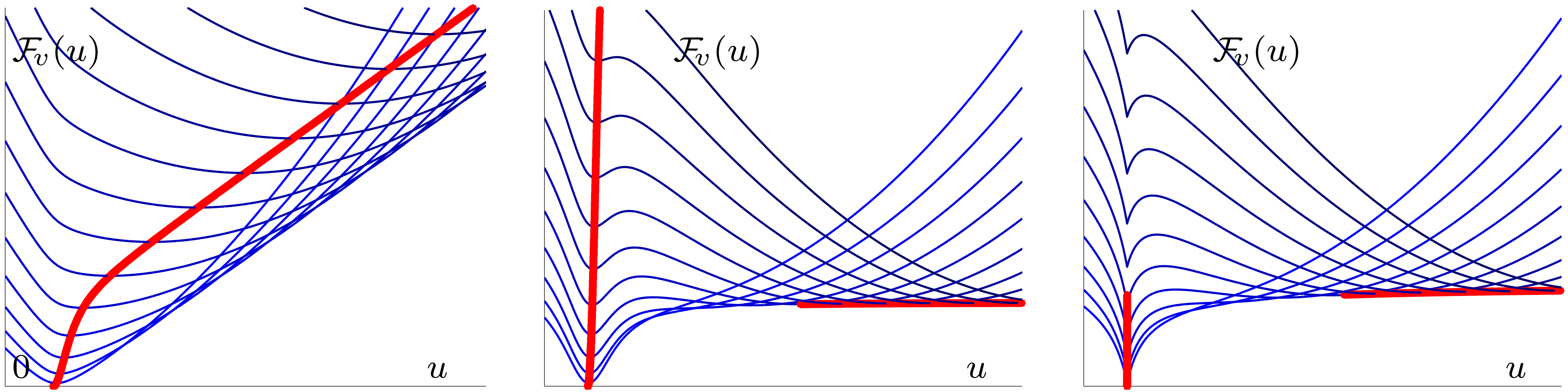
$v = -0.9$ $\quad$ $v = -0.2$ $\quad$ $v = 0.95$ $\quad$ $v = 1.2$

Question **1** `Comment the example in terms of Definitions 2.2 and 2.3.`

Definition 2.4 $\boldsymbol{\mathcal{U}} : \boldsymbol{O} \to \mathbb{R}^{\boldsymbol{p}}$, $\quad O \subset \mathbb{R}^q$ open, is a (strict) 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 – a 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 **2** What these plots reveal about the local / global minimizer functions?

An extension of the Implicit Functions Theorem

Lemma 2.1 Let $f_v : \mathbb{R}^n \to \mathbb{R}$ be $\mathcal{C}^{m \geqslant 2}$.
Let $\widehat{u}$ be such that $\nabla f_v(\widehat{u}) = 0$ and $\nabla^2 f_v(\widehat{u})$ is positive definite.
Then there exist $\rho > 0$ and a unique $\mathcal{C}^{m-1}$ strict local minimizer function
$\mathcal{U} : B(v, \rho) \to \mathbb{R}^n$ such that $\mathcal{U}(v) = \widehat{u}$. [22]

- The lemma can be extended the the whole domain $\mathbb{R}^p$ if $\mathcal{F}_v$ is strongly convex and coercive.
- The usual objective functions do not fulfill these conditions.
- We shall present different extensions of this lemma.

3 Stability of the (local) minimizers under perturbations

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

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

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

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

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

$\varphi'(0^+)>0 \quad \Longrightarrow \quad$ **Φ is nonsmooth on** $\displaystyle\bigcup_i \{u : G_iu=0\}$

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

Question **3** Why this assumption is needed?

$\mathcal{F}_v$ nonconvex $\quad \Longrightarrow \quad$ there may be (many) local minimizers

[Durand, MN 06]

H 3.1 $\varphi : \mathbb{R}_+ \to \mathbb{R}$ is continuous and $\mathcal{C}^{m \geqslant 2}$ on $\mathbb{R}_+ \setminus \{\theta_1, \cdots \theta_n\}$,
edge-preserving, possibly non-convex and $\operatorname{rank}(A) = p$

Local minimizers

Theorem 3.1 Let H3.1 hold. Then there is a closed $\boldsymbol{N} \subset \mathbb{R}^q$ with Lebesgue measure $\mathbb{L}^{\boldsymbol{q}}(\boldsymbol{N}) = \mathbf{0}$ such that $\forall \boldsymbol{v} \in \mathbb{R}^{\boldsymbol{q}} \setminus \boldsymbol{N}$, every (local) minimizer $\hat{\boldsymbol{u}}$ of $\boldsymbol{\mathcal{F}_v}$ is given by $\hat{\boldsymbol{u}} = \boldsymbol{\mathcal{U}}(\boldsymbol{v})$ where $\boldsymbol{\mathcal{U}}$ is a $\mathcal{C}^{m-1}$ (local) minimizer function. [23]

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

Global minimizers

Theorem 3.2 Let H3.1 hold. Then

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

– There is an open subset of $\mathbb{R}^q \setminus \hat{N}$, dense in $\mathbb{R}^q$, where the global minimizer function $\widehat{\mathcal{U}}$ is $\mathcal{C}^{m-1}$-continuous. [24]

Question **5** What can happen if $v \in \hat{N}$? (See the figure on p. 22.)

Question 6 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.

Question 7 For $\mathcal{F}_v$ strictly convex coercive determine N and $\hat{N}$.

Question 8 Comment the sets N and $\hat{N}$ in the sense of Definitions 2.1 and 2. 2

Questions about the assumption $\operatorname{rank}(A) = p$ (homework)

Question 9 Let $\mathcal{F}_v(u) = (u-v)^2 + \varphi(u)$ where $\varphi(u) = \begin{cases} 1-(|u|-1)^2 & \text{if} \quad 0 \leqslant |u| \leqslant 1 \\ 1 & \text{if} \quad |u| > 1 \end{cases}$

Compute the sets N and $\hat{N}$.

Hint: consider the cases $|y| > 1$, $y \in \{-1, 1\}$ and $y \in (-1, 1)$.

Question 10 Let $\mathcal{F}_v(u) = (u_1 - u_2 - v)^2 + \beta(u_1 - u_2)^2$ where $\beta > 0$.

Compute the sets N and $\hat{N}$.

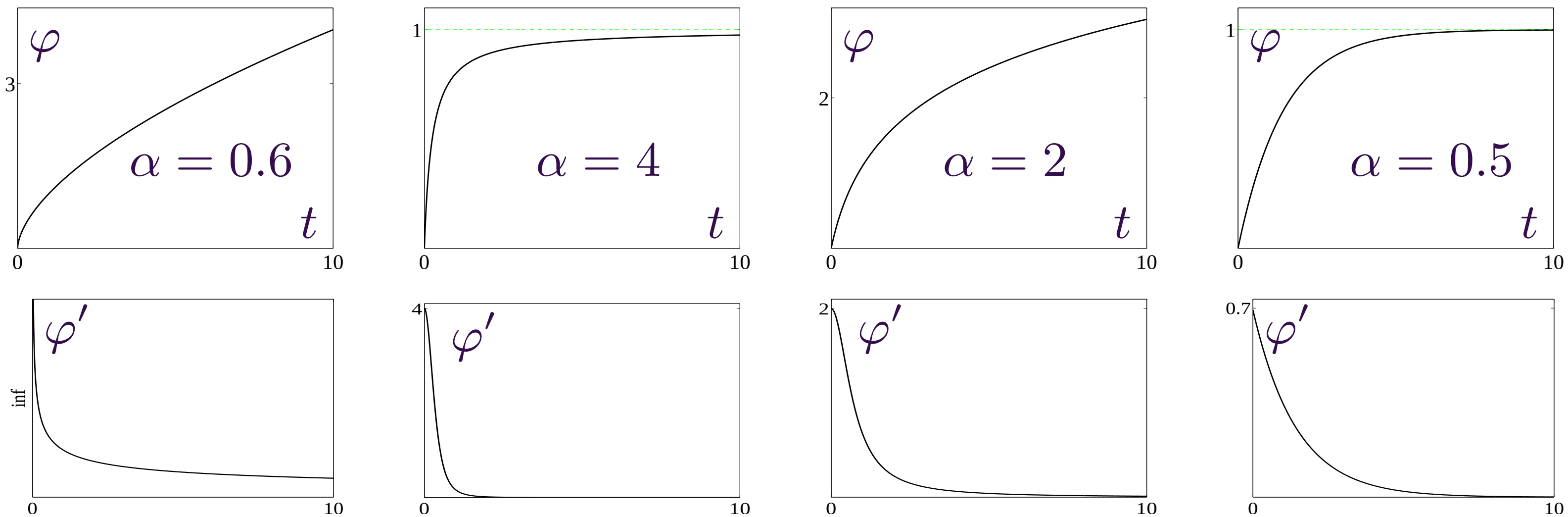
Question 11 Let $\mathcal{F}_v(u) = (u-v)^2 + \varphi(u)$ where $\varphi(u) = \min\{u^2, 1\}$.

Find the local minimizer functions and determine $\hat{N}$.

4 Minimizers under Non-Smooth Regularization

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

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



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

Main result $\mathcal{F}_v(u)=\Psi(u,v)+\beta\sum_{i=1}^{r}\varphi(\|G_iu\|)\quad \Psi\in\mathcal{C}^{m\geqslant 2},\varphi'(0^+)>0$ [MN 97,00,04]

H4.1 φ is piecewise $\mathcal{C}^m$ on $\mathbb{R}_{>0}$, increasing on $\mathbb{R}_{\geqslant 0}$, and $\varphi'(0^+)>0$, and $\Psi(\cdot,v)\sim\mathcal{C}^2$.

Theorem 4.1 Assume H4.1. [25, 26]

For $\widehat{\boldsymbol{u}}$ a local minimizer of $\mathcal{F}_v$ define $\widehat{\boldsymbol{h}}:=\{\boldsymbol{i}:\boldsymbol{G_i}\widehat{\boldsymbol{u}}=\boldsymbol{0}\}$.

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

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

This holds for any $\widehat{u}$ such that $\widehat{h}:=\{i:G_i\widehat{u}=0\}\neq\varnothing$.[a] Consequences:

$$\mathcal{O}_{\widehat{h}}\;:=\left\{v\in\mathbb{R}^q\;:\;G_i\mathcal{U}(v)=0,\;\forall\,i\in\widehat{h}\right\}\quad\Longrightarrow\quad\mathbb{L}^q(\mathcal{O}_{\widehat{h}})>0$$

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

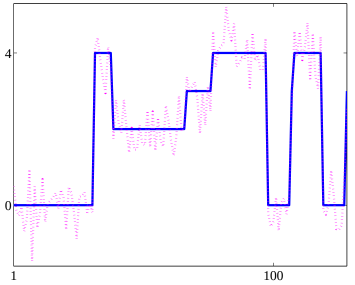
$\{G_i\}=\nabla\quad\Rightarrow\quad\widehat{u}[i]=\widehat{u}[j]$ for many neighbors (i,j) ("stair-casing" effect)

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

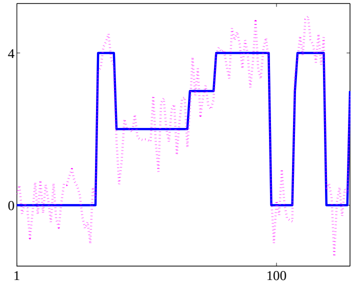
Question **12** $\{G_i\}=$ `second-order differences` $\Longrightarrow$???

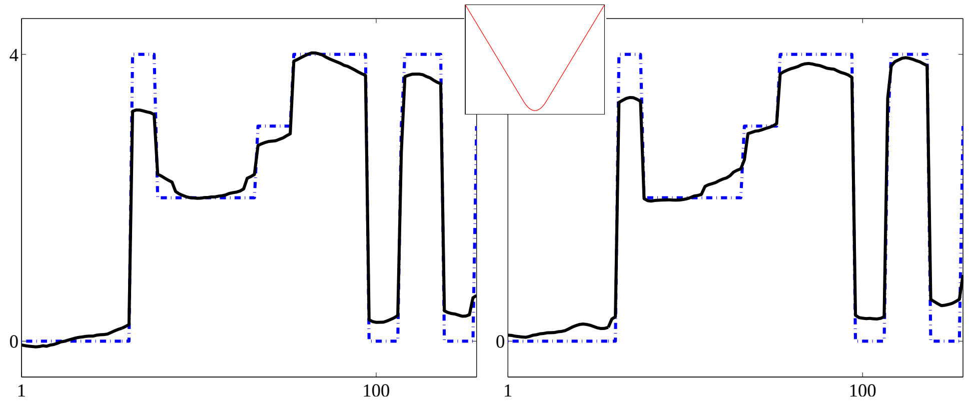
[a] The existence of v yielding $\widehat{h}\neq\varnothing$ is known.

The same original signal with two different noise realizations

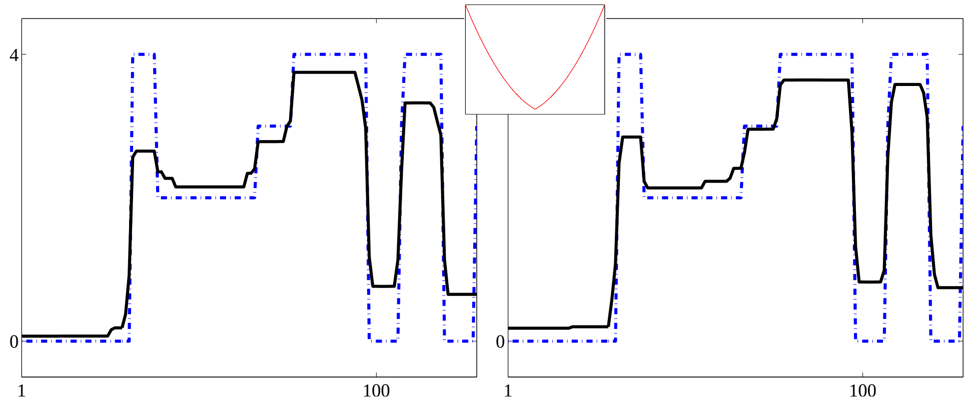


$$\mathcal{F}_v(u) = \|u - v\|^2$$
$$+\beta \sum \varphi(|u[i] - u[i-1]|)$$

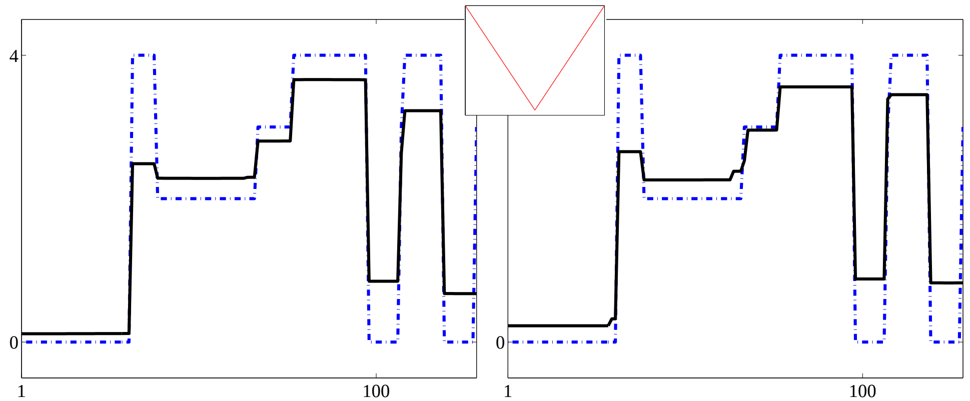




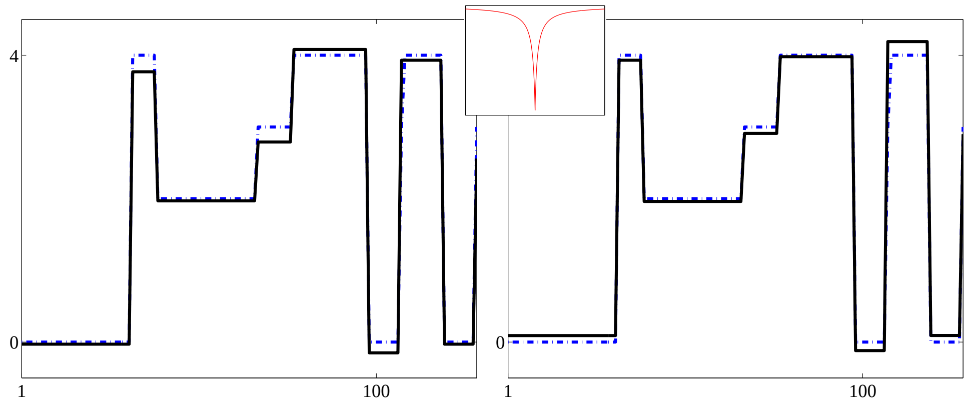
$\varphi(t) = \sqrt{\alpha + t^2}, \quad \varphi'(0) = 0$ **(smooth at 0)**



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



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



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

Analyzing the local minimizers of $\mathcal{F}_v$ under variations of v

$$(\widehat{u}, v) \in \mathbb{R}^p \times \mathbb{R}^q \qquad \widehat{h} := \{i : G_i\widehat{u} = 0\} \quad \text{and} \quad K_{\widehat{h}} := \left\{u \in \mathbb{R}^p \;\middle|\; G_i u = 0 \quad \forall\, i \in \widehat{h}\right\}$$

Can we have minimizers in $K_{\widehat{h}}$?

$$\mathcal{F}_v = f_v + g_v \quad \text{where} \quad f_v(\widehat{u}) := \Psi(\widehat{u}) + \beta \sum_{i \in \widehat{h}^c} \varphi(\|G_i\widehat{u}\|) \quad \text{and} \quad g_v(\widehat{u}) := \beta \sum_{i \in \widehat{h}} \varphi(\|G_i\widehat{u}\|) = 0$$

Necessary condition for a local minimizer function of $\mathcal{F}_v$: check only $K_{\widehat{h}} \cup K_{\widehat{h}}^{\perp}$

Theorem 4.2 Let H4.1 hold. Assume there is $\rho > 0$ so that [26]

(a) $Df_v(\widehat{u})d + \delta g_v(\widehat{u})(d) > 0 \quad \forall\, d \in K_{\widehat{h}}^{\perp} \cap \mathrm{bd}B(v, \rho)$; here $\delta g_v(\widehat{u})(d) = \beta\varphi'(0^+)\sum_{i\in\widehat{h}} \|G_i d\|$

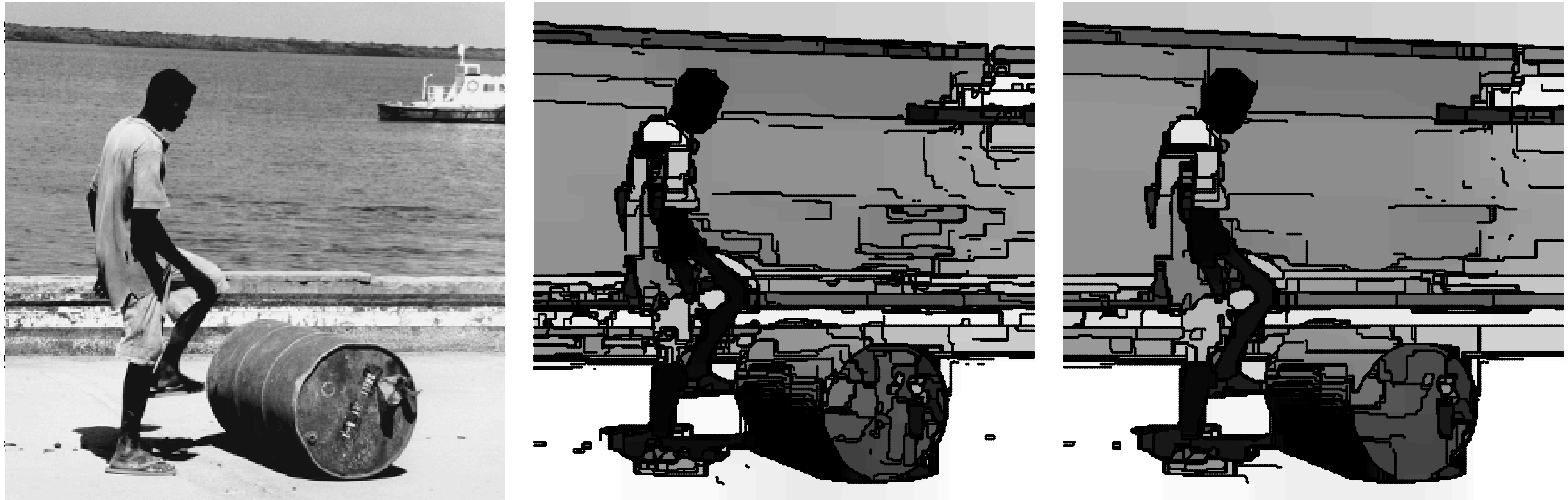
(b) $f_v\big|_{K_{\widehat{h}}}$ has a local minimizer function $\mathcal{U}_{\widehat{h}} : B(v, \rho) \to K_{\widehat{h}}$ continuous at v and $\widehat{u} = \mathcal{U}_{\widehat{h}}(v)$.

Then $\exists\, \rho' \leqslant \rho$ such that $\forall\, v' \in B(v, \rho')$, $\widehat{u}' = \mathcal{U}_{\widehat{h}}(v') \in K_{\widehat{h}}$ is a minimizer of $\mathcal{F}_v$.

Three main ingredients:

- (Fermat's rule) f_v has a local minimum at $\widehat{u} \;\Rightarrow\; \delta f_v(\widehat{u})(d) \geqslant 0,\ \forall\, d \in \mathbb{R}^n$ (directional deriv
- $\boldsymbol{\varphi'(0^+) > 0}$ then $\forall\, \gamma \in (0,1)$ there is $\rho > 0$ such that $\varphi'(t) > \gamma\varphi'(0^+)|t|,\ \forall\, t \in B(0, \rho)$.
- For (b): Lemma 2.1 (p. 23) or Theorem 3.1 (p. 25) or an extension.

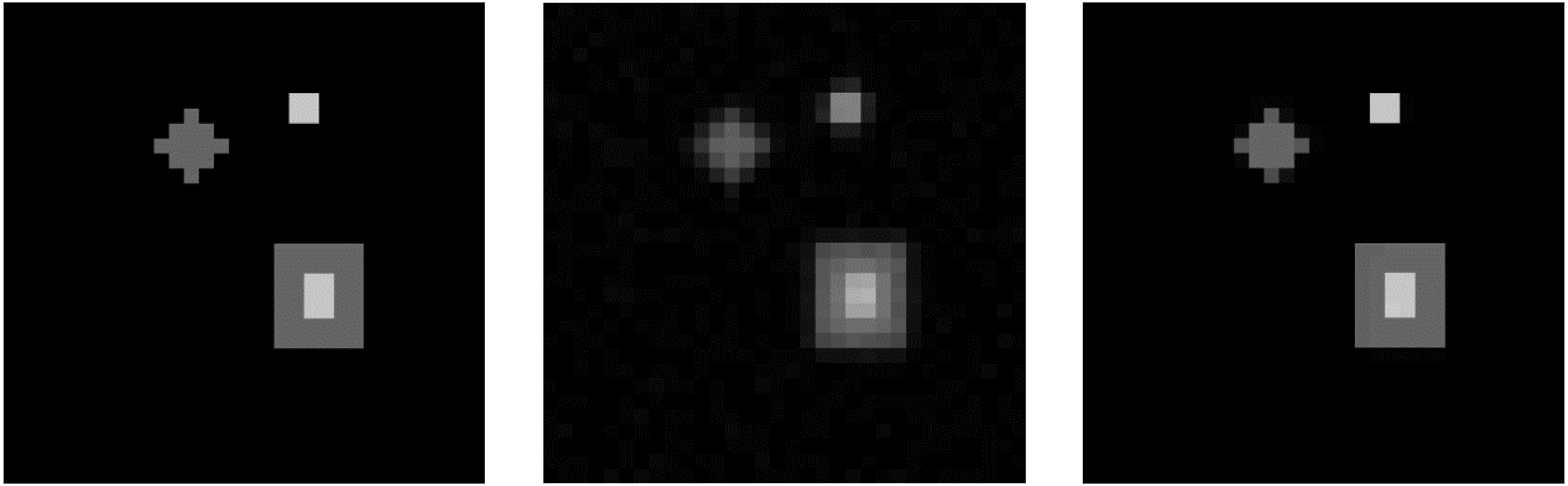
Question **13** `Describe the prior that` $\widehat{u}$ `satisfies for a general` $\{G_i\}$.



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

Black curves between constant (up to 10^{-5}) parts.

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



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 **14** Give $\mathbb{P}(u_o = 0)$.

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 maximum a posteriori (MAP) objective to recover u_o from v reads as

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

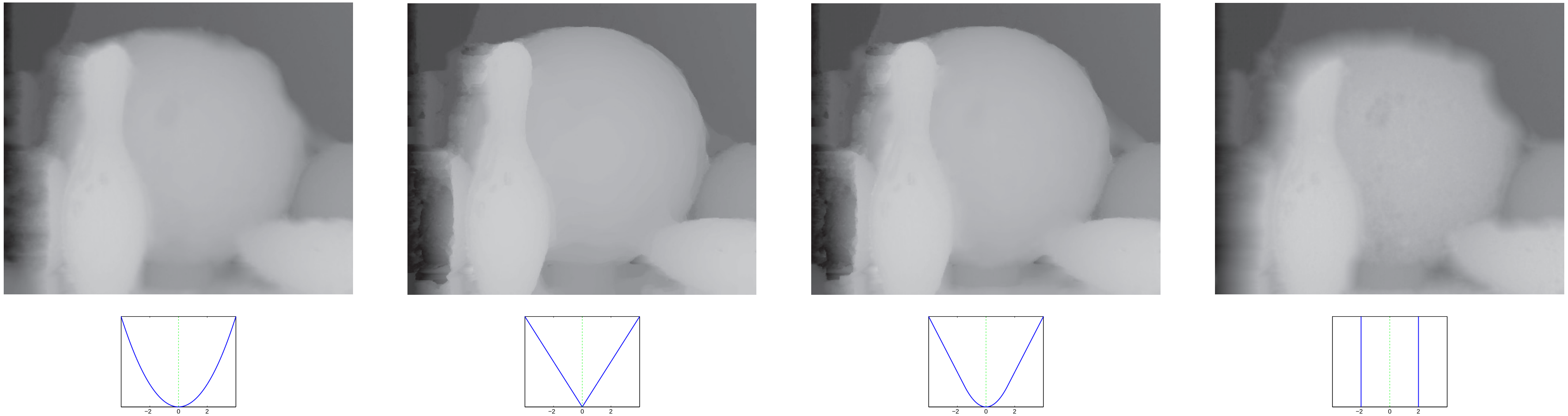
Question **15** Give the minimizer function $\mathcal{U}$ for $\mathcal{F}_v$ (defined on p. 21).

