ℓ_1−concave versus ℓ_1 − TV energies: Questions and challenges

Mila Nikolova

CMLA, ENS Cachan, CNRS
61 Av. President Wilson, F-94230 Cachan, France

Convex Relaxation Methods for Geometric Problems in Scientific Computing

February 11 - 15, 2013, IPAM - UCLA

1. Problem formulation

image $\boldsymbol{u}$ is stored in a vector in $\mathbb{R}^p$ data $\boldsymbol{v} \in \mathbb{R}^q$

$$\begin{aligned} \mathcal{F}(\boldsymbol{u}) &= \|\boldsymbol{A}\boldsymbol{u}-\boldsymbol{v}\|_1 + \boldsymbol{\beta}\sum_{j\in J}\varphi(\|\mathbf{G}_j\boldsymbol{u}\|_2) \\ &= \sum_{i\in I}\big|\boldsymbol{a}_i\boldsymbol{u}-\boldsymbol{v}[i]\big| + \boldsymbol{\beta}\sum_{j\in J}\varphi(\|\mathbf{G}_j\boldsymbol{u}\|_2)\ , \quad \boldsymbol{\beta}>\mathbf{0}\ , \end{aligned}$$

where
$$\begin{aligned} I &\stackrel{\text{def}}{=} \{1,\cdots,q\}\ , \\ J &\stackrel{\text{def}}{=} \{1,\cdots,r\}\ . \end{aligned}$$

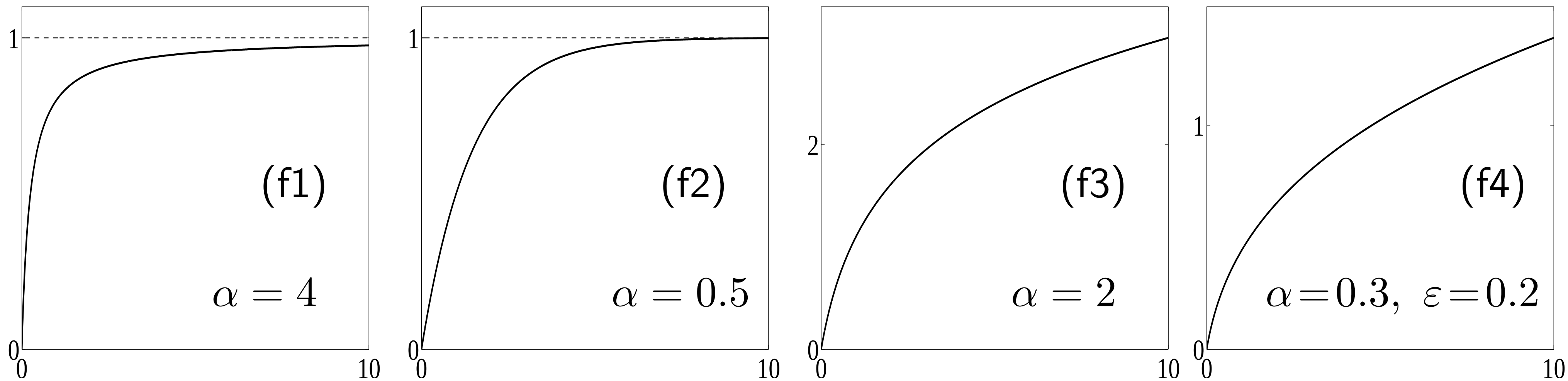
- G_j are matrices or vectors (e.g. discrete gradient operators)
- A is a matrix of any rank with rows $a_i \in \mathbb{R}^{1\times p}$
- $\varphi(t) = t \quad \Rightarrow \quad \ell_1 - \mathrm{TV}$ *[Chan, Esedoglu 2005]*
- In our case:

  φ is concave on $\mathbb{R}_+$

  This family of objective functions has never been considered before

	(f1)	(f2)	(f3)	(f4)
$\varphi(t)$	$\dfrac{\alpha\, t}{\alpha\, t+1}$	$1-\alpha^t$	$\ln(\alpha t+1)$	$(t+\varepsilon)^\alpha$
	$\alpha>0$	$0<\alpha<1$	$\alpha>0$	$0<\alpha<1,\ \varepsilon>0$

Functions $\boldsymbol{\varphi} : \mathbb{R}_+ \to \mathbb{R}_+$



Plots of the PFs φ. Note that (f1) an (f2) are bounded above, (f3) and (f4) are coercive.

Motivation

$\hat{u}$—(local) minimizer of $\mathcal{F}$

- nonsmooth regularization $\sum_{j\in J} \varphi(\|\mathrm{G}_j u\|_2)$ with $\varphi'(0) > 0$ (e.g. TV)

  $\Rightarrow$ **many j such that $\mathbf{G}_j \hat{\boldsymbol{u}} = \mathbf{0}$** *[Nikolova 2000, 2004]*

- ℓ_1 data fidelity $\|Au - v\|_1 = \sum_{i\in I} \big|a_i u - v[i]\big|$

  $\Rightarrow$ **many i such that $a_i \hat{\boldsymbol{u}} = v[i]$** *[Nikolova 2002, 2004]*