Question **16** Determine the set $\{v' \in \mathbb{R} \ : \ \mathcal{U}(v') = 0\}$. 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 Lipschitz

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

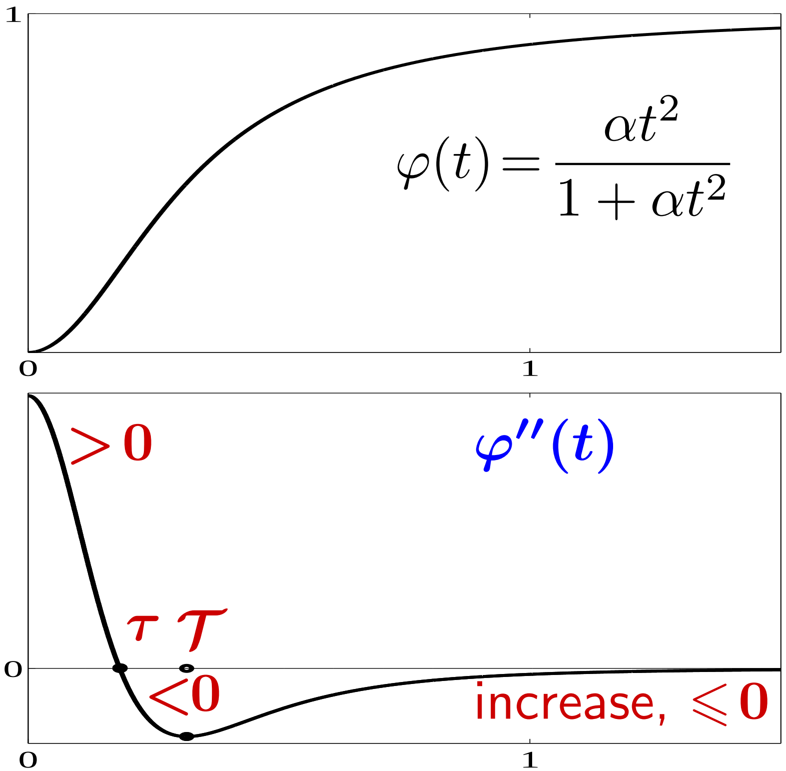
5 Nonconvex Regularization

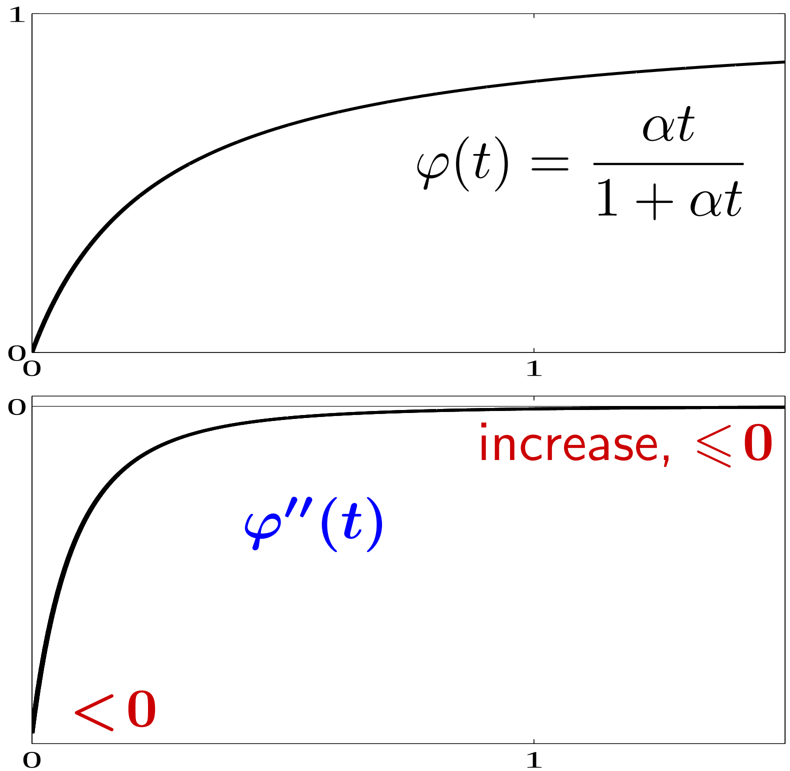
$$\boxed{\mathcal{F}_v(u) = \|Au - v\|^2 + \beta \sum_{i \in J} \varphi(\|G_i u\|)} \qquad J = \{1, \cdots, r\}$$

H5.1 (standard) φ is $\mathcal{C}^2$ on $\mathbb{R}_+$ with $\lim_{t\to\infty} \varphi''(t) = 0$ and

$\varphi'(0) = 0$ (Φ is smooth)

$\varphi'(0^+) > 0$ (Φ is nonsmooth)



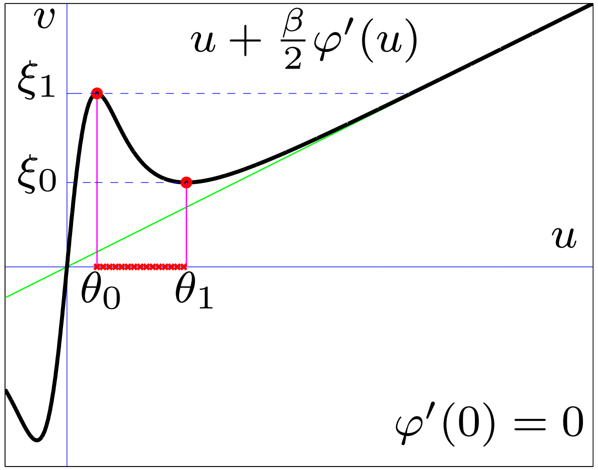


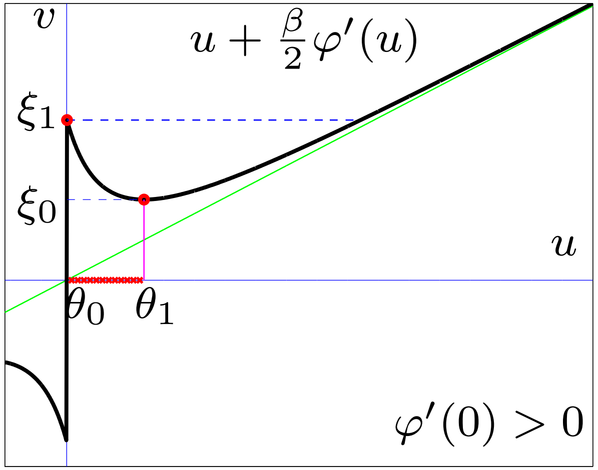
The empirical distribution of ∇u in natural images is nonconvex [Zhu, Mumford 97]

Illustration on $\mathbb{R}$

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

Fermat's rule: $\widehat{u}$ solves $\qquad v = u + \dfrac{\beta}{2}\varphi'(u)$





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

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

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

For $v = \xi$ the global minimizer jumps from $\hat{u}_0$ to $\hat{u}_1$ $\equiv$ decision on smoothing regime

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

Sharp edge property

Theorem 5.1 Assume H5.1 for φ with $\varphi'(0) = 0$ and that the set $\{G_i^\top\}$ is linearly independent. Let $\mu := \max_{i \in J} \|G^\top (GG^\top)^{-1} e_i\|_2$

$$\beta_0 := \frac{2\mu^2 \|A^\top A\|_2}{\varphi''(\mathcal{T})} \quad \text{where} \quad \mu := \max_{i \in J} \|G^\top (GG^\top)^{-1} e_i\|_2$$

With $\beta > \beta_0$ there are associated $\boldsymbol{\theta_0 \in (\tau, \mathcal{T})}$ and $\boldsymbol{\theta_1 > \mathcal{T}}$ such that every local minimizer of $\mathcal{F}_v$ satisfies

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

When β increases, θ_0 decreases and θ_1 increases. [29]

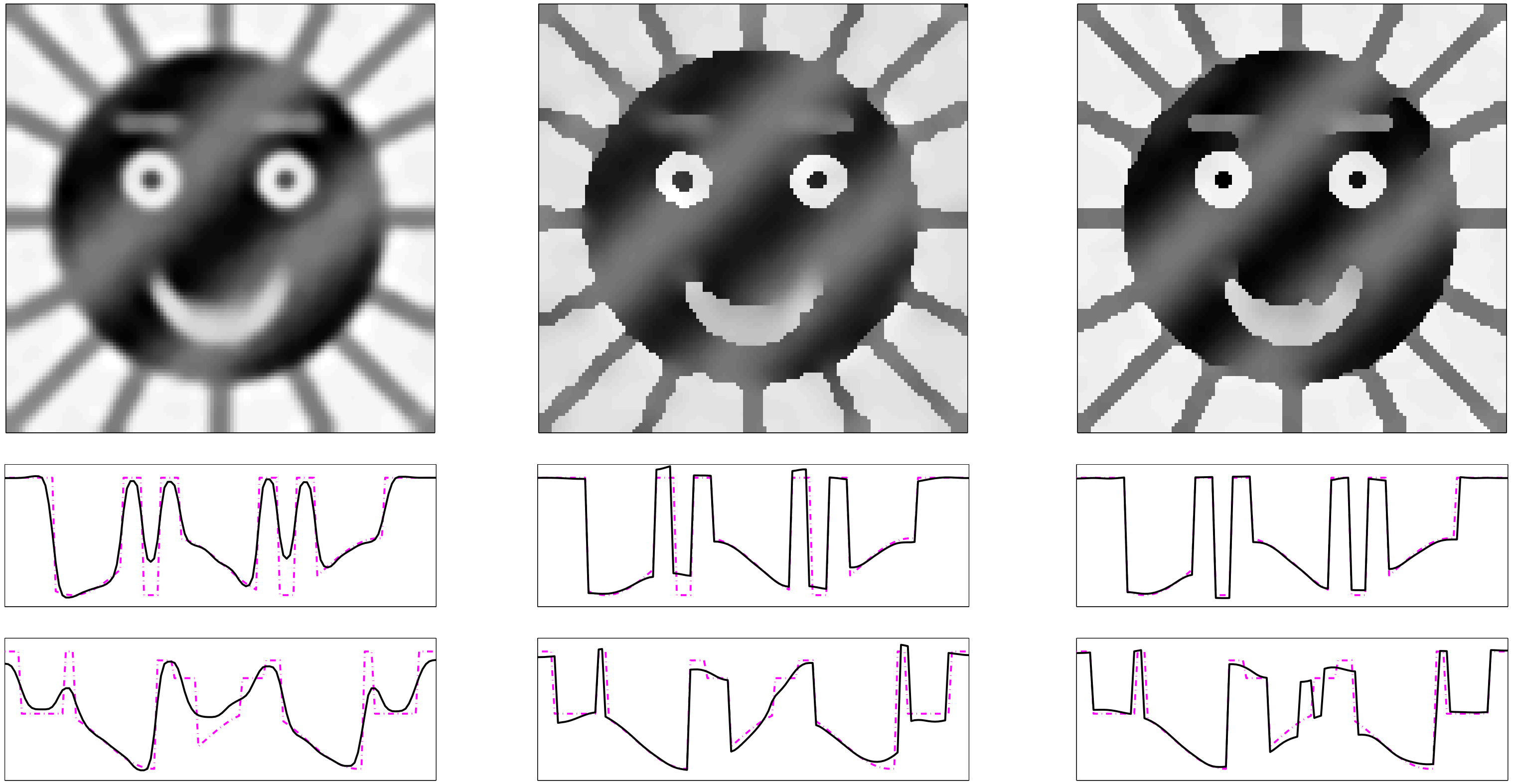
The values of (θ_0, θ_1) and β_0 are independent of v.

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

For $\varphi(t) = \min\{\alpha t^2,\ 1\}$ the theorem holds if $\widehat{u}$ is a global minimizer.

Comparison with Convex Edge-Preserving Smooth Regularization

$\varphi(t) = |t|^{1.4}$ $\qquad$ $\varphi(t) = \alpha t^2/(1+\alpha t^2)$ $\qquad$ $\varphi(t) = \min\{\alpha t^2, 1\}$



Restored images and their rows 54 and 90

Sharp edges and sparsity

Theorem 5.2 Assume H5.1 for φ with $\varphi'(0^+) > 0$. Then there exist $\boldsymbol{\theta_1 > 0}$, as well as β_0 such that for $\beta > \beta_0$ every local minimizer of $\mathcal{F}_v$ satisfies

$$\text{either} \quad \|G_i\hat{u}\| = 0 \quad \text{or} \quad \|G_i\hat{u}\| \geqslant \theta_1 \qquad \forall i \in J$$

In particular, $\beta_0|\varphi''(0^+)| \propto \|A^\top A\|_2$. [29, 30]

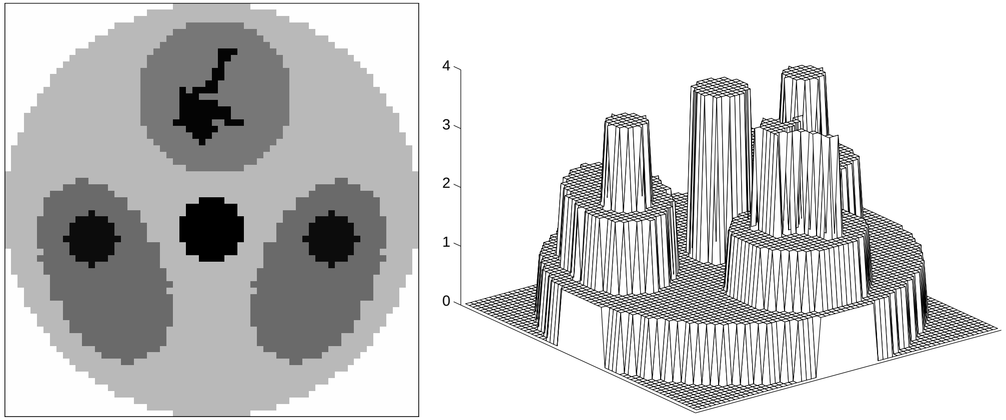
$$\{G_i\} = \nabla \quad \Longrightarrow \quad \begin{array}{rcll} \boldsymbol{\widehat{h}_0} & = & \{\boldsymbol{i : \|G_i\hat{u}\| = 0}\} & \text{constant regions} \\ \boldsymbol{\widehat{h}_1} & = & \{\boldsymbol{i : \|G_i\hat{u}\| \geqslant \theta_1}\} & \text{edges} \end{array}$$

$\Longrightarrow$ $\widehat{u}$ is a fully segmented image where we note that A is a general linear operator.

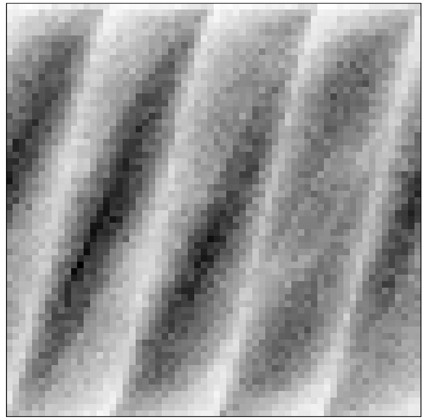
Bound θ_1 for ℓ_p non-Lipschitz, box constraints and $\{G_i\}$ first-order differences in [31].
Analysis, huberization, thrust regions and fast solver for TV^p, $0 < p < 1$ in [32].

Question **17** Explain the features of an image when $\{G_i\}$ are 2nd order differences.

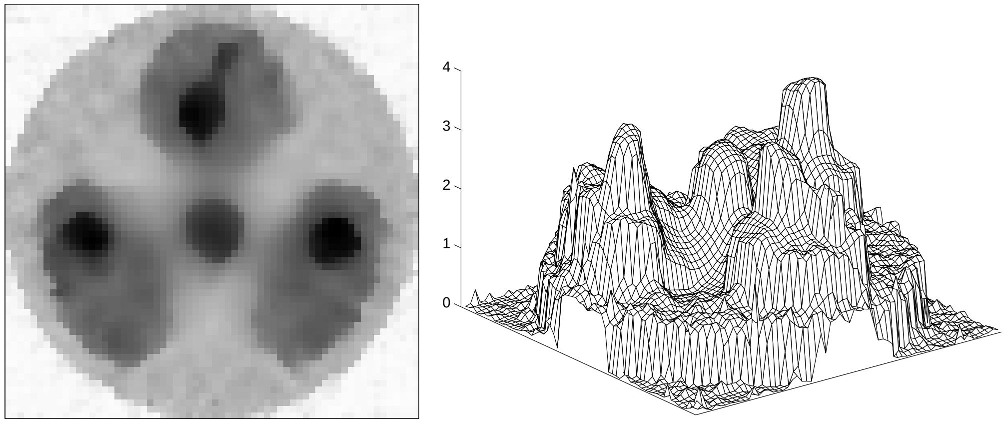
Image Reconstruction in Emission Tomography



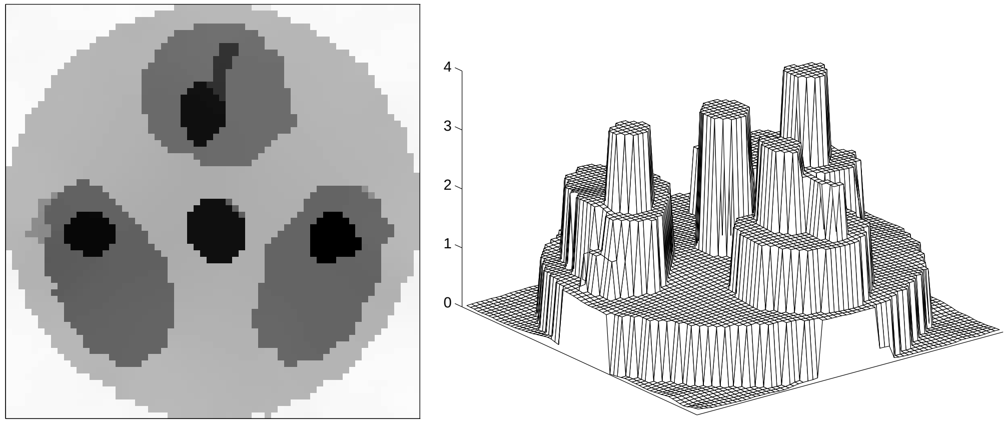
Original phantom



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

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

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

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

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

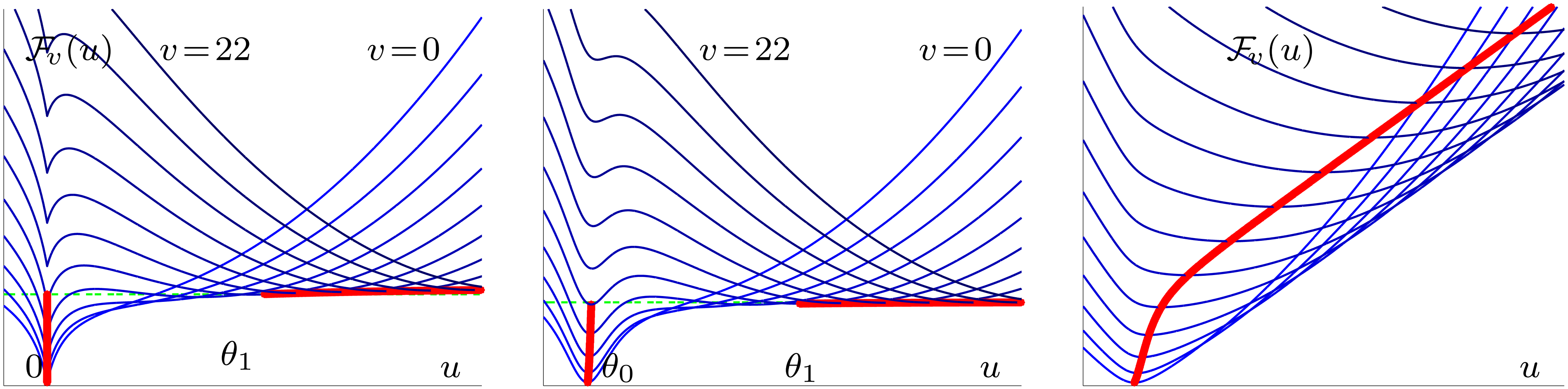
Sketch of the results

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

Moreover $\exists \; \boldsymbol{\xi_1 > 0}$ such that:

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

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



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

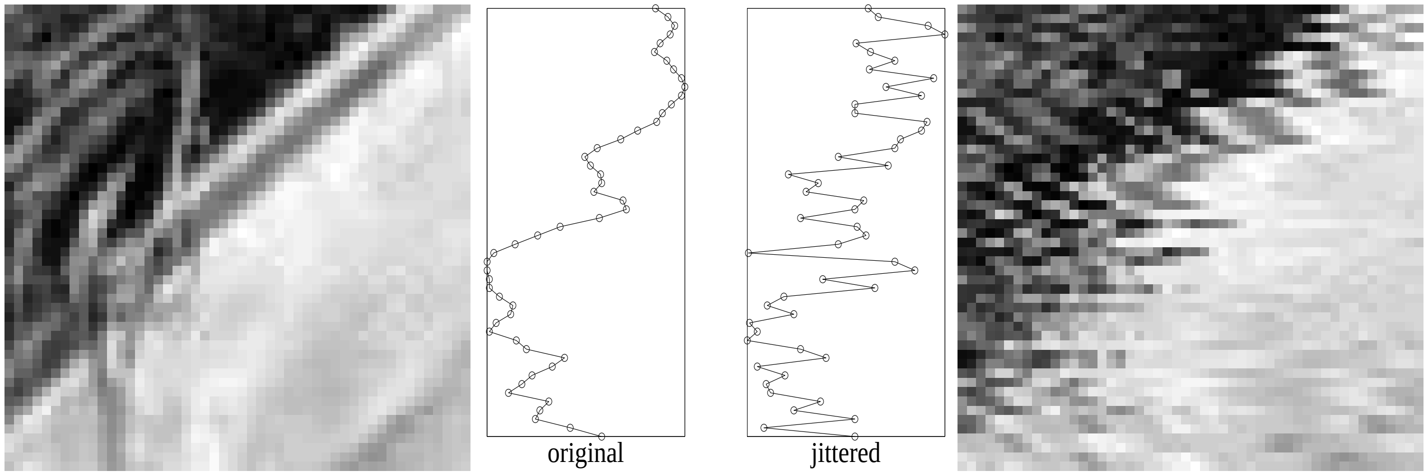
global function ( ) $\qquad$ global minimizer functions ( ) $\qquad$ unique minimizer function ( )

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

Question **18** How to describe the global minimizer when v increases?

One-step real-time dejittering of digital video [MN 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, $|\boldsymbol{d_i}| \leqslant \boldsymbol{M}$, typically $M \leqslant 20$.
- Restore $\hat{\boldsymbol{u}} \equiv$ restore $\hat{\boldsymbol{d}}_{\boldsymbol{i}}$, $\boldsymbol{1 \leqslant i \leqslant m}$



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.

The results of Theorems 4.1 and 5.2 hold for $\beta \to \infty$. Restoration model: minimize the second-order differences between the rows. [33]

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

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

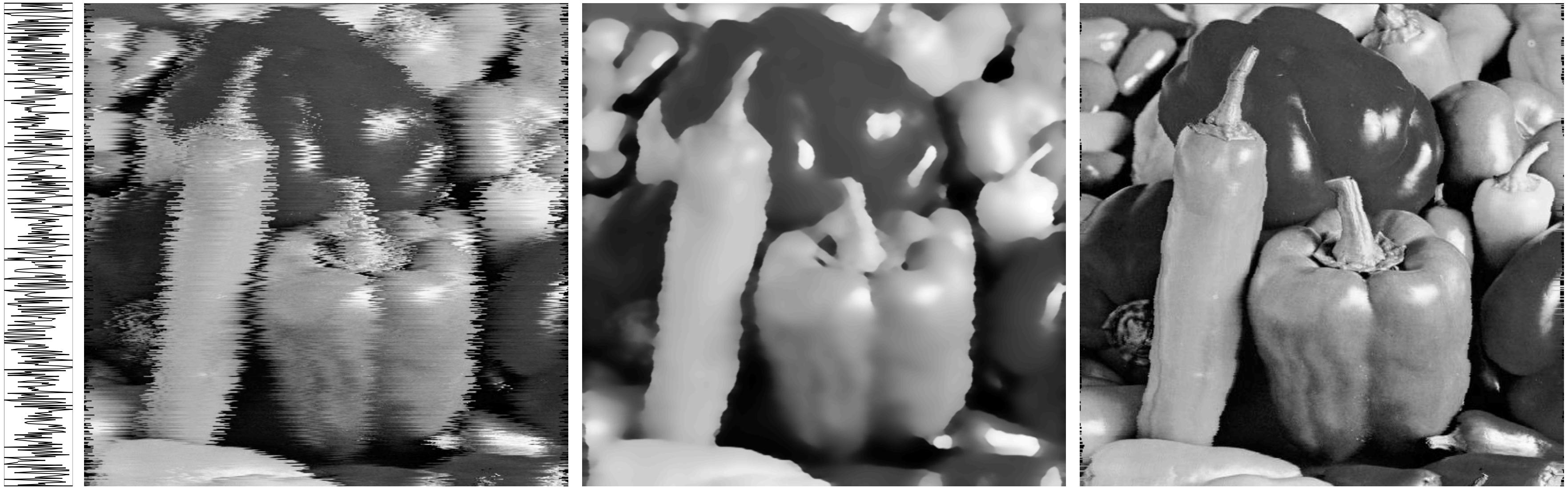
Monte-Carlo experiments – in almost all cases $\alpha = 0.5$ is better.

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

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

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

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



Jitter | Jittered Image | Bayesian TV | Bake & Shake

Original | Column model α=0.5 | Error $u_o - \widehat{u}$

[Kokaram98, Laborelli03, Shen04, Kang06, Scherzer11]

Least squares regularized with ℓ_0 norm [MN [38]]

$A = \big(a_1, \cdots, a_p\big) \in \mathbb{R}^{q\times p} \quad a_i \neq 0 \ \ \forall\, i \quad p > q$

$$\mathcal{F}_v(u) = \|Au - v\|_2^2 + \beta\|u\|_0 \qquad \text{where} \qquad \|u\|_0 := \sharp\Big\{i \in \mathbb{I}_p \ : \ u[i] \neq 0\Big\}$$

$\mathbb{I}_p = \{1, \cdots, p\}$ index set. For $\omega \subset \mathbb{I}_p$ set $\omega^c := \mathbb{I}_p \setminus \omega$ and

$$A_\omega := \big(a_{\omega[1]}, \cdots, a_{\omega[\sharp\omega]}\big) \in \mathbb{R}^{q\times \sharp\omega} \quad u_\omega := (u[\omega[1]], \cdots, u[\omega[\sharp\omega]]) \in \mathbb{R}^{\sharp\omega}$$

Theorem 4.2 Given $v \in \mathbb{R}^q$ and $\omega \subset \mathbb{I}_p$ consider the problem

$$(\mathsf{P}_\omega) \qquad \min_{u\in\mathbb{R}^p} \|Au - v\|_2^2 \quad \text{subject to} \quad u[i] = 0 \ \ \forall i \in \omega^c$$

Let $\widehat{u}$ solve (P_ω). Then for any $\beta > 0$, $\widehat{u}$ is a (local) minimizer of $\mathcal{F}_v$ and $\mathrm{supp}(\widehat{u}) \subseteq \omega$.

Lemma 4.2 Let $\mathcal{F}_v$ have a (local) minimum at $\widehat{u}$. Set $\widehat{\sigma} := \mathrm{supp}(\widehat{u})$. Then $\widehat{u}$ solves $(\mathsf{P}_{\widehat{\sigma}})$.

Solving (P_ω) for some $\omega \subset \mathbb{I}_p$ is equivalent to finding a local minimizer of $\mathcal{F}_v$.
Such a local minimizer is independent of the value of β

How to recognize a strict (local) minimizer of $\mathcal{F}_v$?

Theorem 4.3 Let $\widehat{u}$ be a (local) minimizer of $\mathcal{F}_v$. Set $\boldsymbol{\widehat{\sigma} := \mathrm{supp}(\widehat{u})}$. Then

$$\boldsymbol{\widehat{u}} \quad \textbf{is strict} \quad \Longleftrightarrow \quad \mathbf{rank}\boldsymbol{A_{\widehat{\sigma}}} = \sharp\boldsymbol{\widehat{\sigma}} \leqslant \boldsymbol{p}$$

If $\mathcal{F}_v$ has a strict (local) minimum at $\widehat{u}$, then $\widehat{u}_{\widehat{\sigma}} = \left(A_{\widehat{\sigma}}^T A_{\widehat{\sigma}}\right)^{-1} A_{\widehat{\sigma}}^T v$ and $\widehat{u}_{\mathbb{I}_p \setminus \widehat{\sigma}} = 0$.

All strict minimizers of $\mathcal{F}_v$ are moreover isolated minimizers (see p. 19)

Question **21** Is it difficult to compute a (strict) local minimizer of $\mathcal{F}_v$?

On the global minimizers of $\mathcal{F}_v$

Theorem 4.4 Let $v \in \mathbb{R}^q$ and $\beta > 0$. Then the set $\widehat{U}$ of the global minimizers of $\mathcal{F}_v$ obeys

$$\widehat{U} := \left\{ \widehat{u} \in \mathbb{R}^p \ : \ \widehat{u} = \min_{u \in \mathbb{R}^p} \mathcal{F}_v(u) \right\} \neq \varnothing$$

- every $\widehat{u} \in \widehat{U}$ is an isolated (hence strict) minimizer of $\mathcal{F}_v$ [39]
- every $\widehat{u} \in \widehat{U}$ satisfies $|\widehat{u}[i]| \geqslant \dfrac{\sqrt{\beta}}{\|a_i\|_2} \quad \forall\, i \in \mathrm{supp}(\widehat{u})$

$\Longrightarrow \quad |\widehat{u}[i]| \geqslant \theta_1 \quad \forall\, i \in \mathrm{supp}(\widehat{u})$ for $\theta_1 := \dfrac{\sqrt{\beta}}{\max_i \|a_i\|_2}$

The proof that $\widehat{U} \neq \varnothing$ consists in showing that $\mathcal{F}_v$ is asymptotically level stable. [10, 38]

A Continuous Exact ℓ_0 Penalty [40] [Soubies, Blanc-Féraud, Aubert 15]

There is no minimizers such that $|\widehat{u}[i]| \in \left(0, \frac{\sqrt{\beta}}{\|a_i\|_2}\right)$ – Continuous Exact ℓ_0 penalty

$$\mathcal{F}_v^{\mathrm{CEL0}}(u) \quad := \quad \frac{1}{2}\|Au - v\|^2 + \sum_{i \in \mathbb{I}_p} \varphi(u_i\,; \|a_i\|, \beta)$$

$$\varphi(t\,; a, \beta) \quad = \quad \beta - \frac{a^2}{2}\left(|t| - \frac{\sqrt{2\beta}}{a}\right)^2 \mathbb{1}_{|t| \leqslant \frac{\sqrt{2\beta}}{a}} \quad a \in \mathbb{R}_{>0} \quad t \in \mathbb{R}$$

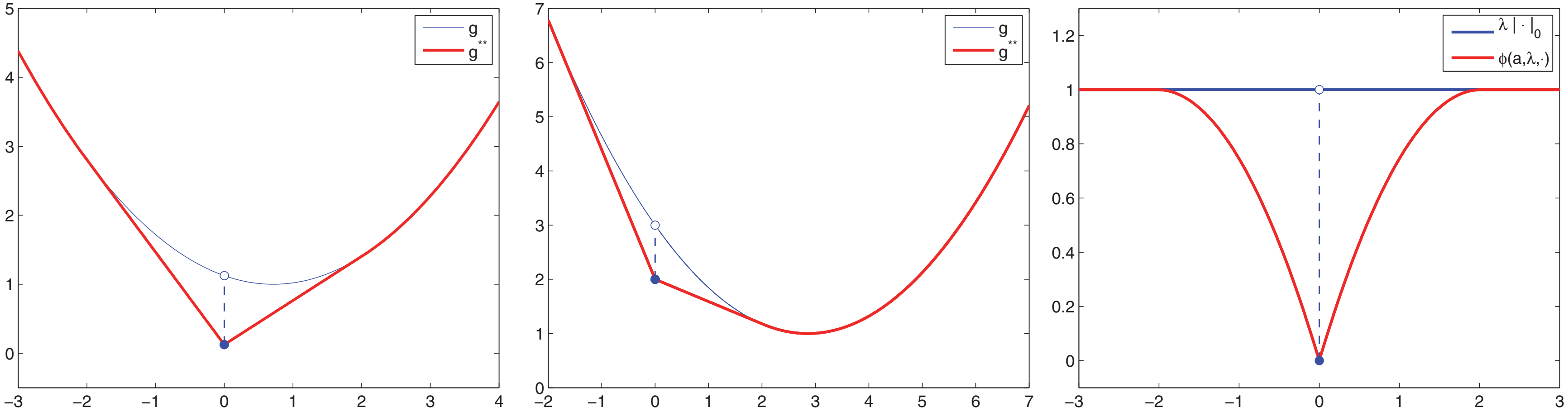


Figure 1. *Plot of g (blue) and $g^{\star\star}$ (red) for $a = 0.7$, $\lambda = 1$, and $d = 0.5$ (left) or $d = 2$ (center). Right: Plot of $\lambda|\cdot|_0$ (blue) and $\phi(a, \lambda; \cdot)$ for $a = 0.7$ and $\lambda = 1$.*

Image credits to the authors Soubies, Blanc-Féraud, Aubert [40]

- $\mathcal{F}_v^{\mathrm{CEL0}}$ and $\mathcal{F}_v^{L_0}$ (p. 48) have the same global minimizers
- Every local minimizer of $\mathcal{F}_v^{\mathrm{CEL0}}$ is a local minimizer of $\mathcal{F}_v^{L_0}$

  $\mathcal{F}_v^{\mathrm{CEL0}}$ has less local (not global) minimizers than $\mathcal{F}_v^{L_0}$
- $\mathcal{F}_v^{\mathrm{CEL0}}$ is continuous nonsmooth and nonconvex
- $u[i] \mapsto \mathcal{F}_v^{\mathrm{CEL0}}(u)$ is convex $\forall\ i$

6 Minimizers relevant to non-smooth data-fidelity

General case

[MN 02]

$$\boxed{\mathcal{F}_v(u) = \sum_i \psi(|a_i u - v[i]|) + \beta\Phi(u), \quad a_i \in \mathbb{R}^{1,p}, \quad \psi'(0^+) > 0}$$

H6.1 $\Phi \in \mathcal{C}^{m\geqslant 2}$ and $\psi \in \mathcal{C}^m(\mathbb{R}_{>0})$ with $\psi'(0^+) > 0$ finite.

Teorem 6.1 Assume H6.1. Let $\hat{u}$ be a local minimizer of $\mathcal{F}_v$. Set $\widehat{h} := \{i : a_i\hat{u} = v[i]\}$. Assume that the set $\{a_i, i \in \widehat{h}\}$ is linearly independent. Then $\exists\ O_{\widehat{h}} \subset \mathbb{R}^q$ open, $\exists\ \mathcal{U} \in \mathcal{C}^{m-1}$ local minimizer function so that

$$v' \in O_{\widehat{h}}, \quad \hat{u}' = \mathcal{U}(v') \quad \Rightarrow \quad a_i\,\hat{u}' = v'[i] \ \ \forall\, i \in \widehat{h} \ \ \text{and} \ \ a_i\,\hat{u}' \neq v'[i] \ \ \forall\, i \in \widehat{h}^c$$

The result holds for any $\widehat{h} \subset \{1, \cdots, q\}$ such that[a] $\widehat{h} \neq \varnothing$ It follows that

$$\mathcal{O}_{\widehat{h}} := \left\{ v \in \mathbb{R}^q \ : \ a_i\mathcal{U}(v) = v[i], \ \ \forall\, i \in \widehat{h} \quad a_i\mathcal{U}(v) \neq v[i], \ \ \forall\, i \in \widehat{h}^c \right\} \quad \Longrightarrow \quad \mathbb{L}^q(\mathcal{O}_{\widehat{h}}) > 0$$

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

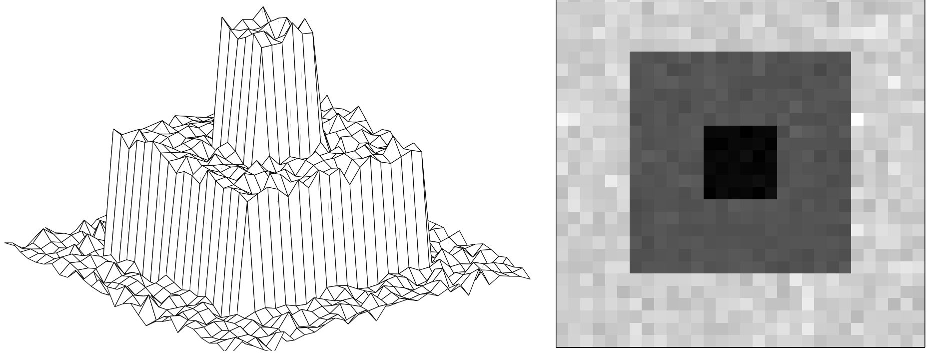
[a] For $\beta = 0$ one has $\sharp\,\widehat{h} = \mathrm{rank}\{a_i, i \in \mathbb{I}_q\}$.

Question 22 Suggest cases when you would like that your minimizer obeys this property.

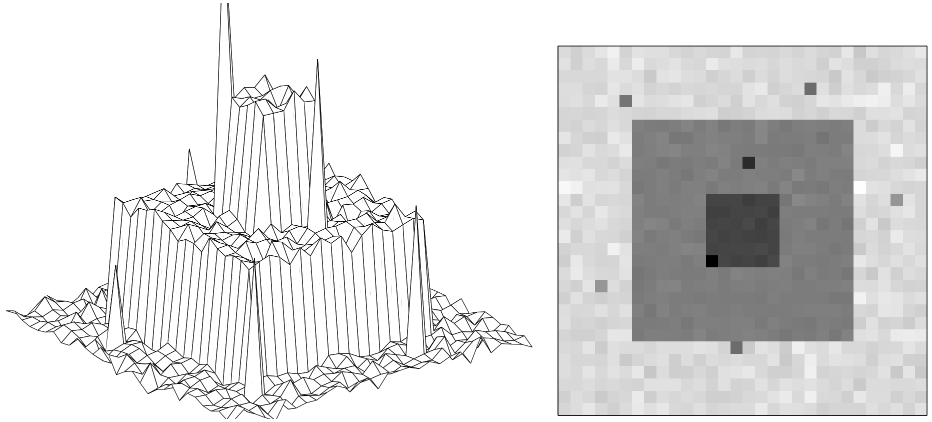
Question 23 Compute the minimizer of $\mathcal{F}_v(u) = |u - v| + \beta u^2$ for $u, v \in \mathbb{R}$ and $\beta > 0$.

Question 24 Find a relationship between the properties of the minimizer

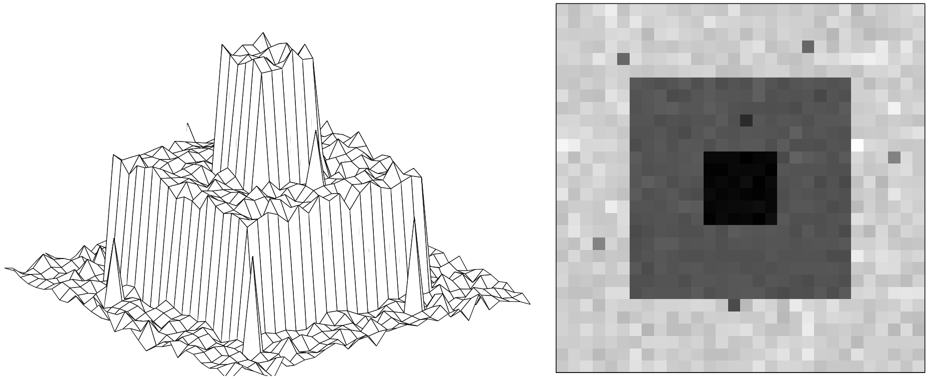
when $\varphi'(0^+) > 0$ (chapter 4, p. 27) and when $\psi'(0^+) > 0$ (this chapter, p. 52)



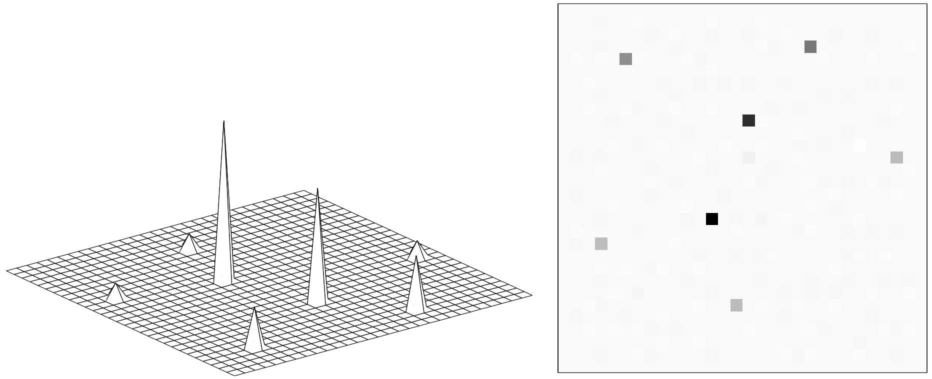
Original u_o



Data $v = u_o$+outliers

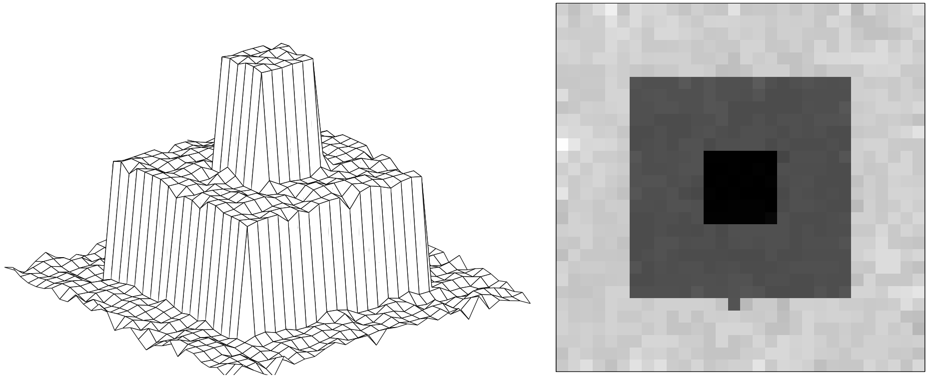


Restoration $\hat{u}$ for $\boldsymbol{\beta = 0.14}$

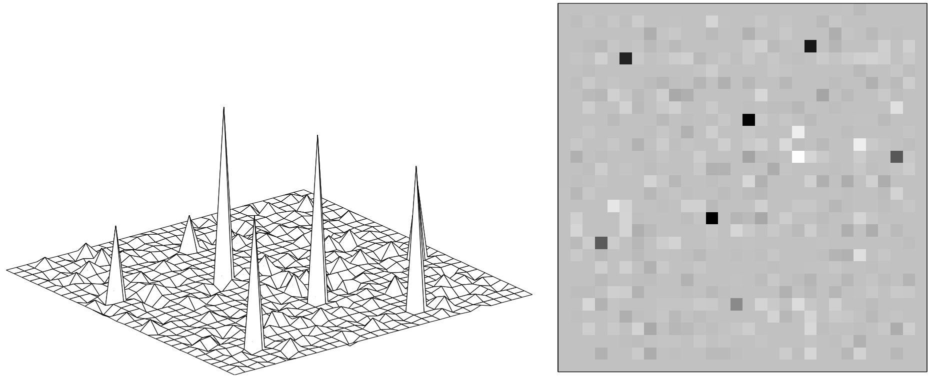


Residuals $v - \hat{u}$

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

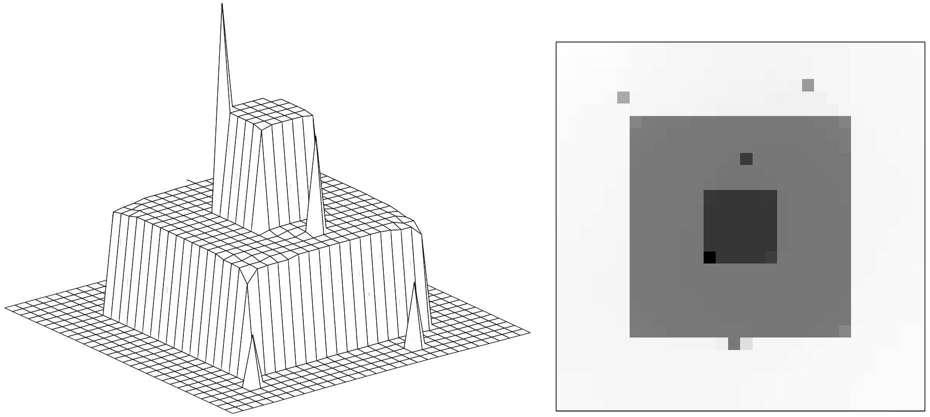


Restoration $\hat{u}$ for $\boldsymbol{\beta = 0.25}$

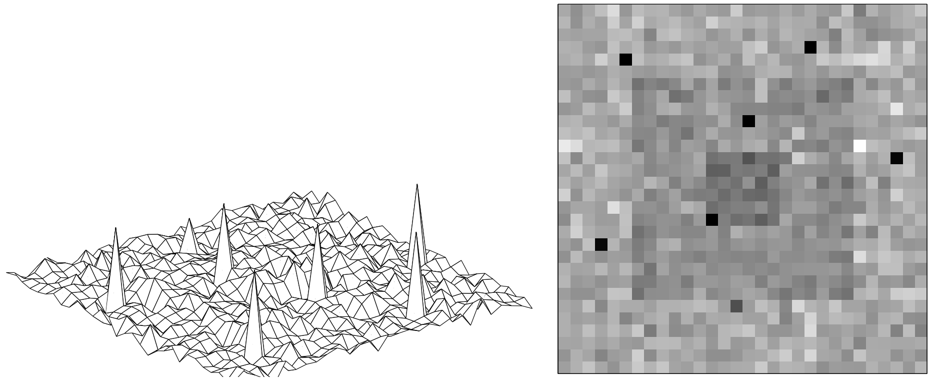


Residuals $v - \hat{u}$

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



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



Residuals $v - \hat{u}$

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

Analyzing the local minimizers of $\mathcal{F}_v$ under variations of v

$$(\widehat{u}, v) \in \mathbb{R}^p \times \mathbb{R}^q \quad \widehat{h} := \{i \ : \ a_i\widehat{u} = v[i]\} \quad \mathcal{K}_{\widehat{h}}(v) := \{u \in \mathbb{R}^p \ : \ a_i\widehat{u} = v[i]\}$$

$$K_{\widehat{h}} := \{u \in \mathbb{R}^p \ : \ a_i\widehat{u} = 0\}$$

$$\mathcal{F}_v = f_v + g_v \ \text{ for } \ f_v(\widehat{u}) = \sum_{i\in\widehat{h}} \psi\left(|a_i\widehat{u} - v[i]|\right) \ \text{ and } \ g_v(\widehat{u}) = \sum_{i\in\widehat{h}^c} \psi\left(|a_i\widehat{u} - v[i]|\right) + \beta\Phi(\widehat{u})$$

Necessary condition for a local minimizer function of $\mathcal{F}_v$ near $\widehat{u}$: check only $\left(K_{\widehat{h}} \cup K_{\widehat{h}}^{\perp}\right)$

Theorem 6.2 Let H 6.1 hold. Given $v \in \mathbb{R}^q$ and $\widehat{u} \in \mathbb{R}^p$, let $\widehat{h} := \{i \in \mathbb{I}_q \ : \ a_i\widehat{u} = v[i]\}$. Suppose that $\{a_i, i \in \widehat{h}\}$ are linearly independent and that

(a) $\quad Dg_v(\widehat{u})d = 0$ and $d^{\mathsf{T}}\left(D^2 g_v(\widehat{u})\right)d > 0 \quad \forall\, d \in K_{\widehat{h}}$

(b) $\quad \delta f_v(\widehat{u})(d) + Dg_v(\widehat{u})d > 0 \quad \forall\, d \in K_{\widehat{h}}^{\perp} \quad \|d\| = 1$

Then $\exists \rho > 0$ and a $\mathcal{C}^{m-1}$ local minimizer function $\mathcal{U} : B(v, \rho) \to \mathbb{R}^p$ obeying $\widehat{u} = \mathcal{U}(v)$ and

$$v' \in B(v, \rho) \quad \Longrightarrow \quad a_i\mathcal{U}(v') = v'[i] \ \ \forall\, i \in \widehat{h} \quad \text{and} \quad a_i\mathcal{U}(v') \neq v'[i] \ \ \forall\, i \in \widehat{h}^c$$

Details

- $g_v(\widehat{u}) = \mathcal{F}_v|_{\mathcal{K}_{\widehat{h}}}(\widehat{u}) = \sum_{i\in\widehat{h}^c} \psi\left(|a_i\widehat{u} - v[i]|\right) + \beta\Phi(\widehat{u})$ is $\mathcal{C}^m$ near $\widehat{u}$
- $f_v(\widehat{u}) = 0$ and $\delta f_v(\widehat{u})(d) = \psi'(0^+)\sum_{i\in\widehat{h}} |a_i d| > 0 \ \ \forall\, d \in K_{\widehat{h}}^{\perp} \setminus \{0\}$
- assumption $\{a_i, i \in \widehat{h}\}$ are linearly independent can fail only if v is in a proper subspace

Other facts

– The existence of a $\mathcal{C}^{m-1}$ local minimizer function shows the stability of the local minimizers of $\mathcal{F}_v$ and extends Lemma 2.1 (p. 23)

– $v' \mapsto \widehat{h}(v')$ is constant on $B(v,\rho)$ hence stable under perturbations.

Set $A := \left(\begin{array}{c} a_1 \\ \dots \\ a_q \end{array}\right)$ and let $\boldsymbol{\psi(t) = t}$. Let $v' \in B(v,\rho)$.

(a) $\Longrightarrow$ $$Dg_v(u)d = \Big(A_{\widehat{h}^c}\{\mathrm{sign}\,(a_i u - v'[i])\}_{i\in\widehat{h}^c} + \beta D\Phi(u)\Big)\, d = 0 \quad \forall\, d \in K_{\widehat{h}}$$

(b) $\Longrightarrow$ $$\sum_{i\in\widehat{h}} |a_i d| + \beta\left(A_{\widehat{h}^c}\{\mathrm{sign}\,(a_i u - v'[i])\}_{i\in\widehat{h}^c} + \beta D\Phi(u)\right) d > 0 \quad \forall\, d \in K_{\widehat{h}}^{\perp}$$

Only $v'[i]$ for $i \in \widehat{h}$ need to be in $B(v,\rho)$ in order to keep $\widehat{h}$ constant; and

$$\forall\, v'[i] \quad i \in \widehat{h}^c \quad \text{such that} \quad \mathrm{sign}\,(a_i u - v'[i]) = \mathrm{sign}\,(a_i u - v[i])$$

cannot change the minimizer. Therefore,

$$\boxed{v'[i] \quad \forall\, i \in \widehat{h}^c \text{ can be outliers}}$$

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

Example

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

The minimizer function is

$$\begin{array}{lcl} |v[i]| \leqslant \dfrac{1}{\beta} & \Longrightarrow & \mathcal{U}[i] = v[i] \\ |v[i]| > \dfrac{1}{\beta} & \Longrightarrow & \mathcal{U}[i] = \dfrac{1}{\beta}\mathrm{sign}(v[i]) \end{array}$$

For any $h \subset \mathbb{I}_p$ the set $\mathcal{O}_h$ in Theorem 6.1 is well defined and reads as

$$\mathcal{O}_h := \left\{ v \in \mathbb{R}^p \ : \ |v[i]| \leqslant \frac{1}{\beta} \ \forall \ i \in h \quad \text{and} \quad |v[i]| > \frac{1}{\beta} \ \forall \ i \in h^c \right\}$$

6 Limits on noise removal using likelihood and regularization

Numerous works on image restoration use data-fidelity $= -\log(\text{Likelihood})$ and regularization.

Context

n noise with known distribution $f_N(n)$

$$v = Au + n$$

$$f_{V|U}(v|u) = f_N(v - Au) \quad \Longrightarrow \quad \Psi(u;v) = -\log\big(\text{Likelihood}(v|u)\big) = -\log f_N(v - Au)$$

How the noise is processed at a minimizer of $\mathcal{F}_v = \Psi + \beta\Phi$?

- We know what we want.
- We want to understand what we do

We can say that the noise is properly cleaned if the residual

$\widehat{n} = v - A\widehat{u}$ has an empirical distribution similar to f_N.

How Φ and β can help ? The maximum a posteriori (MAP) estimator will be evoked, details p. 108

Normal noise and edge-preserving regularization

$$v = Au_o + n \qquad n \sim (0, \sigma^2 I)$$

$$\mathcal{F}_v(u) = \frac{1}{2}\|Au - v\|_2^2 + \beta \sum_i \varphi(\|G_i u\|)$$

For convex edge-preserving potential functions usually[a] $\|\varphi'\|_\infty < \infty$. We can set $\|\varphi'\|_\infty = 1$.

H7.1 φ is piecewise $\mathcal{C}^1$, increasing on $\mathbb{R}_{\geqslant 0}$ and $\|\varphi'\|_\infty$ is finite.

Theorem 7.1. Assume H7.1 with $\|\varphi'\|_\infty = 1$ and $\mathrm{rank} A = q \leqslant p$. [42]
Let $\widehat{u}$ be a (local) minimizer $\widehat{u}$ of $\mathcal{F}_v$. Then

$$\|\widehat{n}\|_\infty = \|A\widehat{u} - v\|_\infty \leqslant \beta \left\|(A^T A)^{-1} A\right\|_\infty \|G\|_1$$

If $G \approx \{\nabla_i\}$ then $\|G\|_1 = 4$ for u an image. Let also $A = I$. Then $\|\widehat{n}\|_\infty \leqslant 4\beta$

$$n \sim \mathcal{N}(0, \sigma I) \quad \Longrightarrow \quad \text{a.s.} \quad \exists |n_i| > 4\beta \quad \Longrightarrow \quad \|\widehat{n}\|_\infty < \|n\|_\infty$$

[a] An exception are the ℓ_p norms for $1 < p < 2$

Sketch of the proof – 1D signal and Φ smooth

$\|\boldsymbol{\varphi'}\|_{\boldsymbol{\infty}} \boldsymbol{= 1}$, $A = I$

$$G := \begin{pmatrix} -1 & 1 & & \\ & \ddots & \ddots & \\ & & -1 & 1 \end{pmatrix} = (G_1^T, \cdots, G_r^T)^T$$

$$\mathcal{F}_v(u) = \frac{1}{2}\|u - v\|_2^2 + \beta \sum_i \varphi(|G_i u|)$$

$$\begin{aligned} \nabla \mathcal{F}_v(\widehat{u}) = 0 \qquad &\Longrightarrow \qquad v - \widehat{u} = \beta G^T \varphi'(G^T \widehat{u}) \\ &\Longrightarrow \qquad \|v - \widehat{u}\|_\infty \leqslant \beta \|G\|_1 \|\varphi'(G^T \widehat{u})\|_\infty = 2\beta \end{aligned}$$

Question **25** `If` $v = u_o + n$ `for` $n \sim \mathcal{N}(0, \sigma^2 I)$ `Gaussian noise, are we sure to clean` v `from this noise by minimizing` $\mathcal{F}_v$`?`

Denoising in a frame domain

$$x = Wu_{\mathrm{o}} + Wn \qquad Wn \sim \mathcal{N}(0, \sigma^2)$$

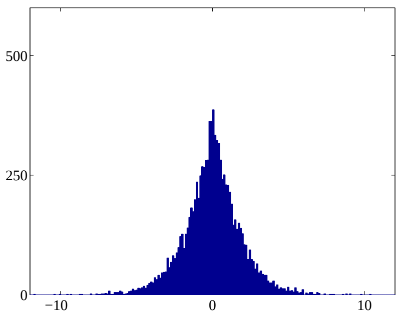
Clean coefficients follow Generalized Gaussians (GG) distributions: [57, 58]

$$f_X(x) = \frac{1}{Z} e^{-\lambda |x|^\alpha}, \quad x \in \mathbb{R}, \qquad \lambda > 0 \quad \alpha > 0$$

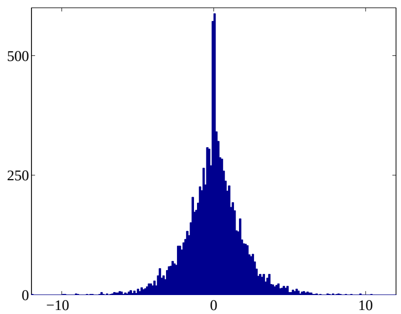
$\widehat{x} = \arg\min_x \mathcal{F}_v(x)$

$$\mathcal{F}_v(x) = \sum_i \Big((x[i] - \langle w_i, v\rangle)^2 + \beta |x[i]|^\alpha \Big) \qquad \beta = 2\sigma^2\lambda$$

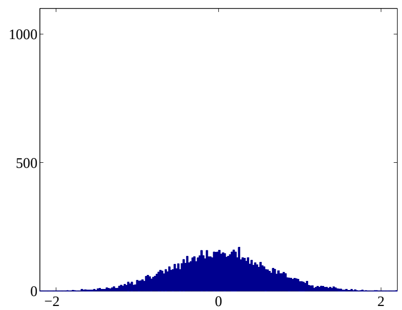
Then $\widehat{u} = W^\dagger \widehat{x}$ where $W^\dagger$ is a left-inverse of W



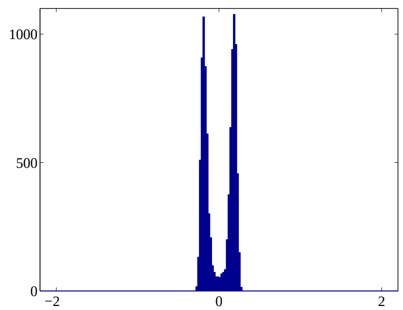
GG for $\alpha = 1.2$, $\lambda = 0.5$



True MAP $\widehat{x}$



$\mathcal{N}(0, \sigma^2)$, $\sigma = 0.6$



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

Histograms for 10 000 independent trials.

Non-smooth at zero noise models

$$f_N(t) = \frac{1}{Z}\exp\left(-\lambda\psi(t)\right) \qquad \psi'(0^-) < \psi'(0^+)$$

$$\mathcal{F}_v(u) = \sum_i \psi(a_i^T u - v[i]) + \beta \sum_i \varphi(\|G_i u\|)$$

ψ is continuous and $\mathcal{C}^2(\mathbb{R}_{>0})$, and φ is $\mathcal{C}^1$

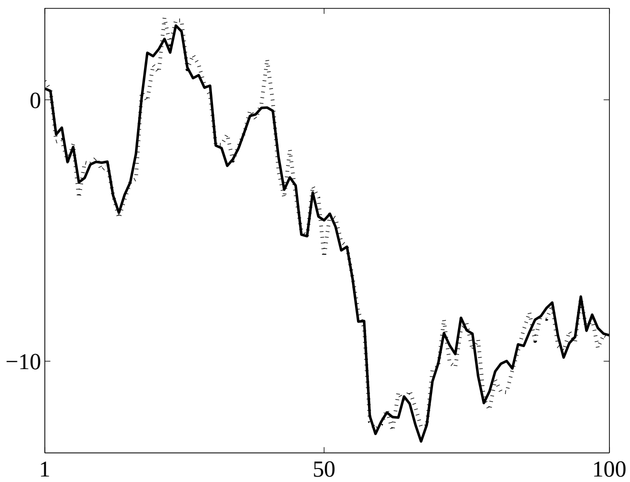
Example: Generalized Gaussian Markov chain under Laplacian noise, MAP denoiser

u_o — Markov chain, $U[i] - U[i+1] \sim f_{\Delta U}$ are i.i.d.

$$f_{\Delta U}(t) = \frac{1}{Z} e^{-\mu |t|^\alpha}$$

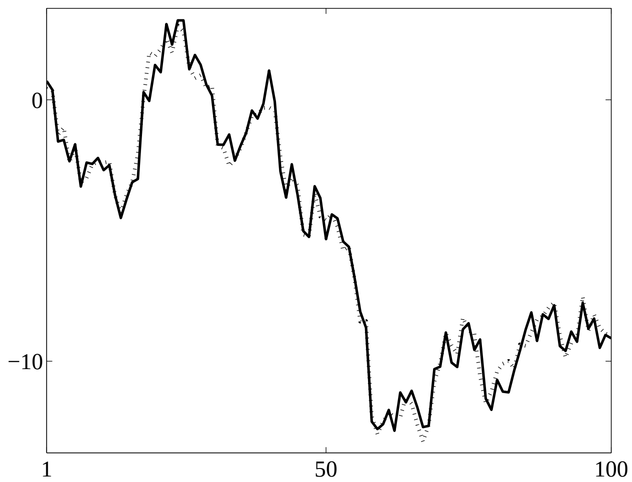
$V = U + N$ where N_i, $1 \leqslant i \leqslant p$ are i.i.d. with $f_N(t) = \frac{\lambda}{2} e^{-\lambda |t|}$

$$\mathcal{F}_v(u) = \sum_{i=1}^{p} \big|u[i] - v[i]\big| + \beta \sum_{i=1}^{p-1} |u[i] - u[i+1]|^\alpha \quad \text{where} \quad \beta = \frac{\mu}{\lambda}$$



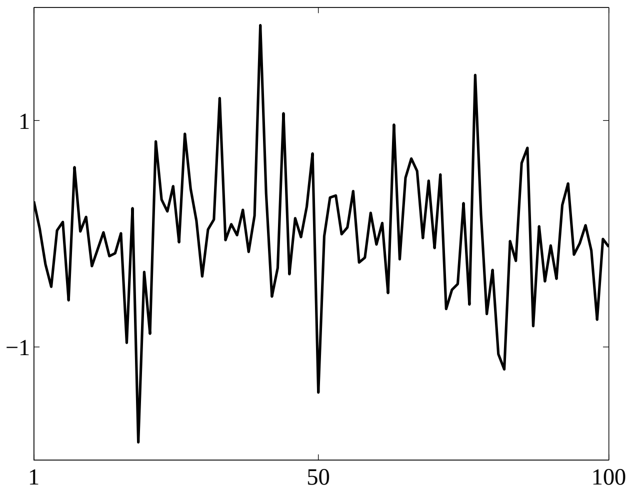
GG Markov chain u_{o} (—) for $\alpha\!=\!1.2$, $\mu\!=\!1$

data $v = u_{\mathrm{o}} + n$ $(\cdots)$

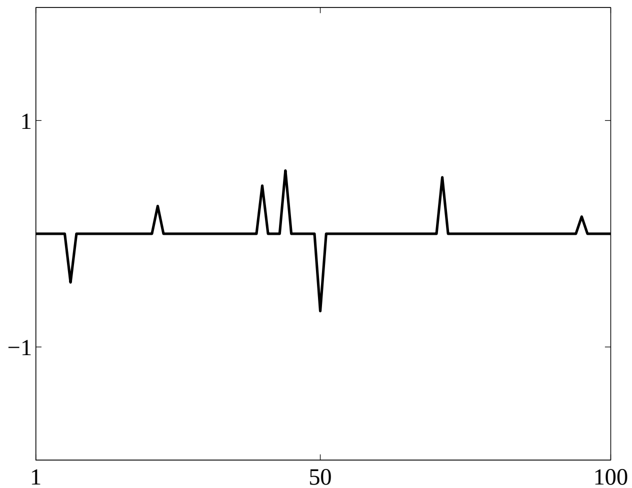


True MAP $\widehat{u}$ (—)

versus the original u_{o} $(\cdots)$



Laplacian i.i.d. noise n for $\lambda = 2.5$



The noise estimate $\widehat{n} = v - \widehat{u}$.

$$u_{\mathrm{o}}[i] \neq v[i] \quad \forall\, i \qquad \sharp\,\{i \ : \ \widehat{n}[i] = 0\} = 93\%$$

From Theorems 6.1 and 2 (p. 52 and p. 56) we know that for $\psi'(0^+) > 0$ and weak assumptions if $\widehat{u}$ is minimizer of $\mathcal{F}_v$, the set $\widehat{h} := \{i : a_i\widehat{u} = v[i]\}$ is typically nonempty and that there is an open subset $\mathcal{O}_{\widehat{h}} \subset \mathbb{R}^q$ and a local minimizer function $\mathcal{U} \in \mathcal{C}^{m-1}$ so that

$$v' \in O_{\widehat{h}}, \quad \hat{u}' = \mathcal{U}(v') \quad \Rightarrow \quad a_i\,\widehat{u}' = v'[i] \;\; \forall\, i \in \widehat{h} \quad \text{and} \quad a_i\,\widehat{u}' \neq v'[i] \;\; \forall\, i \in \widehat{h}^c$$

A consequence:

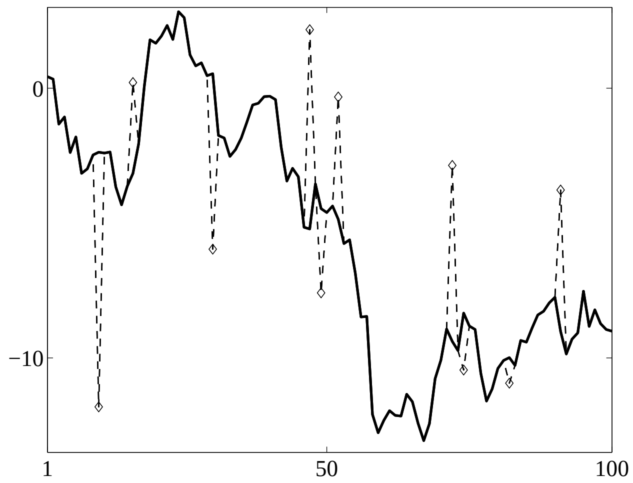
$$\boxed{\begin{array}{c} \mathbb{P}(\widehat{N} = 0) = \mathbb{P}(a_i^T\widehat{U} - V = 0) = \mathbb{P}(V \in O_{\widehat{h}}) = \displaystyle\int_{O_{\widehat{h}}} f_V(v)dv > 0 \\[2ex] \text{whereas} \quad \mathbb{P}(N = 0) = \displaystyle\int f_N(n)\delta(n-0)dn = 0 \end{array}}$$

For all $i \in \widehat{h}$, the regularizer Φ has no influence on the solution.

A Laplace noise model to remove outliers

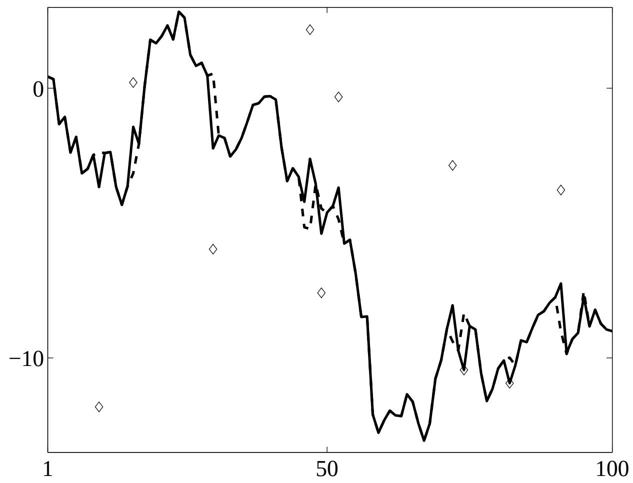
$$\mathcal{F}_v^1(u) = \sum_i |u[i] - v[i]| + \beta \sum_i \sum_{j \in \mathcal{N}_i} \varphi(|u[i] - u[i]|)$$

$\mathcal{N}_i$ neighborhood of pixel i



Original u_o (—), data v (- - -)

with 10% random valued impulse noise.

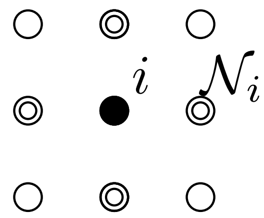


The minimizer $\hat{u}$ of $\mathcal{F}_v^1$ for $\beta = 0.4$ (—)

original u_o (- - -), removed outliers ($\diamond$).

Detection and cleaning of outliers using ℓ_1 data-fidelity [MN 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*

Data v should contain samples that we want to keep ("uncorrupted")

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

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

Theorem 7.2 Let φ be $\mathcal{C}^1$ and convex. Then $\mathcal{F}_v$ has a minimum at $\widehat{u}$ iff

$$\forall\, i \in \widehat{h} \qquad \left|\sum_{j\in\mathcal{N}_i} \varphi'(v[i]-\widehat{u}[j])\right| \leqslant \frac{1}{\beta}$$

$$\forall\, i \in \widehat{h}^c \qquad \sum_{j\in\mathcal{N}_i} \varphi'(\widehat{u}[i]-\widehat{u}[j]) = \frac{\sigma_i}{\beta} \qquad \sigma_i = \operatorname{sign}\left(\sum_{j\in\mathcal{N}_i^2} \varphi'(y[i]-\widehat{u}[j])\right)$$

where $\widehat{h} := \{i \ : \ \widehat{u}[i] = v[i]\}$

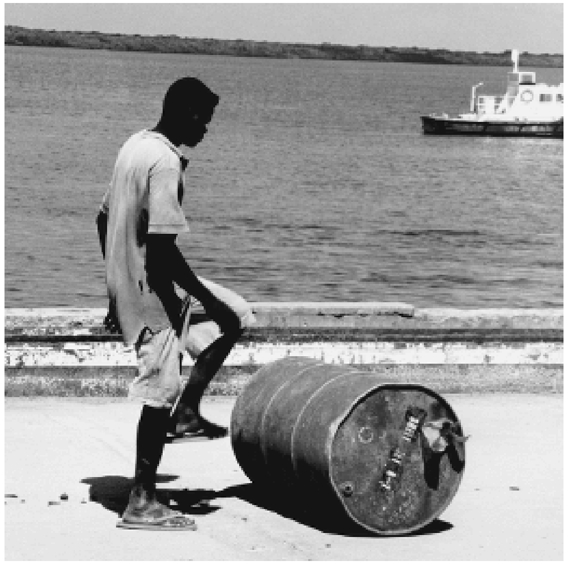
Theorem 7.3 Let φ be strictly convex and $\mathcal{F}_v$ has a minimum at $\widehat{u}$. Consider $\widehat{h} \subset \{1,\ldots,p\}$ and $\sigma_i \in \{-1,1\}$ for any $i \in \widehat{h}^c$ as in Theorem 7.2. Then there is $\rho > 0$ such that for

$$\widetilde{O}_{\widehat{h}} := \left\{ v \in \mathbb{R}^p \;\middle|\; \begin{array}{cc} |v'[i]-v[i]| \leqslant \rho & \forall\, i \in \widehat{h} \\ \sigma_i v'[i] \geqslant \sigma_i v[i] - \rho & \forall\, i \in \widehat{h}^c \end{array} \right\} \subset O_{\widehat{h}}$$

every $\mathcal{F}_{v'}$ reaches its minimum at a $\widehat{u}'$ obeying

$$\begin{array}{rcll} \widehat{u}'[i] & = & v'[i] & \forall\, i \in \widehat{h} \\ \widehat{u}'[i] & \neq & v'[i] & \forall\, i \in \widehat{h}^c \end{array}$$

The components $v[i]$ for $i \in \widehat{h}^c$ are outliers; they can take arbitrary values



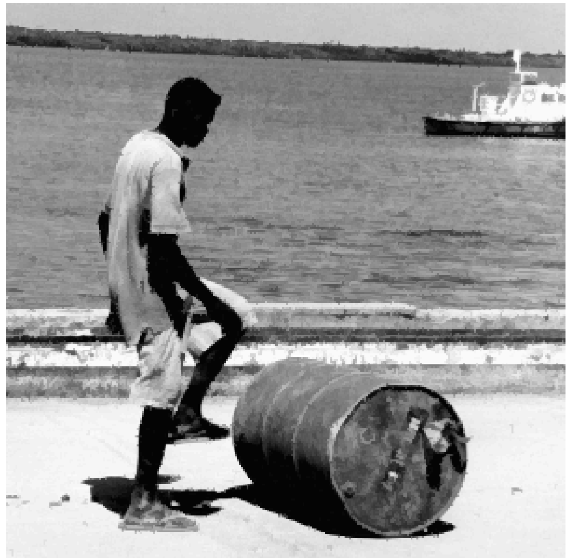
Original image u_o

10% random-valued noise

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

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

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



ℓ_1 data term ($\|\hat{u}-u_o\|_2 = 2934$)

Normal noise removal using a frame and ℓ_1 data-fidelity [Durand, MN 07]

- Data: $v = u_o + n$ where n is centered iid Gaussian noise
- Frame coefficients: $y = Wv = Wu_o + \tilde{n}$ with $\tilde{n}$ centered iid Gaussian noise
- Hard thresholding $\boldsymbol{y_T[i]} := \begin{cases} \mathbf{0} & \text{if } |\boldsymbol{y[i]}| \leqslant \boldsymbol{T} \\ \boldsymbol{y[i]} & \text{if } |\boldsymbol{y[i]}| > \boldsymbol{T} \end{cases}$

  Keep relevant information if T small but outliers appear
- $W^\dagger$ =left inverse of W
- $\tilde{u} = W^\dagger y_T$ — Gibbs oscillations and frame-shaped artifacts
- Hybrid objective methods—combine fidelity to y_T with prior $\Phi(u)$

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

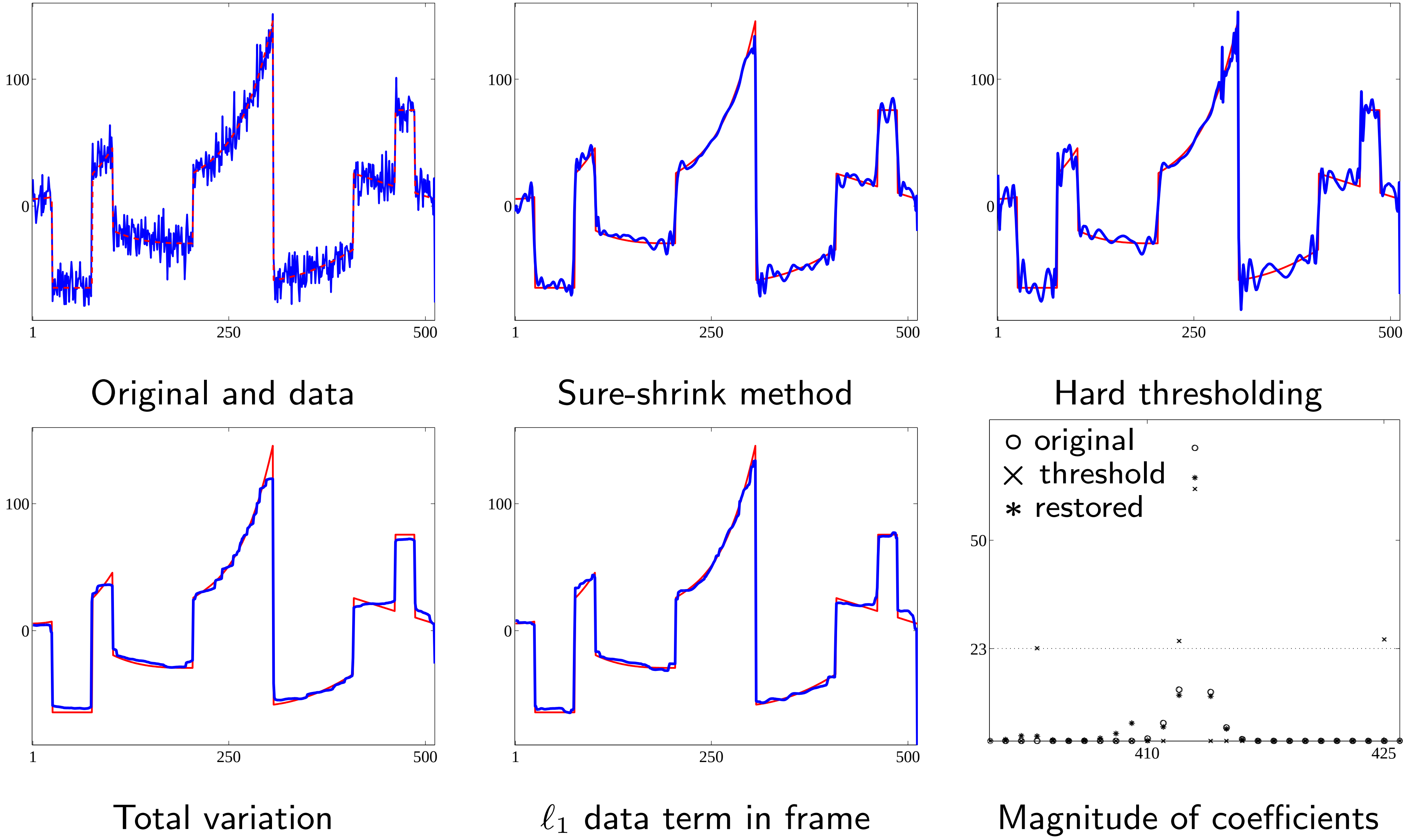
Desiderata: $\boldsymbol{\mathcal{F}_y}$ convex and

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

Then:

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

Motivation: "good" coefficients fitted exactly, "bad" coefficients corrected by the prior.



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

8. Nonsmooth data-fidelity and regularization

Consequence of §4 and §6: if Φ and Ψ non-smooth, $\begin{cases} \boldsymbol{G_i \hat{u} = 0} & \text{for} \quad \boldsymbol{i} \in \boldsymbol{\hat{h}_\varphi} \neq \varnothing \\ \boldsymbol{a_i \widehat{u} = v[i]} & \text{for} \quad \boldsymbol{i} \in \boldsymbol{\hat{h}_\psi} \neq \varnothing \end{cases}$

L_1-TV objective

[T. Chan, S. Esedoglu 05]

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

– $\exists\, \widehat{u} = 1\!\!1_\Sigma$ (Ω convex $\Rightarrow$ $\Sigma \subset \Omega$ and $\widehat{u}$ unique for almost every $\beta > 0$)

– contrast invariance: if $\widehat{u}$ minimizes for $v = 1\!\!1_\Omega$ then $c\widehat{u}$ minimizes $\mathcal{F}_{cv}$

– critical values β^* $\begin{cases} \beta < \beta^* & \Rightarrow \quad \text{objects in } \widehat{u} \text{ with good contrast} \\ \beta > \beta^* & \Rightarrow \quad \text{they suddenly disappear} \end{cases}$

$\Longrightarrow$ data-driven scale selection

Binary images by L1 − TV

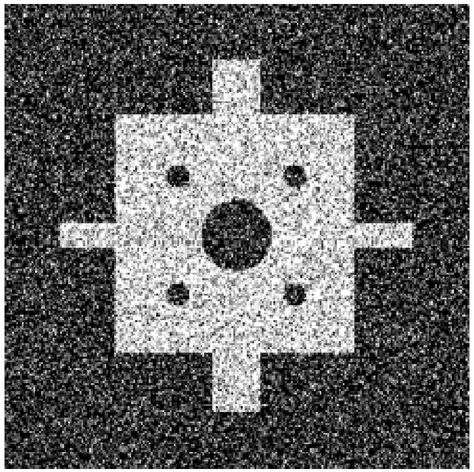
[T. Chan, S. Esedoglu, MN 06]

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

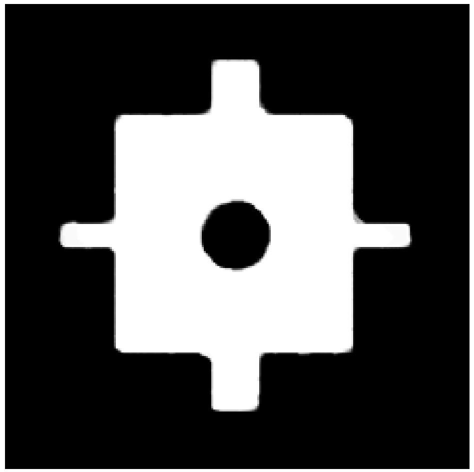
$$\hat{\Sigma} = \arg\min_{\Sigma} \left\{ \left\| \mathbb{1}_\Sigma - \mathbb{1}_\Omega \right\|_2^2 + \beta \mathrm{TV}(\mathbb{1}_\Sigma) \right\} \qquad \text{nonconvex problem} \qquad (\star)$$

usual techniques (curve evolution, level-sets) fail

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



Data



Restored

This work gave rise to numerous convex relaxation methods to solve non-convex imaging problems

Multiplicative noise removal on frame coefficients [Durand, Fadili, MN 09]

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

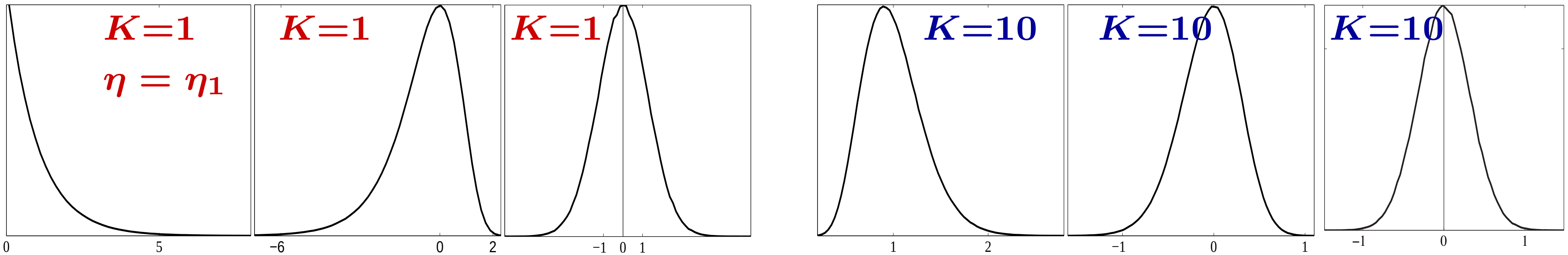
– Original image: $\boldsymbol{S_o}$

– One shot: $\boldsymbol{\Sigma_k = S_o \eta_k}$

– Data: $\boldsymbol{\Sigma = \frac{1}{K}\sum_{k=1}^{K}\Sigma_k = S_o \frac{1}{K}\sum_{k=1}^{K}\eta_k = S_o\,\eta}$ where $\mathrm{pdf}(\eta) =$ Gamma density

– Log-data: $\boldsymbol{v = \log\Sigma = \log S_o + \log\eta = u_0 + n}$

– Frame Coefficients: $\boldsymbol{y = Wv = Wu_0 + Wn}$ (W curvelets)



$\mathbf{pdf}(\boldsymbol{\eta}) = \mathbf{pdf}(\boldsymbol{\eta_k})$ $\mathbf{pdf}(\boldsymbol{n})$ $\mathbf{pdf}\big(\boldsymbol{Wn}\big)$ $\mathbf{pdf}(\boldsymbol{\eta})$ $\mathbf{pdf}(\boldsymbol{n})$ $\mathbf{pdf}\big(\boldsymbol{Wn}\big)$

Question **26** Comment the noise distribution of Wn

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

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

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

$$\boxed{\begin{array}{l} \mathcal{F}_y(x) = \lambda_0 \sum_{i \in I_0} \big|x[i]\big| + \lambda_1 \sum_{i \in I_1} \big|x[i] - y[i]\big| + \|W^\dagger x\|_{\mathrm{TV}} \\ \hat{S} = B \exp\left(W^\dagger \hat{x}\right), \text{ where } W^\dagger \text{ left inverse, } B \text{ bias correction} \end{array}}$$

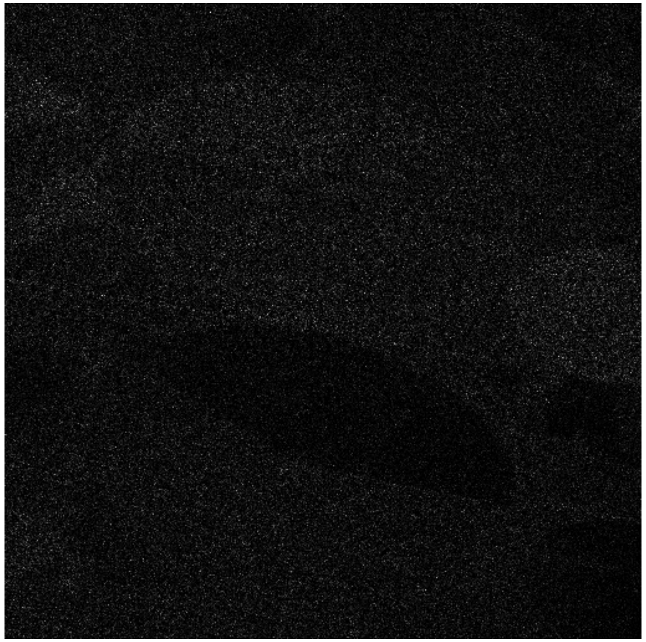
Some comparisons

– **BS** [Chesneau,Fadili,Starck 08]: Block-Stein thresholds the curvelet coefficients, $\approx$ minimax(large class of images with additive noises), optimal threshold $\mathfrak{T} = 4.50524$

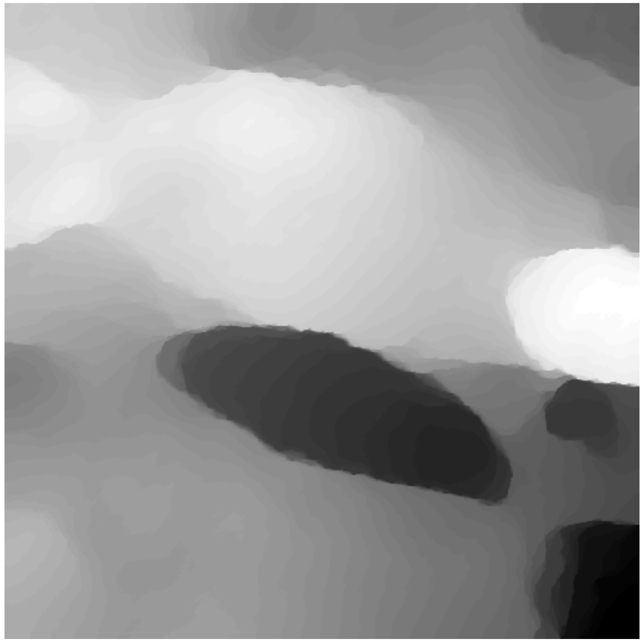
– **MAP** [Aubert,Aujol 08]: $\Psi = -$ Log-Likelihood(Σ), $\Phi = \mathrm{TV}(\Sigma)$

– **ISS** [Shi,Osher 08]: relaxed inverse scale-space for $\mathcal{F}_v(u) = \|v - u\|_2^2 + \beta \mathrm{TV}(u)$ $\approx$ MAP(u)

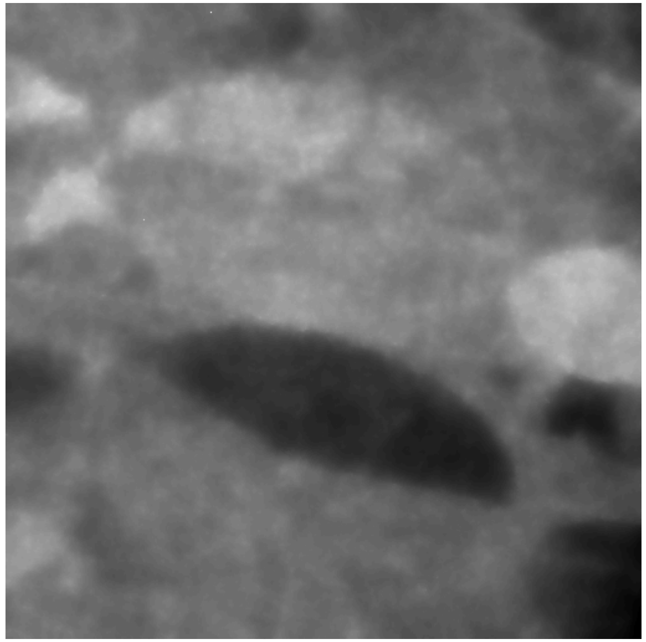
stopping rule: $k^* = \max\{k \in \mathrm{N} : \mathrm{Var}(u^{(k)} - u_o) \geqslant \mathrm{Var}(n)\}$.



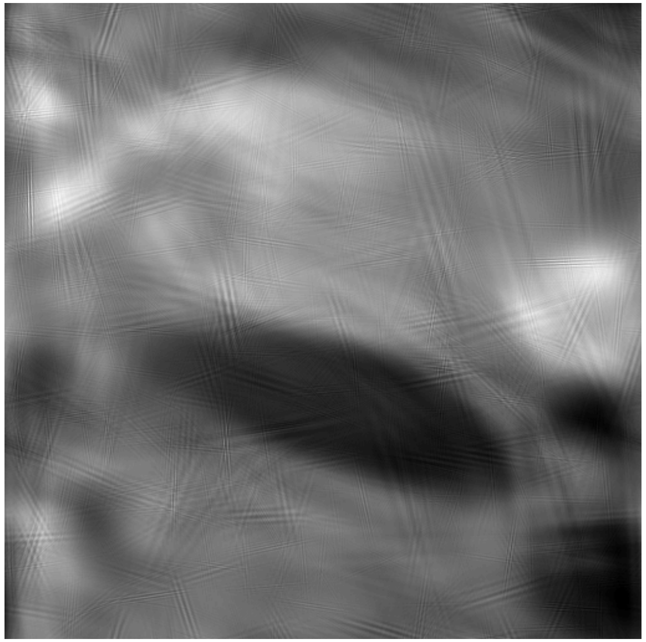
Noisy Fields $K = 1$ (512×512)



ISS: PSNR=9.59

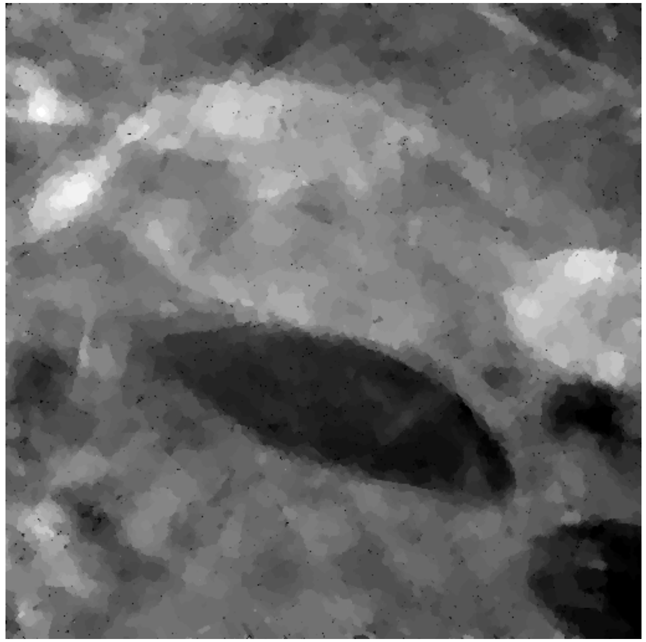


MAP: PSNR=15.74

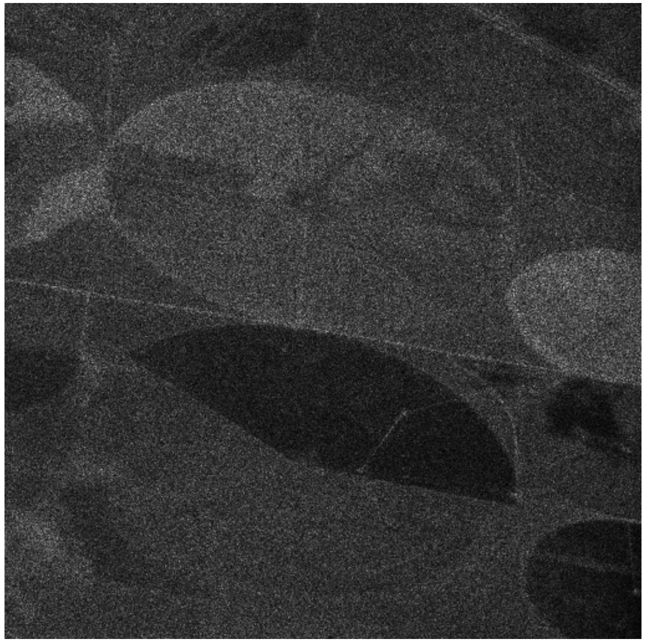


BS: PSNR=22.52

Fields (original)

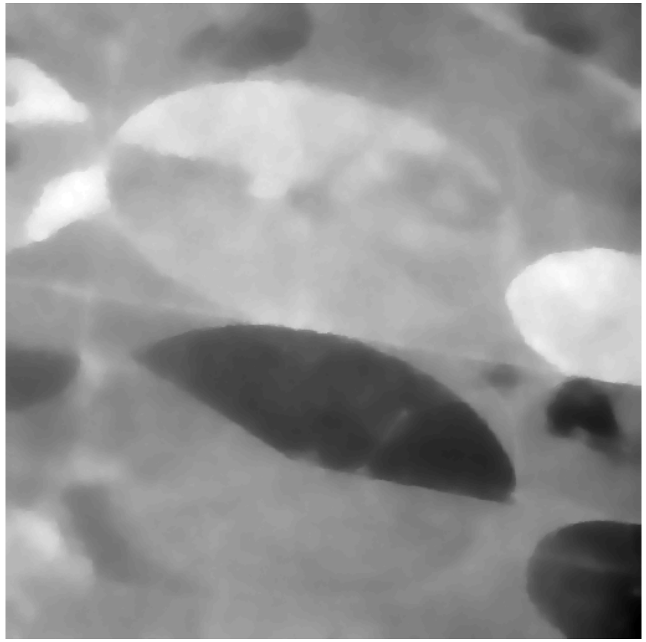


ℓ_1-TV: PSNR=22.89

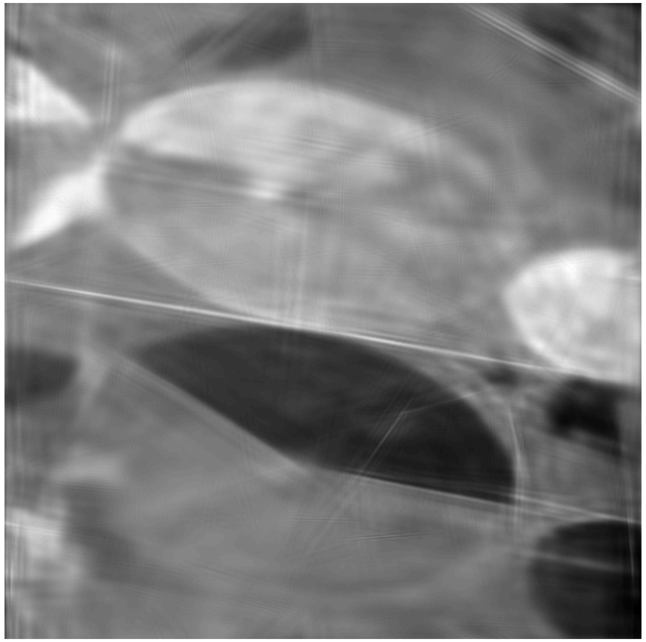


Noisy $K = 10$

ISS: PSNR=25.36

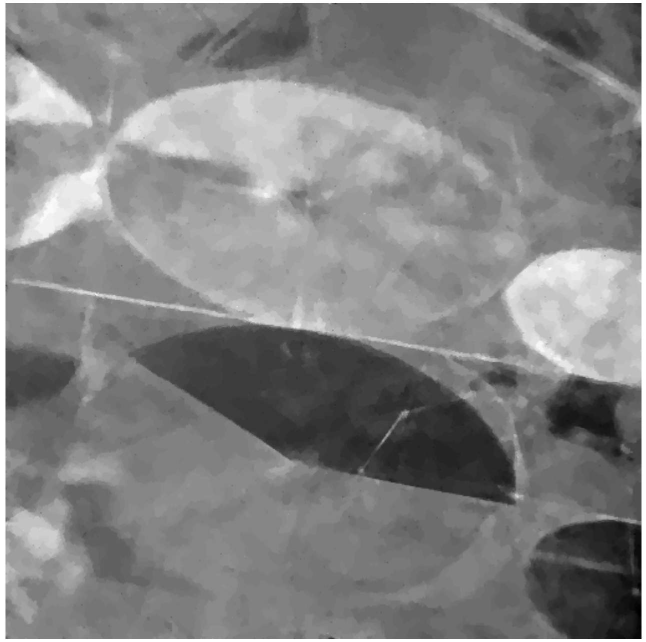


MAP: PSNR=17.13



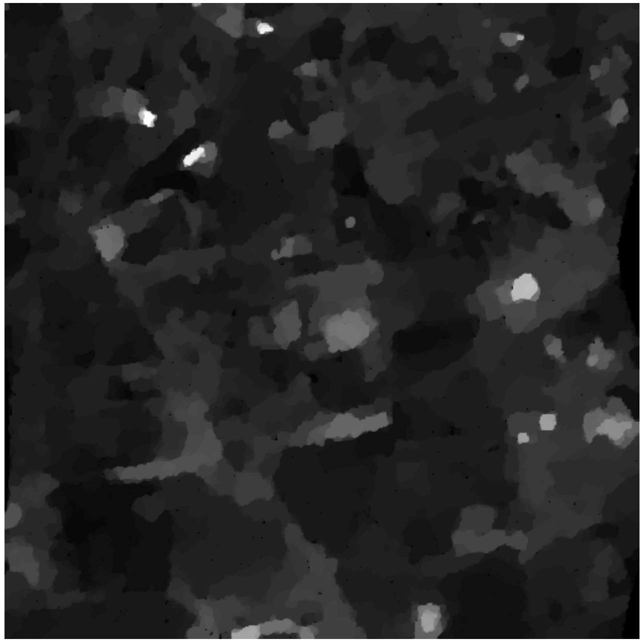
BS: PSNR=27.24

Fields (original)



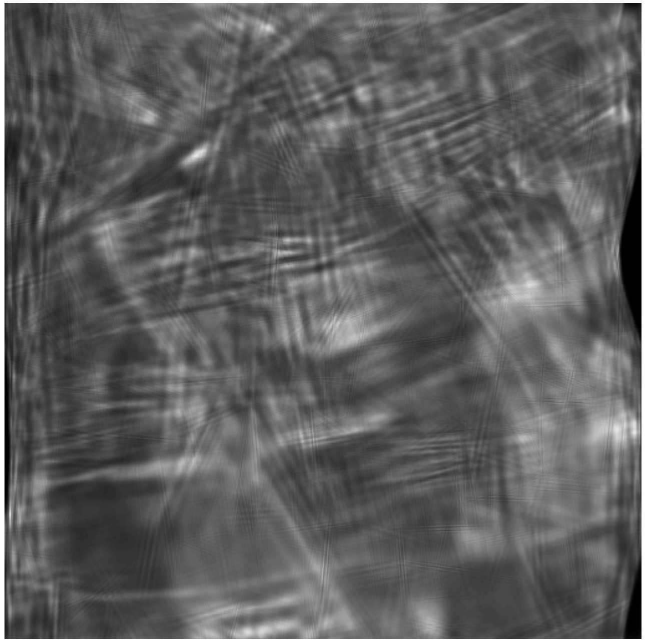
ℓ_1-TV: PSNR=28.04

Noisy City $K = 1$ (512×512)



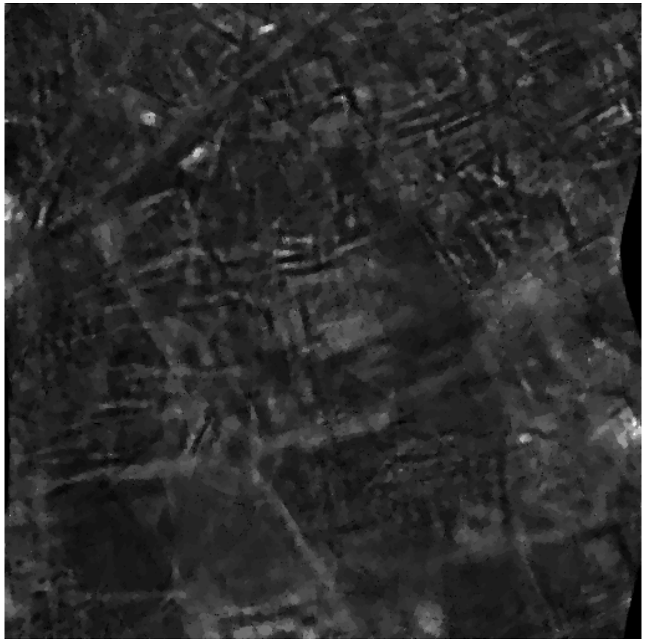
SO: PSNR=18.39

MAP: PSNR=22.18



BS: PSNR=22.25

City (original)

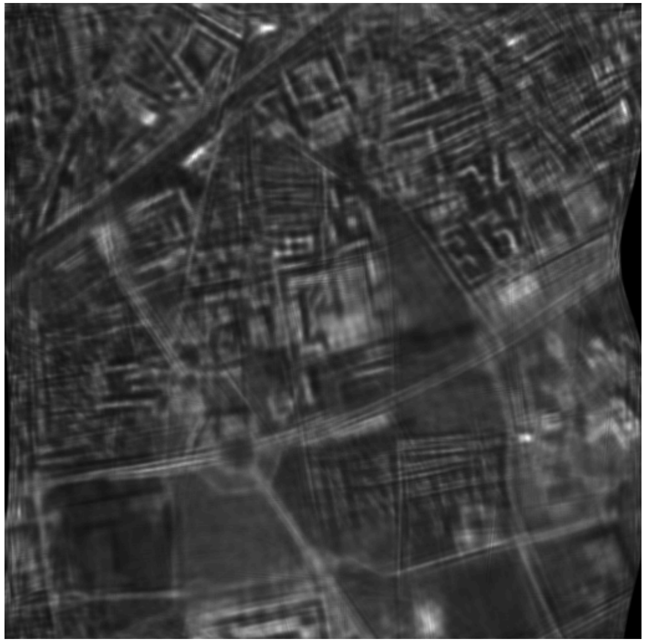


ℓ_1-TV: PSNR=22.64

Noisy $K=4$

ISS: PSNR=24.40

MAP: PSNR=24.55



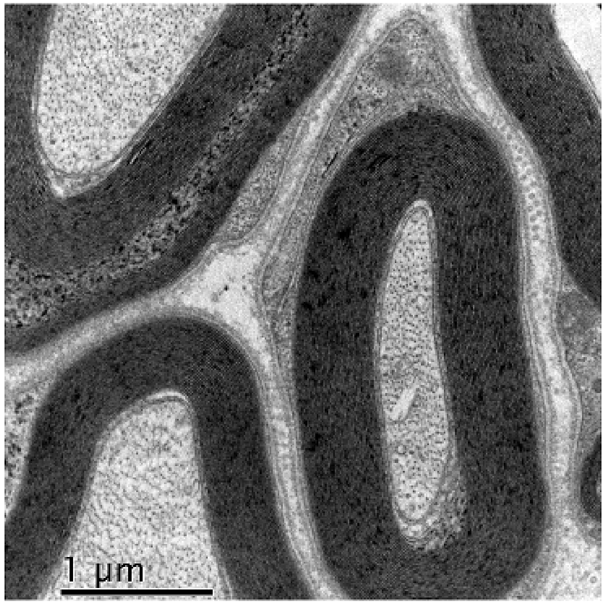
BS: PSNR=24.92

City (original)

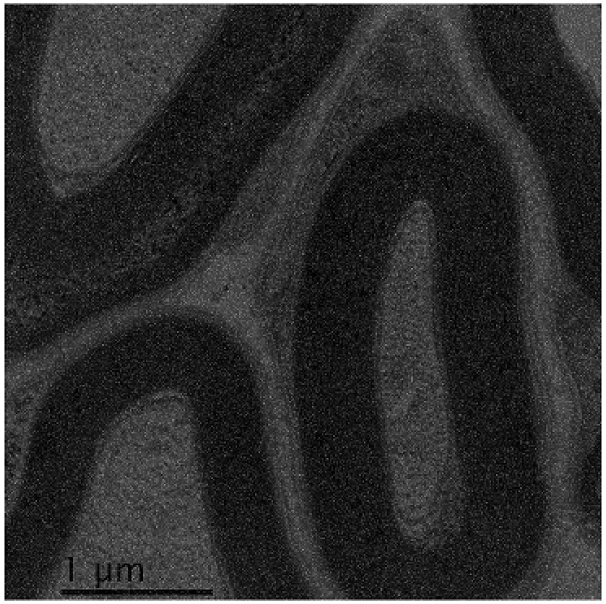
ℓ_1-TV: PSNR=25.84

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/

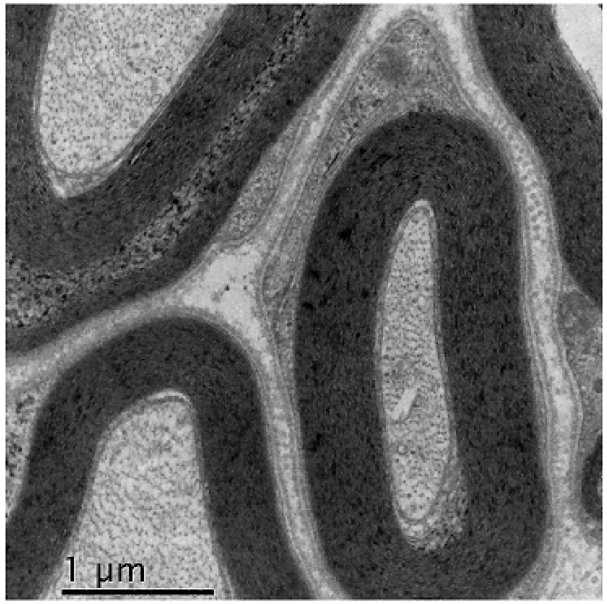
Scanning transmission electron microscopy (2048×2048 image)



true image



noisy image



restoration

ℓ_1 data-fidelity with concave regularization

[MN, Ng, Tam 13]

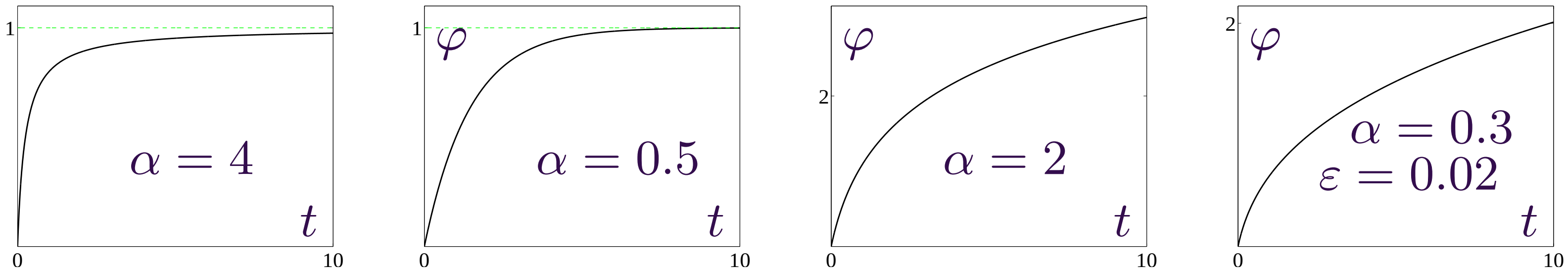
$$\mathcal{F}_v(u) = \sum_{i \in I} \left| a_i u - v[i] \right| + \beta \sum_{j \in J} \varphi(\|G_j u\|_2), \quad \varphi'(0^+) > 0, \ \varphi''(t) < 0, \ \forall t \geqslant 0$$

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

No conditions on the rank of the matrix formed by the rows a_i

H8.2 φ is strictly concave on $[0, +\infty)$, increasing, $\varphi'' \leqslant 0$ and $\lim_{t \to \infty} \varphi''(t) \nearrow 0$

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

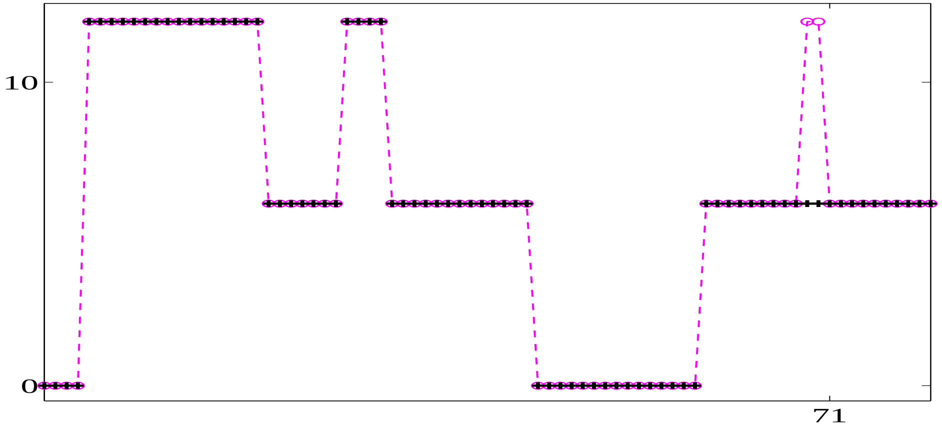


Motivation

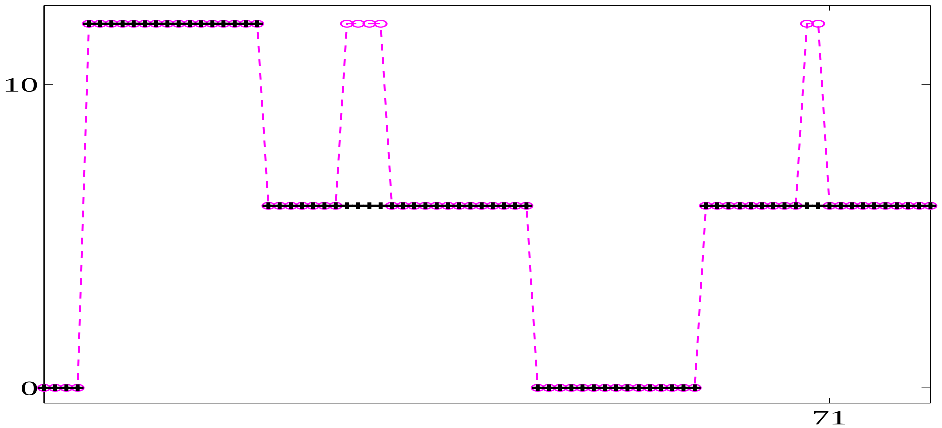
- New family of objective functions
- $\mathcal{F}_v$ can be seen as an extension of $L1 - \mathrm{TV}$
- $\widehat{u}$ – (local) minimizer of $\mathcal{F}_v$ $\overset{?}{\Longrightarrow}$ many i, j such that $a_i \widehat{u} = v[i]$ and $G_j \widehat{u} = 0$

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 $\alpha = 4$

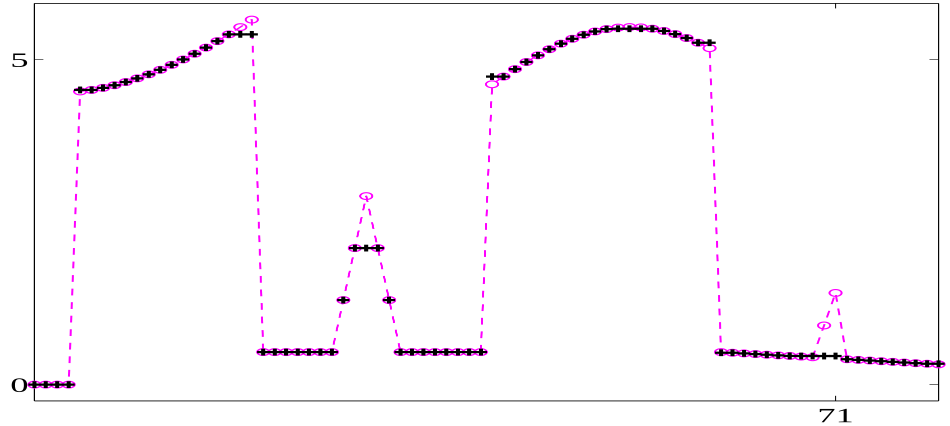


$\beta \in \{78, \cdots, 156\}$

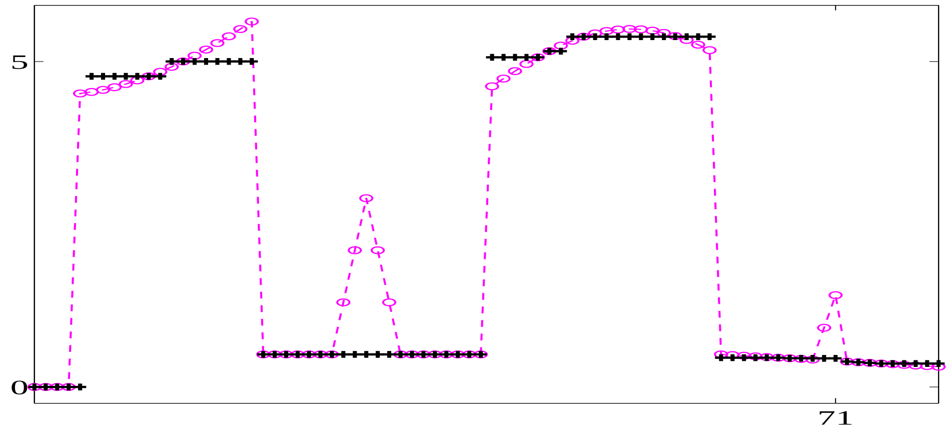


$\beta \in \{157, \cdots, 400\}$

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

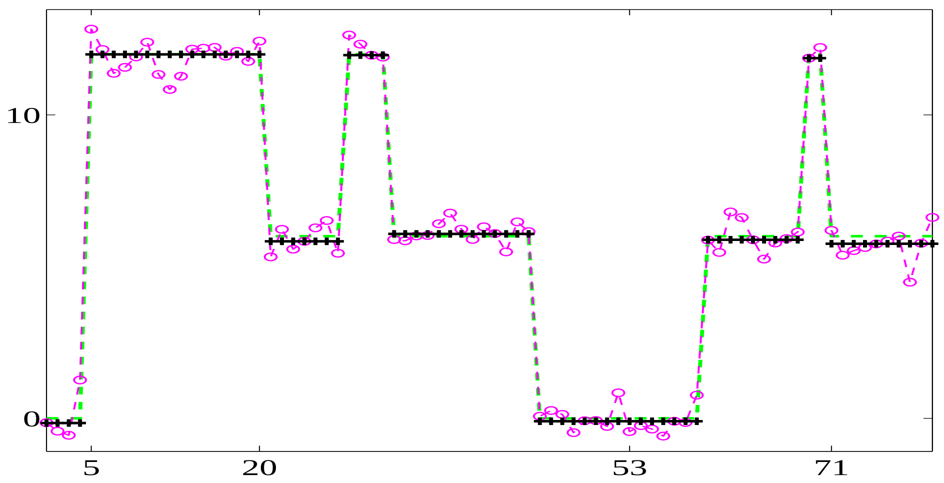


$\beta \in 0.1 \times \{10, \cdots, 14\}$

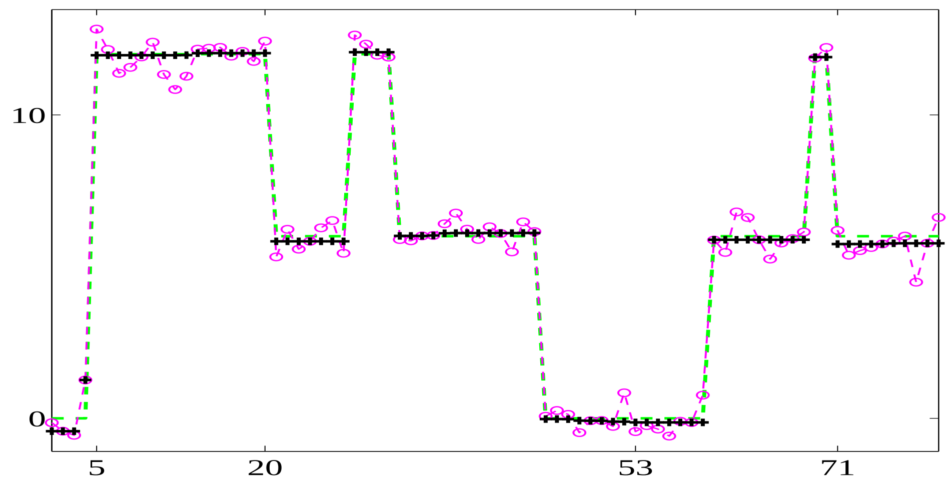


$\beta \in 0.1 \times \{16, \cdots, 30\}$

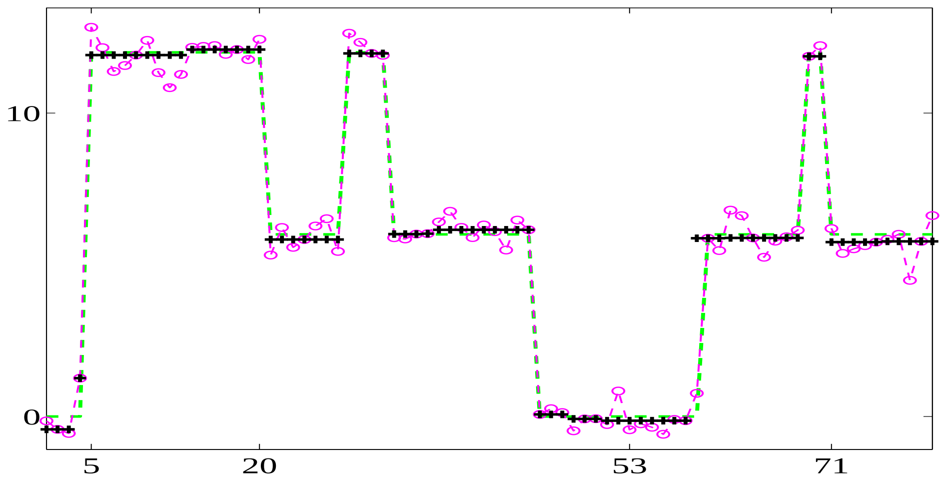
Data samples (ooo), Minimizer samples $\widehat{u}[i]$ (+++).



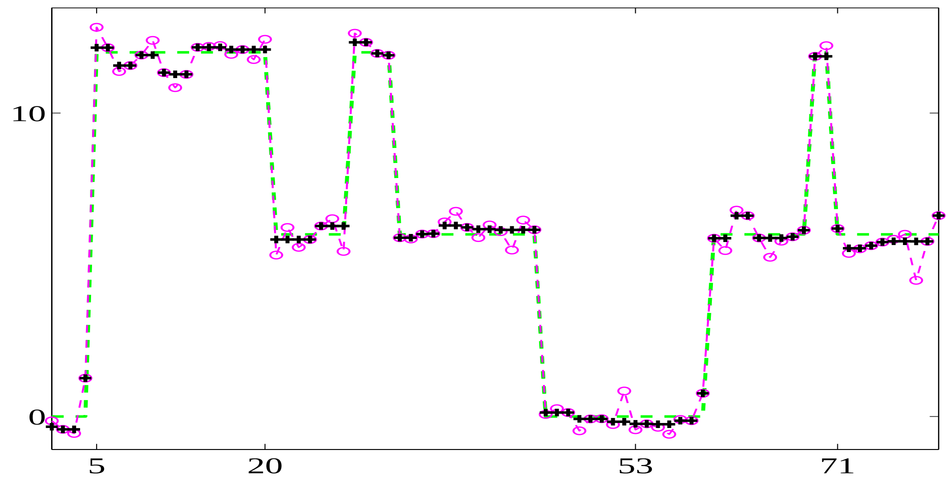
(a) $\varphi(t) = \frac{\alpha\, t}{\alpha\, t+1}$, $\alpha = 4$, $\beta = 3$



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



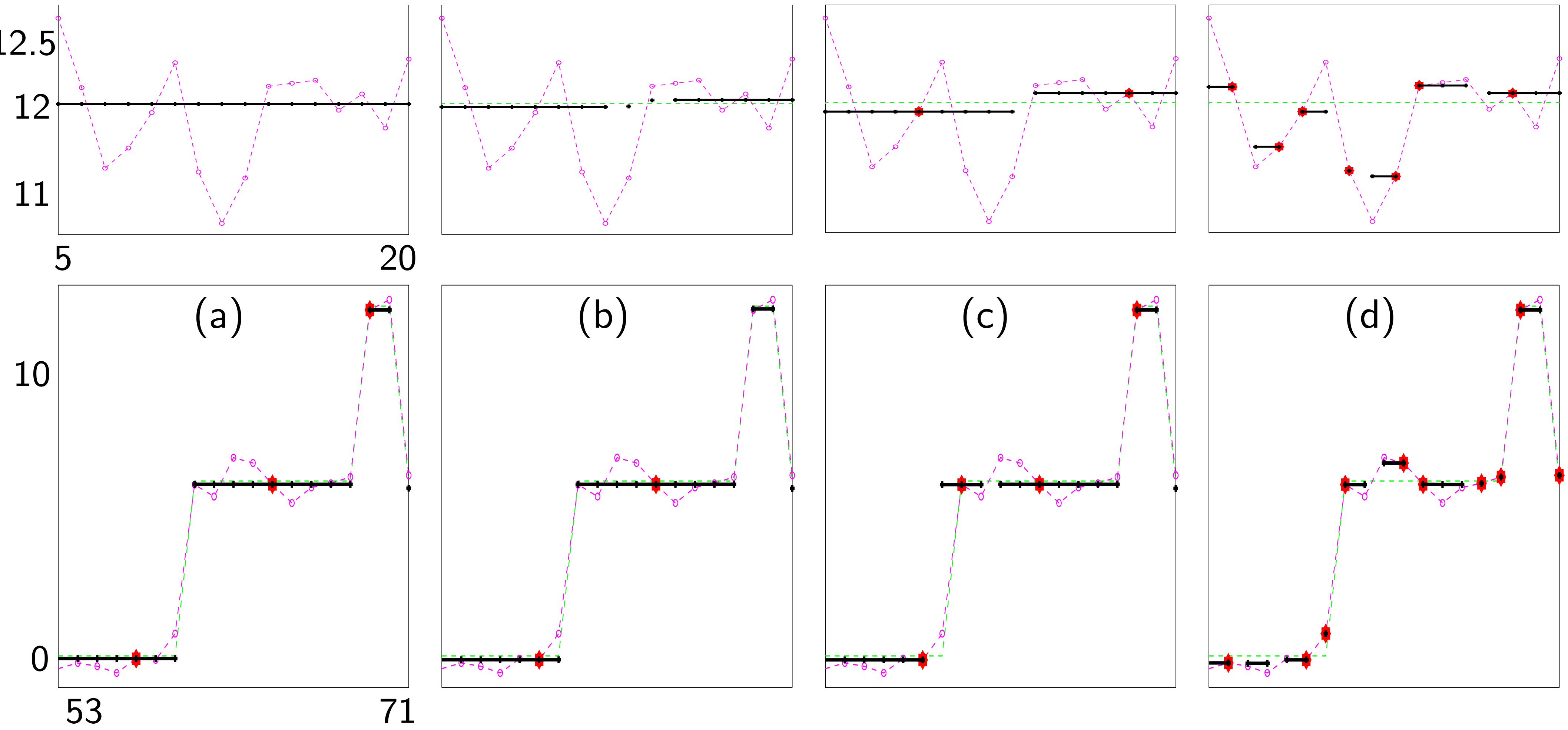
(c) $\varphi(t) = \ln(\alpha t + 1)$, $\alpha = 2$, $\beta = 1.3$



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

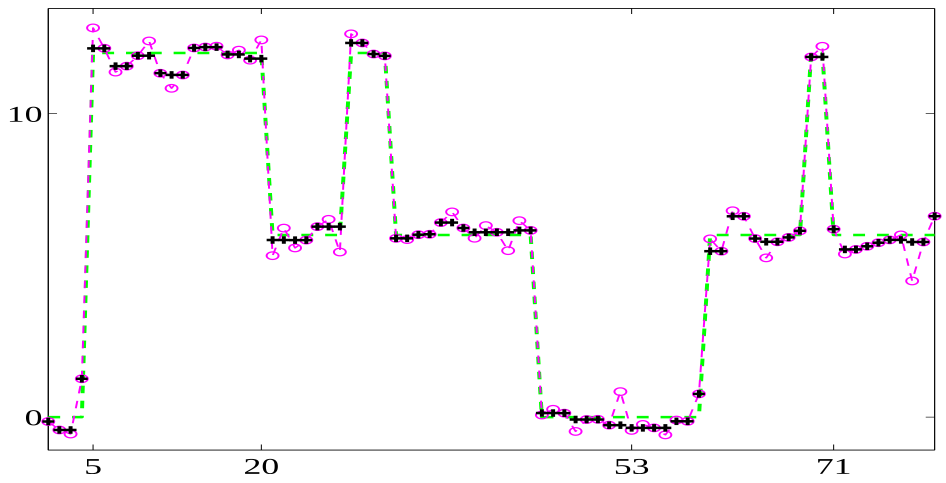
Denoising: Data samples (ooo) are corrupted with Gaussian noise. Minimizer samples $\widehat{u}[i]$ (+++). Original (– – –). $\boldsymbol{\beta}$—the largest value so that the gate at 71 survives.

Zooms



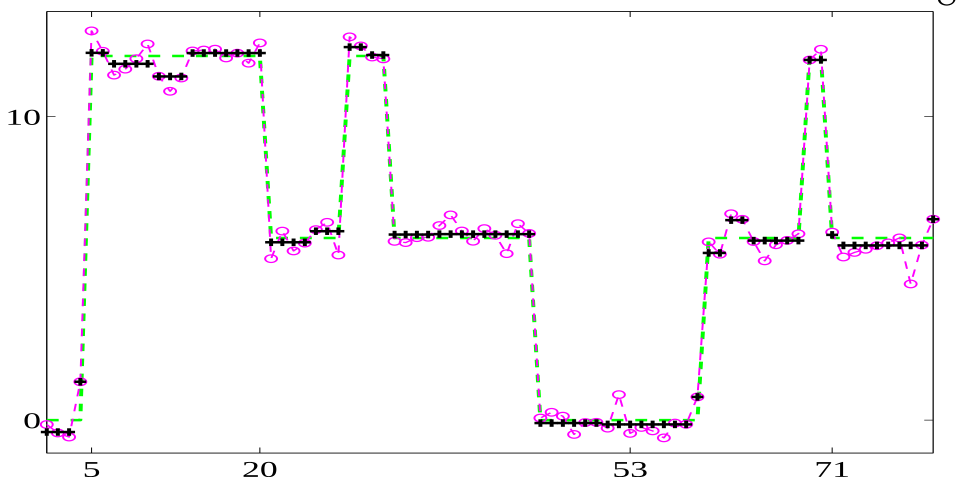
Constant pieces—solid black line.

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



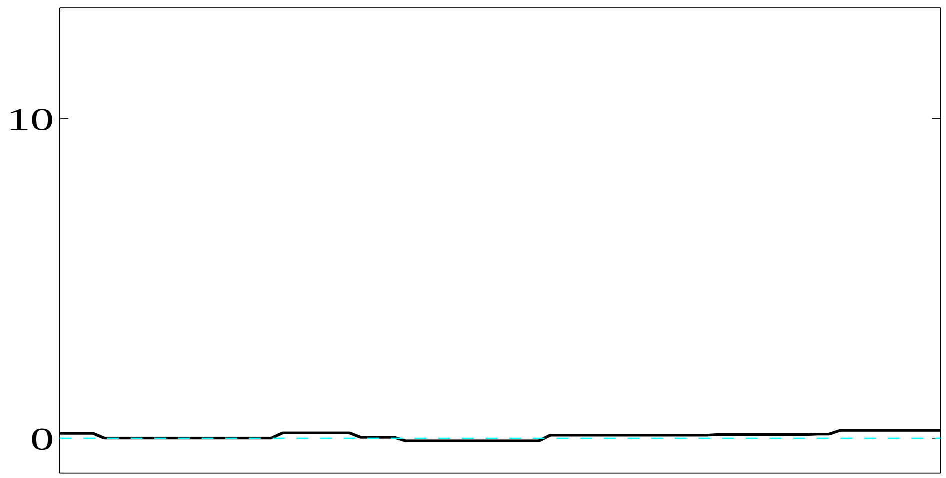
$\varphi(t) = t,\ \beta = 0.8 \quad (\ell_1 - \mathrm{TV})$

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



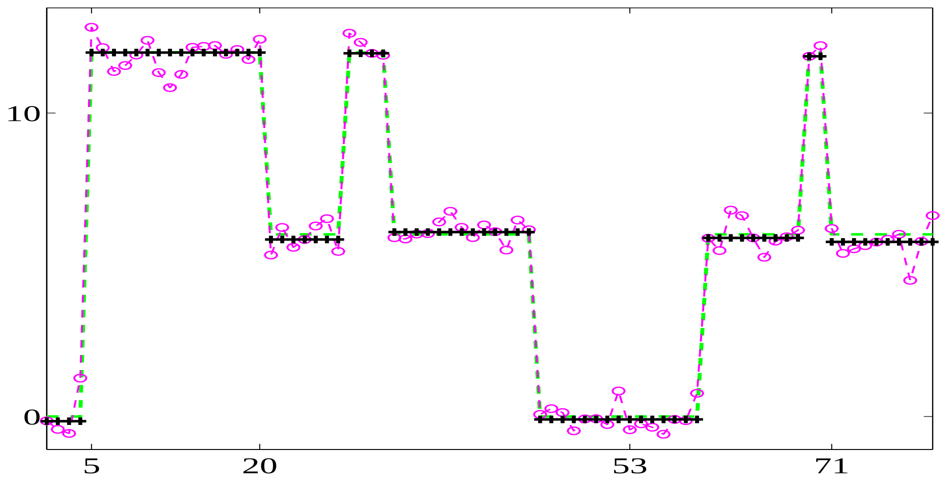
the minimizer for $\varphi(t) = \frac{\alpha\ t}{\alpha\ t+1},\ \alpha = 4,\ \beta = 3$

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



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

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



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

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

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

Question **27** Can you sketch the main properties of the minimizers of $\mathcal{F}_v$?

Question **28** What seems being the role of the asymptotic of φ?

Numerical evidence:

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

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

Result known for the minimizers of $L_1 - \mathrm{TV}$ [9]

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 29 Does $\mathcal{F}_v$ have a global minimizer for any v?

Question 30 Determine $\varphi''(u)$ for $u \in \mathbb{R} \setminus \{0\}$.

Question 31 Show that $\forall\, v \in \mathbb{R}$, any minimizer $\widehat{u}$ of $\mathcal{F}_v$ obeys $\widehat{u} \in \{0, v\}$.

Question 32 Can you extend this result to the other φ on p. 83?

Facts

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

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

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

(b) Then $\widehat{\boldsymbol{u}}$ is the **unique** point solving the liner system

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

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

(c) Contrast invariance' of (local) minimizers

(d) 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**

(e) Each (local) minimizer of $\mathcal{F}_v$ is strict

Proposition 8.1. Let H8.2 hold and $\widehat{u}$ is a local minimizer of $\mathcal{F}_v$. Then $\widehat{I}_0 \cup \widehat{J}_0 \neq \varnothing$.

$$\begin{aligned} \mathcal{K}_{\widehat{u}} &= \{w \in \mathbb{R}^p \;:\; a_i w = v[i] \;\; \forall\, i \in \widehat{I}_0 \;\text{ and }\; G_j w = 0 \;\; \forall\, j \in \widehat{J}_0\} \qquad (\diamond) \\ K_{\widehat{u}} &= \{w \in \mathbb{R}^p \;:\; a_i w = 0 \qquad \forall\, i \in \widehat{I}_0 \;\text{ and }\; G_j w = 0 \;\; \forall\, j \in \widehat{J}_0\} \end{aligned}$$

$$\widehat{u} \in \mathcal{K}_{\widehat{u}} \quad \text{and} \quad \widehat{u} + w \in \mathcal{K}_{\widehat{u}} \;\; \forall\, w \in K_{\widehat{u}}$$

$$F := \mathcal{F}_v|_{\mathcal{K}_{\widehat{u}}} \qquad F(u) = \sum_{i \in \widehat{I}_0^c} |a_i u - v[i]| + \beta \sum_{j \in \widehat{J}_0^c} \varphi(\|G_i u\|)$$

Lemma 8.1. Suppose also that $\dim K_{\widehat{u}} > 1$. Then $\quad w^\intercal D^2 F(\widehat{u}) w < 0 \quad \forall\, w \in K_{\widehat{u}}$.

Details on the main results: Under additional assumption, $\exists \rho > 0$ such that

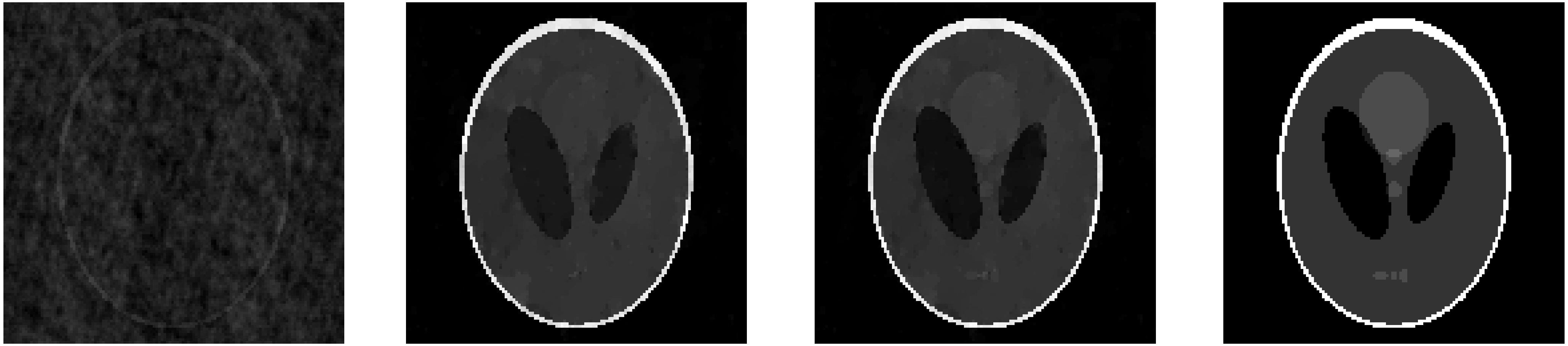
$$\forall\, w \in K_{\widehat{u}} \cap B(0, \rho) \qquad F(\widehat{u}) = \mathcal{F}(\widehat{u}) = \mathcal{F}(\widehat{u} + w) = F(\widehat{u} + w)$$

Then F should have a (local) minimum at $\widehat{u}$ and

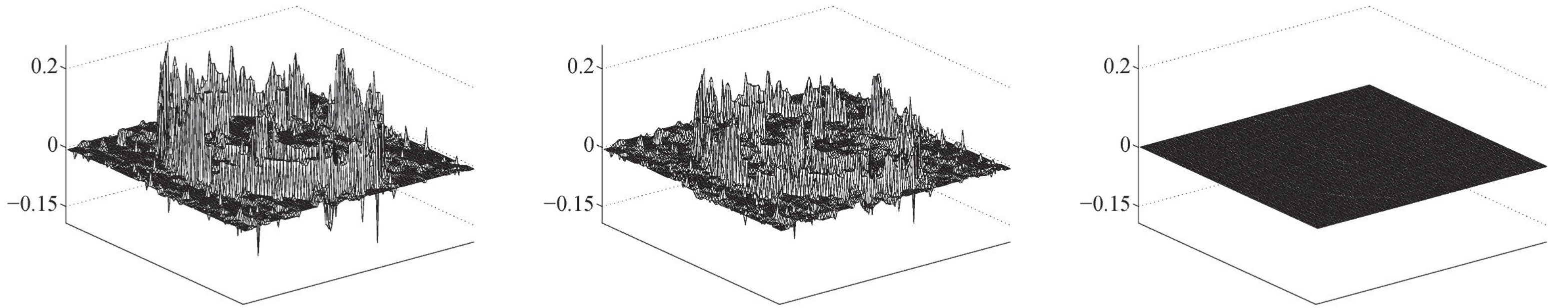
satisfy $w^\intercal D^2 F(\widehat{u}) w > 0 \;\forall\, w \in K_{\widehat{u}} \cap B(0, \rho)$ – impossible by Lemma 8.1.

$$\Longrightarrow \quad \dim K_{\widehat{u}} = 0 \quad \overset{\exists \widehat{u} \;\; (\mathrm{a})}{\Longrightarrow} \quad \mathcal{K}_{\widehat{u}} = \{\widehat{u}\} \quad \overset{(\diamond)}{\Longrightarrow} \quad (\mathrm{b}), \;\; (\mathrm{d}) \text{ and } (\mathrm{e})$$

MR Image Reconstruction from Highly Undersampled Data



0-filling Fourier | $\|\cdot\|_2^2 + \mathrm{TV}$ | $\|\cdot\|_1 + \mathrm{TV}$ | ℓ_1-concave

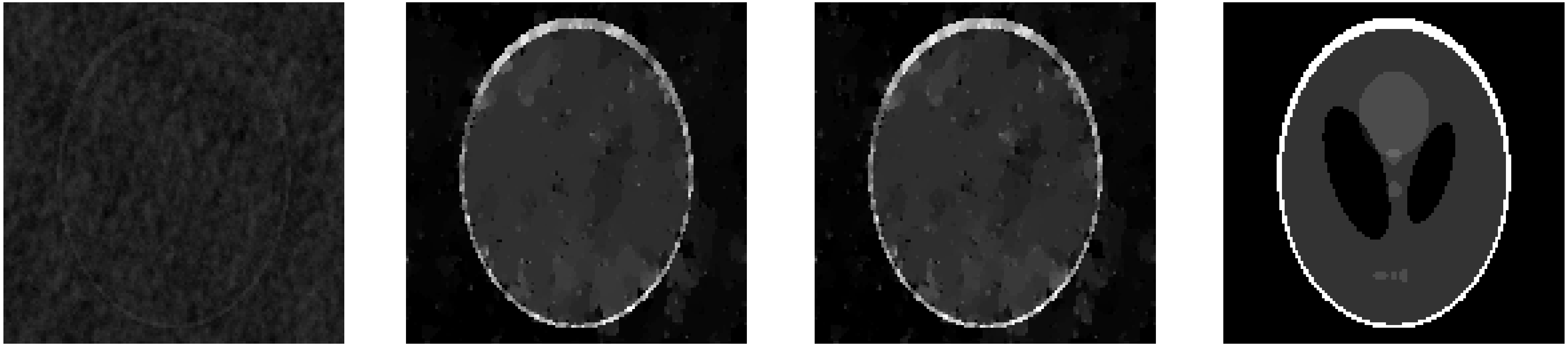


Reconstructed images from **7% noisy** randomly selected samples in the k-space.

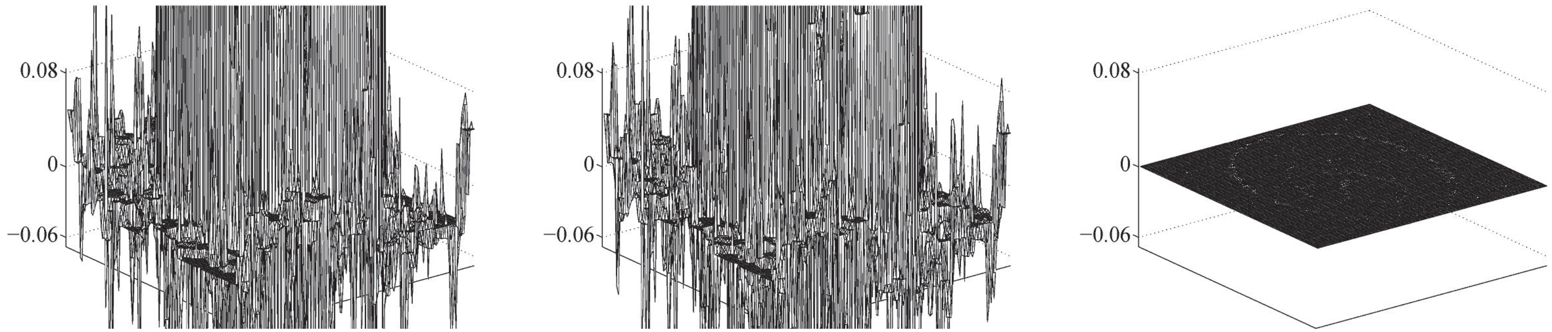
ℓ_1-concave for $\varphi(t) = \dfrac{\alpha t}{\alpha t + 1}$.

Here the best CS recommendation is $\|\cdot\|_2^2 + \mathrm{TV}$. Observe $\|\cdot\|_1 + \mathrm{TV}$.

MR Image Reconstruction from Highly Undersampled Data



0-filling Fourier | $\|\cdot\|_2^2$+TV | $\|\cdot\|_1$+TV | ℓ_1-concave



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

ℓ_1-concave for $\varphi(t) = \dfrac{\alpha t}{\alpha t + 1}$.

9. Fully smoothed $\ell_1 - \mathrm{TV}$

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

$$\boldsymbol{\Psi(u,v) = \sum_{i=1}^{p} \psi_{\alpha_1}(u[i] - v[i])} \quad \text{and} \quad \boldsymbol{\Phi(u) = \sum_{i} \varphi_{\alpha_2}(|G_i u|)}$$

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

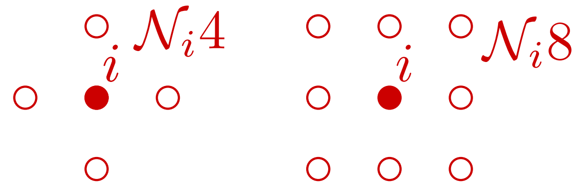
$\varphi(\cdot) := \varphi(\cdot, \alpha_2)$

$(\alpha_1, \alpha_2) > 0$

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

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

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



$(\boldsymbol{\psi}, \boldsymbol{\varphi})$ 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 $\mathcal{C}^{m\geqslant 2}$-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 $\alpha > 0$

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

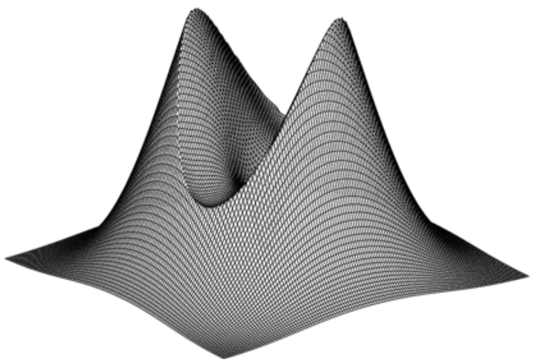
$$\theta'(t, \alpha) := \frac{d}{dt}\theta(t, \alpha)$$

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

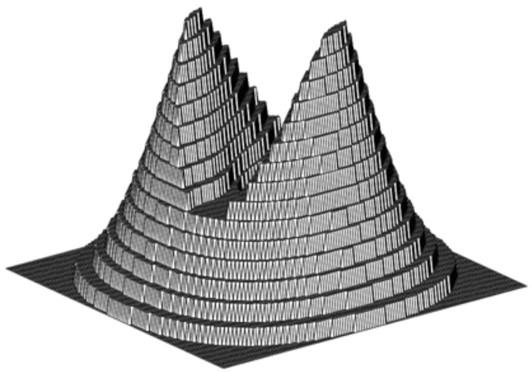
$\Longrightarrow \quad \boldsymbol{\mathcal{F}_v}$ is a fully smoothed $\ell_1 - \mathrm{TV}$ objective.

Goal: to obtain a restoration $\widehat{u}$ of v whose pixels are all different from each other while being close to v but "better" than v

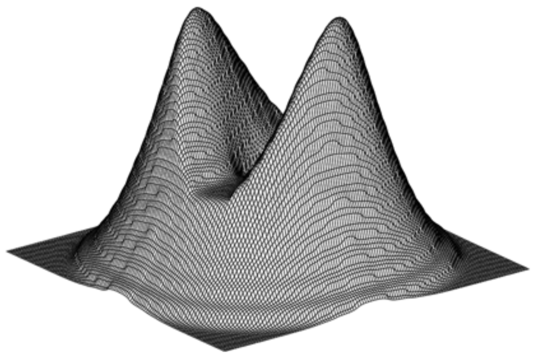
- By H1 $\widehat{u}$ should be nowhere constant
- H2 enables the recovery of edges and details
- $\widehat{u}$ will remain close to v by "nearly L1" data term
- Some removal of the quantization noise is expected



Real-valued original



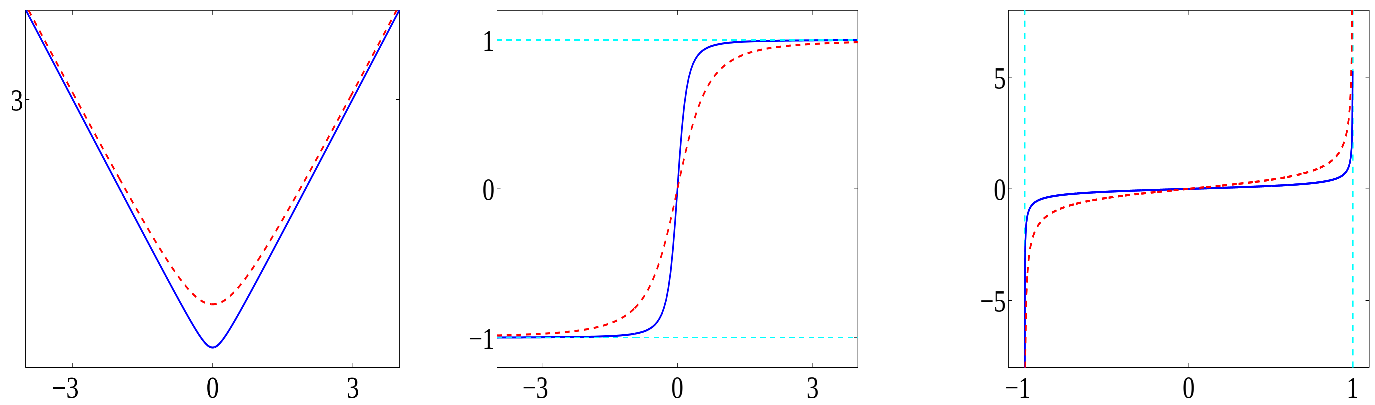
v quantized on $\{0, \cdots, 15\}$



Restored $\widehat{u}$

	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.



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

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

[MN, Wen, R. Chan 12]

Proposition 9.1 Let $\mathcal{F}_v$ satisfy H1. Then $\forall\ \beta$, $\mathcal{F}_v(\mathbb{R}^p)$ has a unique minimizer function $\mathcal{U}:\mathbb{R}^p \to \mathbb{R}^p$ which is $\mathcal{C}^{m-1}$ and $D\mathcal{U}(v) \in \mathbb{R}^{p\times p}$ satisfies $\mathrm{rank} D\mathcal{U}(v) = p \quad \forall\ v \in \mathbb{R}^p$

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

Any 1st-order difference operator G_i belongs 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 33 How to interpret the sets $N_{\mathcal{G}}$ and N_I?

Details about $N_{\mathcal{G}}$

- $f_g(v) := g\,\mathcal{U}(v)$ then $f_g \sim \mathcal{C}^{m-1}$;
- $D\mathcal{U}(v)$ invertible, $\mathrm{rank} f_g(v) = 1$ and f_g does not have critical points;
- $N_g := f_g^{-1}(0) = \{v \in \mathbb{R}^p\ :\ g\,\mathcal{U}(v) = 0\}$ (by extension of the Constant Rank Theorem)
- N_g – manifold with $\dim N_g = p-1$, closed because $f_g \sim \mathcal{C}^{m-1}$ hence $\mathbb{L}(N_g) = 0$

Theorem 9.1 Let $\mathcal{F}_v$ satisfy H1. Then 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 $\mathbb{R}^p$

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

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

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

Question **34** Describe the consequences if $\ell_1 - \mathrm{TV}$ is approximated by a smooth function like $\mathcal{F}_v$.

Bounds on the minimizer

[Bauss, MN, Steidl 13]

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

  Example how to find $(\psi')^{-1}$

  Let $\psi(t) = |t| - \alpha \log\left(1 + \dfrac{|t|}{\alpha}\right)$

  $$y := \psi'(t) = \mathrm{sign}(t) - \frac{\alpha}{\alpha + |t|}\mathrm{sign}(t) = \frac{t}{\alpha + |t|}$$

  $\mathrm{sign}(y) = \mathrm{sign}(t)$

  $$y\alpha + y|t| = t = y\alpha + y\,t\,\mathrm{sign}(y) \quad \Rightarrow \quad t(1-|y|) = \alpha y \quad \Rightarrow \quad t = \frac{\alpha y}{1-|y|} \equiv (\psi')^{-1}(y)$$

  Question **35** Compute $(\theta')^{-1}$ for all functions on p. 96.

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

Theorem 9.2 Let H1 and H2 hold. Assume that $\beta < \dfrac{1}{\|G\|_1}$. Then

$$\|\widehat{u} - v\|_\infty \leqslant \left(\psi'_{\alpha_1}\right)^{-1}\left(\beta\|G\|_1\right) \quad \forall\, v \in \mathbb{R}^p$$

Furthermore, $\quad \|\widehat{u} - v\|_\infty \nearrow \left(\psi'_{\alpha_1}\right)^{-1}\left(\beta\|G\|_1\right)$ as $\alpha_2 \searrow 0$.

Sketch of the proof

From Fermat's rule $\widehat{u}$ satisfies $\nabla_u \Psi(\widehat{u}, v) = -\beta \nabla_u \Phi(\widehat{u})$. Componentwise, using that $|\varphi'_{\alpha_2}| \leqslant 1$:

$$\psi'_{\alpha_1}(\widehat{u}[i] - v[i]) = -\beta\Big(G^{\mathsf{T}}\varphi'_{\alpha_2}(G\widehat{u})\Big)[i] \quad \forall\, i$$

$$|\,\widehat{u}[i] - v[i]\,| = \left|\left(\psi'_{\alpha_1}\right)^{-1}\Big(\beta\big(G^{\mathsf{T}}\varphi'_{\alpha_2}(G\widehat{u})\big)[i]\Big)\right| \leqslant \left(\psi'_{\alpha_1}\right)^{-1}(\beta\|G\|_1) \quad \forall\, i$$

- The upper bound depends only on ψ_{α_1} and β.
- $\|G\|_1 = 4$ for 1st-order horizontal and vertical differences between adjacent pixels.
- The value $\|\widehat{u} - v\|_\infty - \left(\psi'_{\alpha_1}\right)^{-1}\left(\beta\|G\|_1\right)$ depends on v and on α_2 and can be computed.
- $\|\widehat{u} - v\|_\infty \leqslant \delta$ for any $\alpha_1 \in (0, \widehat{\alpha}_1]$ and there does not exist $\alpha_1 > \widehat{\alpha}_1$ such that $\|\widehat{u} - v\|_\infty \leqslant \delta$ holds true.

Examples

$$\eta := \|G\|_1 \qquad \text{and} \qquad b(\beta, \alpha_1) := \left(\psi'_{\alpha_1}\right)^{-1}(\beta\eta)$$

$$\psi(t) = \sqrt{t^2 + \alpha_1} \qquad b(\beta, \alpha_1) = \sqrt{\frac{\alpha_1 (\beta\eta)^2}{1 - (\beta\eta)^2}} \qquad \widehat{\alpha}_1 = \delta^2 \left(\frac{1}{(\beta\eta)^2} - 1 \right)$$

$$\psi(t) = |t| - \alpha_1 \log\left(1 + \frac{|t|}{\alpha_1}\right) \qquad b(\beta, \alpha_1) = \frac{\alpha_1 \beta\eta}{1 - \beta\eta} \qquad \widehat{\alpha}_1 = \delta \left(\frac{1}{\beta\eta} - 1 \right)$$

Full control on the minimizer with respect to the parameters.

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: $H_v[k] = \frac{1}{p} \sharp \left\{ v[i] = k \;:\; i \in \{1, \cdots, p\} \right\} \quad \forall\, k \in \{0, \cdots, L-1\}$
- Target histogram: $\zeta = (\zeta[1], \cdots, \zeta[L])$
- Goal of histogram specification (HS): convert v into $\widehat{u}$ so that $H_{\widehat{u}} = \zeta$

  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 $\zeta[k] = p/L \quad \forall\, k \in \{0, \cdots L-1\}$

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.

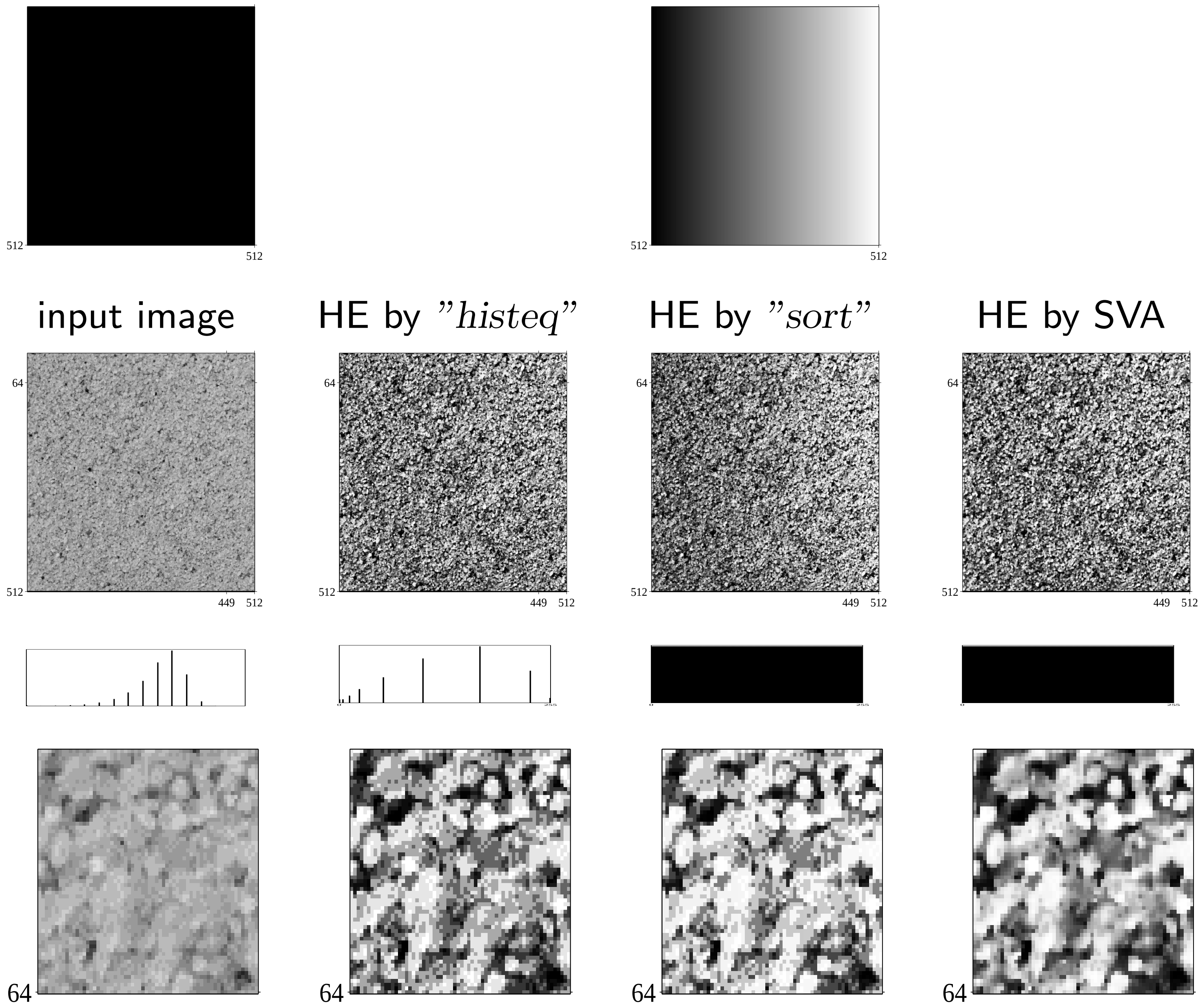
– Heavy and high level of failure.

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

– Minimize $\mathcal{F}_v$ for a parameter choice yielding $\|\widehat{u} - v\|_\infty \lessapprox 0.1$. [50]

– Faithful order and fast algorithm. [54]

Histogram Equalization (HE) using Matlab and SVA ordering



Fringe removal [Soncco, MN 16]

Multiplicative Image Decomposition for Hyperspectral Imaging

$$v = u \circ (1 + f) + n$$

- u panchromatic (fringe-less) image
- f image containing the interferometric pattern, $-1 \leqslant v \leqslant 1$
- n noise (small)

Fast solver based on fully smoothed L1-TV with constraint on FT(f)

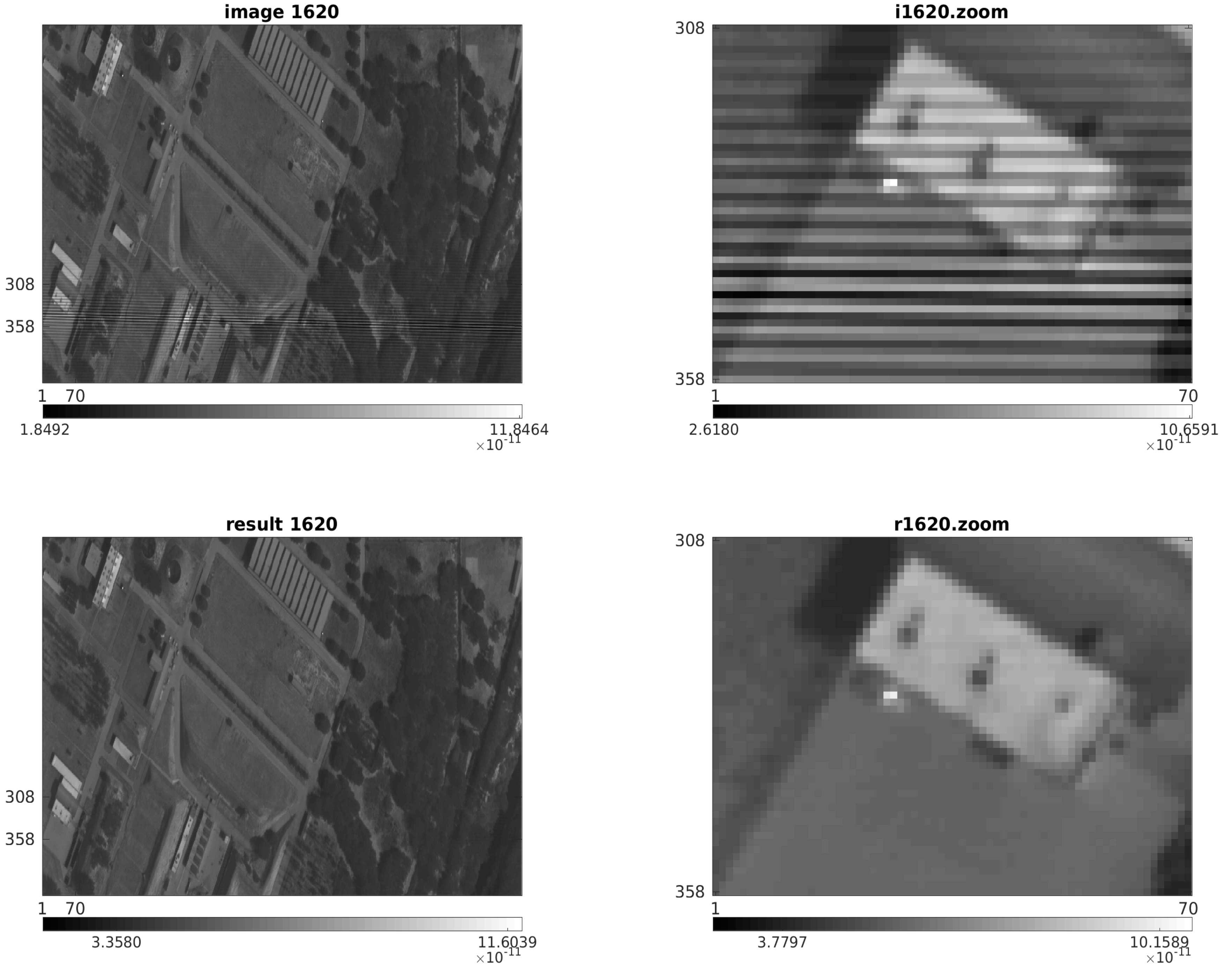


image 1620
308
358
1
70
1.8492
11.8464
×10-11
i1620.zoom
308
358
1
70
2.6180
10.6591
×10-11
result 1620
308
358
1
70
3.3580
11.6039
×10-11
r1620.zoom
308
358
1
70
3.7797
10.1589
×10-11

10 Combining models

Bayesian estimators

- $U,\ V$ random variables
- Likelihood $f_{V|U}(v|u)$
- Prior $f_U(u) = C\exp\{-\lambda\Phi(u)\}$
- Loss function $L(u, u')$ – measures the cost of estimating u' instead of u

Bayes estimation: minimize the risk $\mathbb{E}_{u|v}\left(L(u, u')\right)$

$$\arg\min_{u'} \mathbb{E}_{u|v}\left(L(u, u')\right) \qquad \text{using the posterior } f_{U|V}(u|v)$$

$$L(u, u') = \|u - u'\|^2 \quad \Longrightarrow \quad \widehat{u}_{\mathrm{PM}} = \mathbb{E}(u|v) = \int u f_{U|V}(u|v)du \ \text{ posterior mean (PM)}$$

$$L(u, u') = \mathbb{1}_{u=u'} \quad \Longrightarrow \quad \widehat{u}_{\mathrm{MAP}} = \arg\max_{u} f_{U|V}(u|v) \ \text{ maximum a posteriori (MAP)}$$

Other loss-functions can be better according to the problem...

Well known fact: $f_{V|U}(v|u)$ and $f_U(u)$ have normal distributions $\quad \Longrightarrow \quad \widehat{u}_{\mathrm{PM}} = \widehat{u}_{\mathrm{MAP}}$

Question **36** Prove this fact

MAP estimators to combine data-production and prior models

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

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

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

MAP gives a direct connection to variational regularization objectives

$\Longrightarrow$ The objectives considered so far are usually interpreted as MAP estimators

There exist realist models for data-acquisition $\boldsymbol{f_{V|U}}$ and for priors $\boldsymbol{f_U}$

If a MAP solution $\widehat{u}$ had to be faithful (coherent), then [MN 07]

– The main features of $\widehat{u}$ should match the prior $C \exp\big(-\Phi(u) \big)$;

– The empirical distribution of the recovered noise should fit the data-production model.

Analytical facts on the minimizers $\quad \Longrightarrow \quad$ both models ($f_{V|U}$ and f_U) are violated

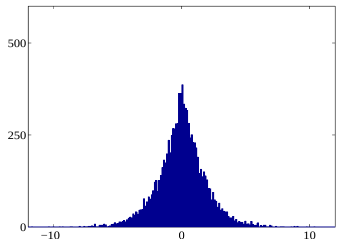
Example: MAP shrinkage in a frame domain

- Noisy wavelet coefficients $\boldsymbol{y} = \boldsymbol{W}\boldsymbol{v} = \boldsymbol{W}\boldsymbol{u_o} + \boldsymbol{n} = \boldsymbol{x_o} + \boldsymbol{n}$, $n \sim \mathcal{N}(0, \sigma^2 I)$
- 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)

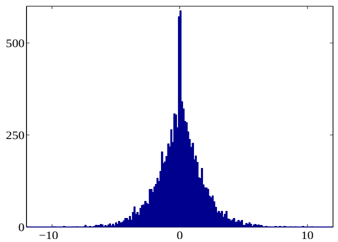
  Experiments have shown that $\boldsymbol{\alpha \in (0, 1)}$ for many real-world images [56, 57, 58]
- MAP restoration $\iff$ $\boldsymbol{\hat{x}[i] = \arg\min_{t \in \mathbb{R}} \left((t - y[i])^2 + \lambda |t|^{\alpha} \right), \ \forall i}$

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

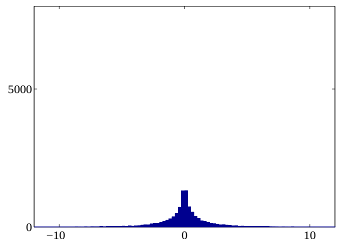
(1) sample $x \sim f_X$ and $n \sim \mathcal{N}(0, \sigma^2)$, (2) form $y = x + n$, (3) compute the **true** MAP $\hat{x}$



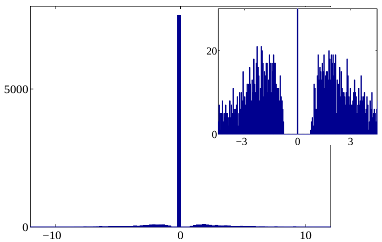
GG, $\alpha = 1.2, \lambda = \frac{1}{2}$



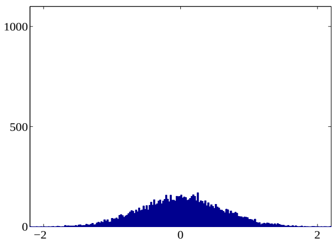
The true MAP $\hat{x}$



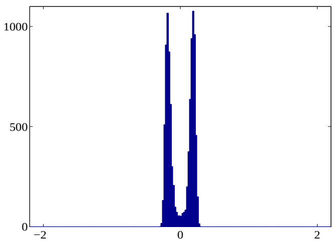
GG, $\alpha = \frac{1}{2}$, $\lambda = 2$



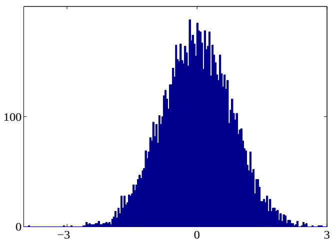
True MAP $\hat{x}$



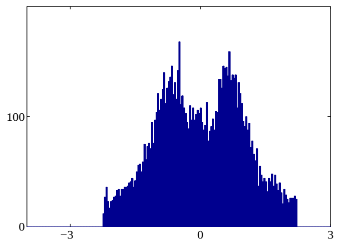
Noise $\mathcal{N}(0, \sigma^2)$



Recovered noise



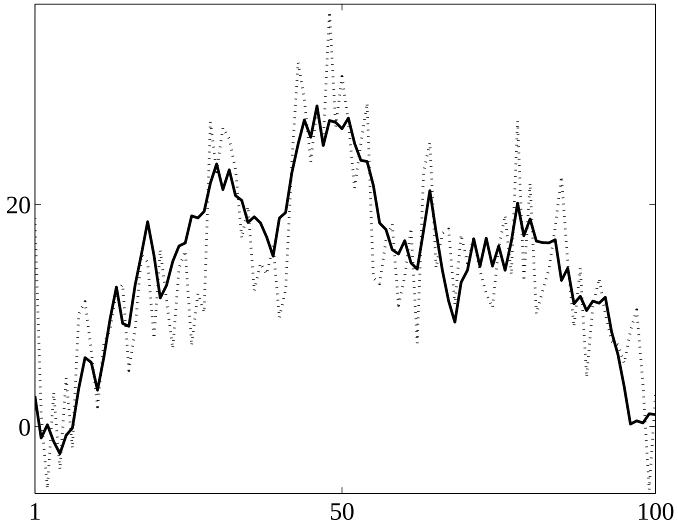
Noise $\mathcal{N}(0, \sigma^2)$



Recovered noise

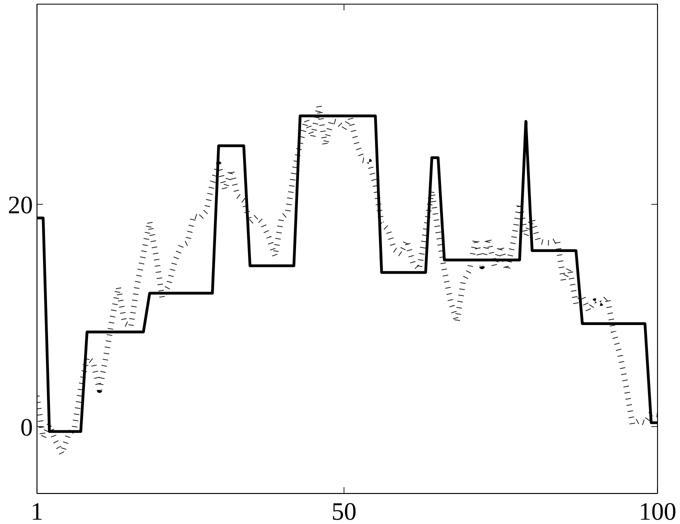
Example: MAP signal recovery with known distributions and parameters

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



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 $(\cdots)$.

Instead: focus on the effective models [55]

Effective model: the properties that the minimizers $\widehat{u}$ of the objective $\mathcal{F}_v$ satisfy

$-\log f_U$ continuous and non-smooth, $\varphi'(0^+)>0$ Ch. 4, p. 27

$$\begin{array}{rcl} \mathbb{P}(G_i u=0) & = & 0, \quad \forall i \\ v\in\mathcal{O}_{\hat h} \Rightarrow \left[G_i\hat u=0,\forall i\in\hat h\right] \Rightarrow \mathbb{P}(G_i\hat u=0,\forall i\in\hat h) & \geqslant & \mathbb{P}(v\in\mathcal{O}_{\hat h})>0 \end{array}$$

Effective prior: $G_i\hat u=0$ for some (many) i. (If $\{G_i\}=\nabla$ – locally constant images)

$-\log f_{U|V}$ continuous and nonsmooth, $\psi'(0^+)>0$ Ch. 6, p. 52

$$\begin{array}{rcl} \mathbb{P}(a_i\, u=v_i) & = & 0 \quad \forall i \\ v\in\mathcal{O}_{\hat h} \Rightarrow \left[a_i\,\hat u=v_i,\ \forall i\in\hat h\right] \Rightarrow \mathbb{P}\Big(a_i\,\hat u=v_i,\forall i\in\hat h\Big) & \geqslant & \mathbb{P}(V\in\mathcal{O}_{\hat h})>0 \end{array}$$

Effective model: some data entries are fitted exactly.

$-\log f_U$ (resp., φ) continuous and nonconvex Ch. 5, p. 37

$$\begin{array}{rcl} \mathbb{P}(\theta_0<\|G_i u\|<\theta_1) & > & 0, \quad \forall i \\ \mathbb{P}\Big(\theta_0<\|G_i\widehat{u}\|<\theta_1\Big) & = & 0, \ \forall i \end{array}$$

Effective prior: $\|G_i u\|\geqslant\theta_1-\theta_0$. (If $\{G_i\}=\nabla$ – high edges).

$-\log f_U$ nonconvex, nonsmooth, continuous, $\varphi'(0^+)>0$ and $\varphi''\leqslant 0$ Ch. 5, p. 39

$$\begin{array}{rcl} \mathbb{P}(0<\|G_i u\|<\theta_1) & > & 0, \quad \forall i \\ \mathbb{P}\Big(0<\|G_i\hat U\|<\theta_1\Big) & = & 0, \ \forall i \end{array}$$

Effective prior: $\|G_i u\|\geqslant\theta_1$. (If $\{G_i\}=\nabla$ – constant regions separated by edges $>\theta_1$).

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

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

MAP gives a direct connection to variational regularization objectives

"Theoretical drawback": MAP takes the maximum, "forgets" the rest of the posterior

– PM seems statistically more sound but higher numerical complexity

The relevant loss function has a clear meaning:
PM is unbiased with respect to $f_{U|V}(u|v)$

posterior mean (PM) $\equiv$ conditional mean (CM) $\equiv$ minimum mean-square error (MMSE)

Normal noise: MAP and PM can be equal but for different priors [Gribonval 11]

Theorem [Gribonval 11] Let $V = U + N$ where $N \sim \mathcal{N}(0, I)$ and U be independent. Then:

- For any prior $p_U(u)$, the estimator $\widehat{u}_{\text{PM}}$ with prior $p_U(u)$ equals $\widehat{u}_{\text{MAP}}$ where MAP correspond to a prior $f_U(u) = C \exp\big(-\Phi(u)\big)$
- vice-versa, for certain regularizers Φ the relevant $\widehat{u}_{\text{MAP}}$ equals $\widehat{u}_{\text{PM}}$ for a different prior $p_U(u)$
- In general $p_U(u) \neq C \exp\big(-\Phi(u)\big)$

In regularized least squares, one must be cautious when interpreting the regularizer in terms of prior in a statistical sense

A detailed study of the PM in the case of TV regularizer in [Louchet, Moisan 13]
In particular, there is no stair-casing – a major concern with TV for 20 years

What is the difference is between MAP and PM estimates? [Burger, Lucke 14]

"PM estimate is classically preferred for being the Bayes estimator for the mean squared error cost, while the MAP estimate is classically discredited for being only asymptotically the Bayes estimator for the uniform cost function." [61]

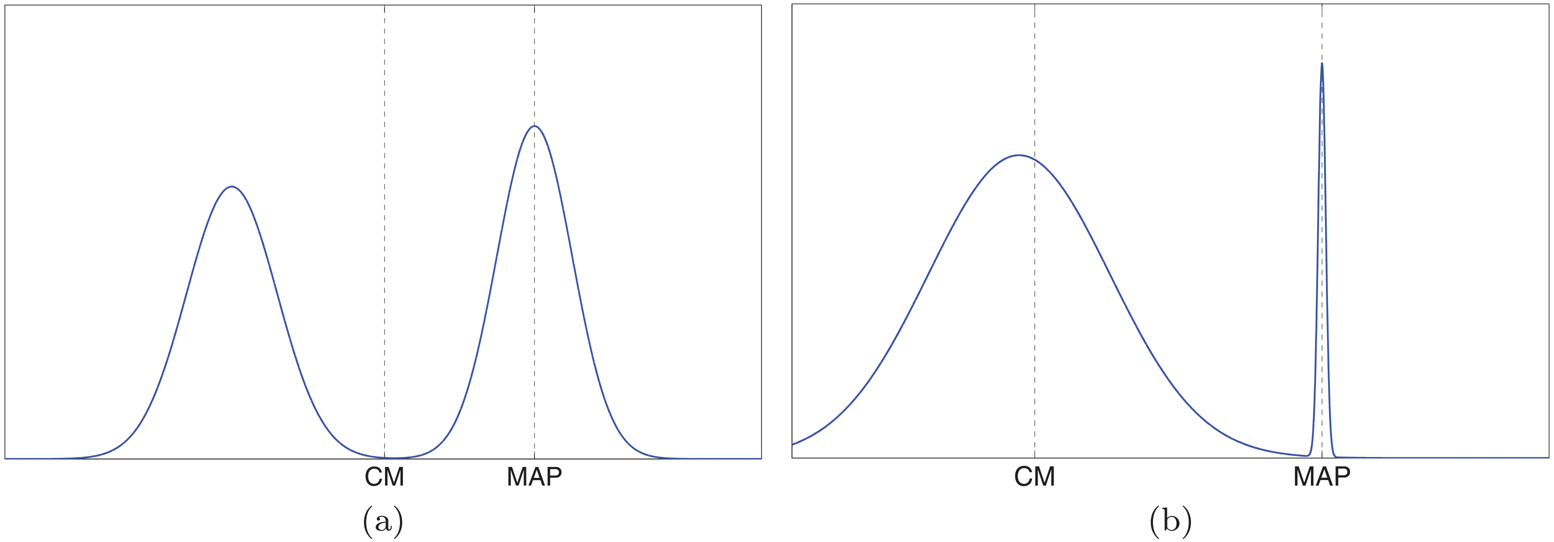
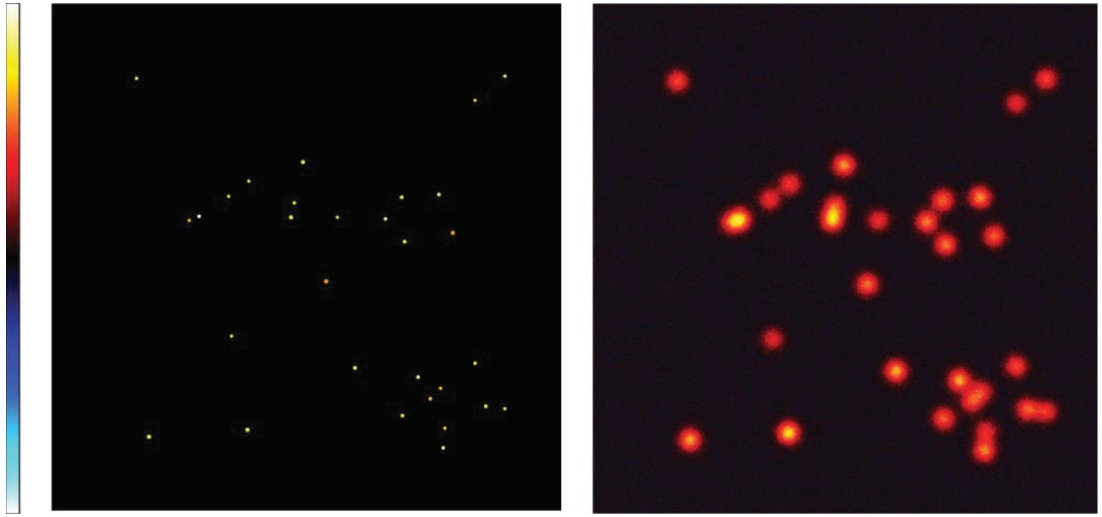


Figure 4. Hypothetical, bimodal distributions to show that neither of the estimates is better in general.

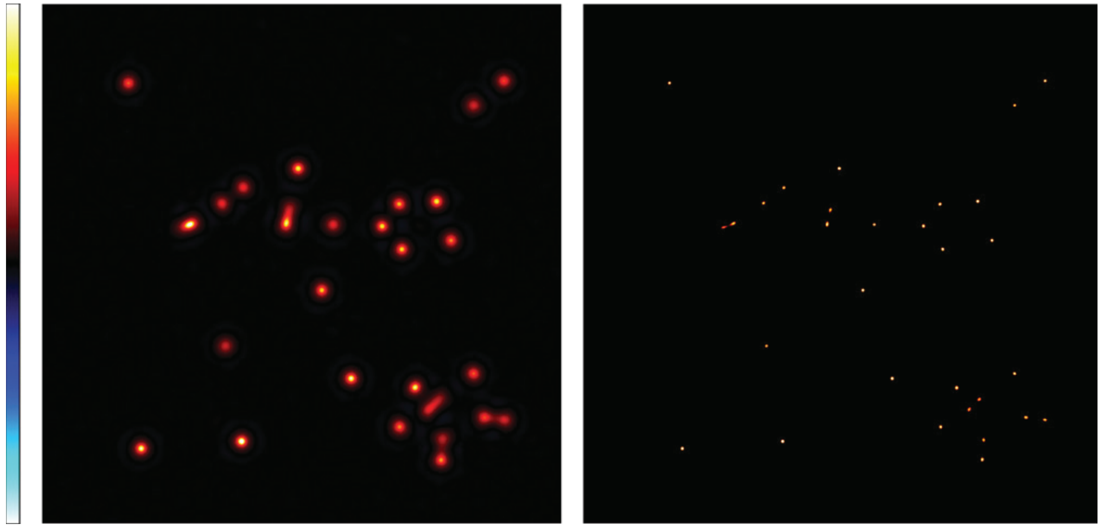
Image credits to the authors Burger and Lucke "Maximum a posteriori estimates in linear inverse problems with log-concave priors are proper Bayes estimators", Inverse Problems, 2014

"Which of them is "better" in general, or for a specific task? - a matter of constant debate"



(a) (b) Unknown function $\tilde{u}$ (c) Data f

Figure 5. A simple 2D deblurring example.



(a) (b) CM estimate (c) MAP estimate

Figure 6. CM and MAP estimate for the 2D deblurring example.

Image credits to the authors Burger and Lucke [61]

Rehabilitation of the MAP for linear problems with sparsity-promoting convex priors [61]

Φ – sparsity promoting and convex – constructed using ℓ_1 norms

Definition 10.1 Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be convex.
The Bregman distance between $u, w \in \mathbb{R}^n$ is

$$D_f^g(u,w) := f(u) - f(w) - \langle g, u - w \rangle \quad g \in \partial f(w)$$

where $\partial f(w)$ belongs to the subdifferential of f at w.

Using Bregman distance, $f_{U|V}(u|v)$ can be rewritten in a MAP-centered form.

[61, Theorem 2] $\quad \mathbb{E}\left[D_\Phi(\widehat{u}_{\mathrm{MAP}}, u)\right] \leqslant \mathbb{E}\left[D_\Phi(\widehat{u}_{\mathrm{PM}}, u)\right]$

– Bregman distance is better suited than L2 norm when Φ is not quadratic

– With the Bregman distance, MAP outperforms PM in terms of theoretical statistics for sparse images

11. Concluding remarks

Combining models remains an open problem

How to solve?

- Non-local multiscale data-adaptive models [62]
- Strong priors based on dictionaries, splines, etc... [63]
- Posterior-sampling based methods [64]
- Construction of specialized $\mathcal{F}_v$ whose minimizers fulfill the requirements

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

12 Some References

References

[1] A. Tikhonov and V. Arsenin, *Solutions of Ill-Posed Problems*, Winston, Washington DC, 1977.

[2] R. B. Potts, "Some generalized order-disorder transformations", *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 48, no. 1, pp. 106–109, 1952.

[3] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images", *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-6, no. 6, pp. 721–741, Nov. 1984.

[4] D. Mumford and J. Shah, "Boundary detection by minimizing functionals", in *Proceedings of the IEEE Int. Conf. on Computer Vision and Pattern Recognition*, 1985, pp. 22–26.

[5] L. Rudin, S. Osher, and C. Fatemi, "Nonlinear total variation based noise removal algorithm", *Physica*, vol. 60 D, pp. 259–268, 1992.

[6] P. Charbonnier, L. Blanc-Féraud, G. Aubert, and M. Barlaud, "Deterministic edge-preserving regularization in computed imaging", *IEEE Trans. Image Process.*, vol. 6, no. 2, pp. 298–311, Feb. 1997.

[7] K. Bredies, K. Kunich, and T. Pock, "Total generalized variation", *SIAM J. Imaging Sci.*, vol. 3, no. 3, pp. 492–526, 2010.

[8] M. Nikolova, "Minimizers of cost-functions involving nonsmooth data-fidelity terms. Application to the processing of outliers", *SIAM J. Numer. Anal.*, vol. 40, no. 3, pp. 965–994, 2002.

[9] T. Chan and S. Esedoglu, "Aspects of total variation regularized l^1 function approximation", *SIAM J. Appl. Math.*, vol. 65, pp. 1817–1837, 2005.

[10] A. Auslender and M. Teboulle, *Asymptotic Cones and Functions in Optimization and Variational Inequalities*, Springer, New York, 2003.

[11] M. C. Robini and I. E. Magnin, "Optimization by stochastic continuation", *SIAM J. Imaging Sci.*, vol. 3, no. 4, pp. 1096–1121, 2010.

[12] M. C. Robini and P.-J. Reissman, "From simulated annealing to stochastic continuation: a new trend in combinatorial optimization", *J. Global Optim.*, vol. 56, no. 1, pp. 185–215, May 2013.

[13] J. Lellmann and C. Schnörr, "Continuous multiclass labeling approaches and algorithms", *SIAM J. Imaging Sci.*, vol. 4, no. 4, pp. 1049–1096, 2011.

[14] E. Bae, J. Yuan, and X.-C. Tai, "Global minimization for continuous multiphase partitioning problems using a dual approach", *Int J Comput Vis*, vol. 92, 2011.

[15] A. Chambolle, D. Cremers, and T. Pock, "A convex approach to minimal partitions", *SIAM J. Imaging Sci.*, vol. 5, no. 4, pp. 1113–1158, 2012.

[16] F. Fogel, R. Jenatton, F. Bach, and A. d'Aspremont, "Convex relaxations for permutation problems", *SIAM. J. Matrix Anal Appl*, vol. 36, no. 4, pp. 1465–1488, 2015.

[17] H. Attouch, J. Bolte, and B. F. Svaiter, “Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forwardbackward splitting, and regularized gaussseidel methods”, *Math. Program.*, vol. 137, no. 1, Feb. 2013.

[18] Z. Peng, T. Wu, Y. Xu, M. Yan, and W. Yin, “Coordinate friendly structures, algorithms and applications”, *Annals of Mathematical Sciences and Applications*, vol. 1, no. 1, pp. 57–119, 2016.

[19] J. Nocedal and S. Wright, *Numerical Optimization*, Springer, New York, 2 edition, 2006.

[20] L. C. Evans and R. F. Gariepy, *Measure theory and fine properties of functions*, Studies in Advanced Mathematics. CRC Press, Roca Baton, FL, 1992.

[21] J.-B. Hiriart-Urruty and C. Lemaréchal, *Convex analysis and Minimization Algorithms, vol. I and II*, Springer-Verlag, Berlin, 1996.

[22] A. Fiacco and G. McCormic, *Nonlinear programming*, Classics in Applied Mathematics. SIAM, Philadelphia, 1990.

[23] S. Durand and M. Nikolova, “Stability of minimizers of regularized least squares objective functions I: study of the local behaviour”, *Appl. Math. Optim.*, vol. 53, no. 2, pp. 185–208, Mar. 2006.

[24] S. Durand and M. Nikolova, “Stability of minimizers of regularized least squares objective functions II: study of the global behaviour”, *Appl. Math. Optim.*, vol. 53, no. 3, pp. 259–277, May 2006.

[25] M. Nikolova, “Local strong homogeneity of a regularized estimator”, *SIAM J. Appl. Math.*, vol. 61, no. 2, pp. 633–658, 2000.

[26] M. Nikolova, "Weakly constrained minimization.Application to the estimation of images and signals involving constant regions", *J. Math. Imaging Vis.*, vol. 21, no. 2, pp. 155–175, Sep. 2004.

[27] T. Pock, D. Cremers, H. Bischof, and A. Chambolle, "Global solutions of variational models with convex regularization", *SIAM J. Imaging Sci.*, vol. 3, no. 2, pp. 1122–1145, 2010.

[28] S. C. Zhu and D. Mumford, "Prior learning and gibbs reaction–diffusion", *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 11, pp. 1236–1250, Nov. 1997.

[29] M. Nikolova, "Analysis of the recovery of edges in images and signals by minimizing nonconvex regularized least-squares", *SIAM J. Multiscale Model Simul*, vol. 4, no. 3, pp. 960–991, 2005.

[30] M. Nikolova, M. Ng, and C. P. Tam, "A fast nonconvex nonsmooth minimization method for image restoration and reconstruction", *IEEE Trans. Image Process.*, vol. 19, no. 12, pp. 3073–3088, Dec. 2010.

[31] X. Chen, M. K. Ng, and C. Zhang, "Non-lipschitz ℓ_p-regularization and box constrained model for image restauration", *IEEE Trans. Image Process.*, vol. 21, no. 12, pp. 4709–4721, Dec. 2012.

[32] M. Hintermüller and T. Wu, "Nonconvex TV^q-models in image restoration: Analysis and a trust-region regularization–based superlinearly convergent solver", *SIAM J. Imaging Sci.*, 2013.

[33] M. Nikolova, "One-iteration dejittering of digital video images", *J. Vis. Commun. Image Rep.*, vol. 20, no. 4, pp. 254–274, 2009.

[34] A. Kokaram, *Motion picture restoration*, Springer-Verlag, 1998.

[35] J. Shen, “Bayesian video dejittering by bv image model”, *SIAM J. Appl. Math.*, vol. 64, no. 5, pp. 1691–1708, 2004.

[36] S. H. Kang and J. Shen, “Video dejittering by bake and shake”, *Image and vision computing*, vol. 24, no. 2, pp. 143–152, 2006.

[37] F. Lenzen and O. Scherzer, “Partial differential equations for zooming, deinterlacing and dejittering”, *Int J Comput Vis*, vol. 92, no. 2, pp. 162–176, Apr. 2011.

[38] M. Nikolova, “Description of the minimizers of least squares regularized with ℓ_0-norm. uniqueness of the global minimizer”, *SIAM J. Imaging Sci.*, vol. 6, no. 2, pp. 904–937, 2013.

[39] M. Nikolova, “Relationship between the optimal solutions of least squares regularized with ℓ_0-norm and constrained by k-sparsity”, *Appl. Comput. Harmon. Anal.*, vol. 41, no. 1, pp. 237–265, July 2016.

[40] E. Soubies, L. Blanc-Féraud, and G. Aubert, “A continuous exact ℓ_0 penalty (CEL0) for least squares regularized problem”, *SIAM J. Imaging Sci.*, vol. 8, no. 3, pp. 1607–1639, 2015.

[41] L. Bar, A. Brook, N. Sochen, and N. Kiryati, “Deblurring of color images corrupted by impulsive noise”, *IEEE Trans. Image Process.*, vol. 16, no. 4, pp. 1101–1111, Apr. 2007.

[42] M. Nikolova, “Analytical bounds on the minimizers of (nonconvex) regularized least-squares”, *Inverse Probl Imaging*, vol. 1, no. 4, pp. 661–677, 2007.

[43] S. Durand and M. Nikolova, "Denoising of frame coefficients using ℓ_1 data-fidelity term and edge-preserving regularization", *SIAM J. Multiscale Model Simul*, vol. 6, no. 2, pp. 547–576, 2007.

[44] T. Chan, S. Esedoglu, and M. Nikolova, "Algorithms for finding global minimizers of image segmentation and denoising models", *SIAM J. Appl. Math.*, vol. 66, no. 5, pp. 1632–1648, 2006.

[45] S. Durand, J. Fadili, and M. Nikolova, "Multiplicative noise removal using ℓ_1 fidelity on frame coefficients", *J. Math. Imaging Vis.*, vol. 19, no. 12, pp. 201–226, Dec. 2010.

[46] C. Chesneau, J. Fadili, and J.-L. Starck, "Stein block thresholding for image denoising", *Appl. Comput. Harmon. Anal.*, vol. 28, no. 1, pp. 67–88, Jan. 2010.

[47] G. Aubert and J.-F. Aujol, "A variational approach to remove multiplicative noise", *SIAM J. Appl. Math.*, vol. 68, no. 4, pp. 925–946, Jan. 2008.

[48] J. Shi and S. Osher, "A nonlinear inverse scale space method for a convex multiplicative noise model", *SIAM J. Imaging Sciences*, vol. 1, no. 3, pp. 294–321, 2008.

[49] M. Nikolova, M. Ng, and C. P. Tam, "On ℓ_1 data fitting and concave regularization for image recovery", *SIAM J. Sci. Comput.*, vol. 35, no. 1, pp. 397–430, 2013.

[50] M. Nikolova, Y.-W. Wen, and R. Chan, "Exact histogram specification for digital images using a variational approach", *J. Math. Imaging Vis.*, vol. 46, no. 3, pp. 309–325, July 2013.

[51] F. Bauss, M. Nikolova, and G. Steidl, "Fully smoothed ℓ_1-TV models: Bounds for the minimizers and parameter choice", *J. Math. Imaging Vis.*, vol. 48, no. 2, pp. 295–307, Jan. 2014.

[52] D. Coltuc, P. Bolon, and J.-M. Chassery, "Exact histogram specification", *IEEE Trans. Image Process.*, vol. 15, no. 6, pp. 1143–1152, 2006.

[53] Y. Wan and D. Shi, "Joint exact histogram specification and image enhancement through the wavelet transform", *IEEE Trans. Image Process.*, vol. 16, no. 9, pp. 2245–2250, 2007.

[54] M. Nikolova and G. Steidl, "Fast ordering algorithm for exact histogram specification", *IEEE Trans. Image Process.*, vol. 23, no. 12, pp. 5274–5283, Dec. 2014.

[55] M. Nikolova, "Model distortions in bayesian map reconstruction", *Inverse Probl Imaging*, vol. 1, no. 2, pp. 399–422, 2007.

[56] E. P. Simoncelli, *Bayesian denoising of visual images in the wavelet domain*, Lecture Notes in Statistics, Vol. 41. Springer Verlag: Berlin, 1999.

[57] M. Belge, M. Kilmer, and E. Miller, "Wavelet domain image restoration with adaptive edge-preserving regularization", *IEEE Trans. Image Process.*, vol. 9, no. 4, pp. 597–608, Apr. 2000.

[58] A. Antoniadis, D. Leporini, and J.-C. Pesquet, "Wavelet thresholding for some classes of non-gaussian noise", *Statistica Neerlandica*, vol. 56, no. 4, pp. 434–453, Dec. 2002.

[59] R. Gribonval, "Should penalized least squares regression be interpreted as maximum a posteriori estimation ?", *IEEE Trans. Signal Process.*, vol. 59, no. 5, pp. 2405–2410, May 2011.

[60] C. Louchet and L. Moisan, "Posterior expectation of the total variation model: Properties and experiments", *SIAM J. Imaging Sci.*, vol. 6, no. 4, pp. 2640–2684, 2013.

[61] M. Burger and F. Lucka, "Maximum a posteriori estimates in linear inverse problems with log-concave priors are proper bayes estimators", *Inverse Problems*, vol. 30, no. 11, pp. 1–21, 2014.

[62] M. Lebrun, A. Buades, and J. M. Morel, "A nonlocal Bayesian image denoising algorithm", *SIAM J. Imaging Sci.*, vol. 6, no. 3, pp. 1665–1688, 2013.

[63] M. Unser, J. Fageot, and J.-P. Ward, "Splines are universal solutions of linear inverse problems with generalized 66 tv regularization", *arXiv*, vol. 0, no. 0, pp. 1–16, 2016.

[64] G. Papandreou and A. Yuille, "Perturb-and-map random fields: using discrete optimization to learn and sample from energy models", *ICCV – IEEE*, vol. 1, no. 1, pp. 193–200, 2011.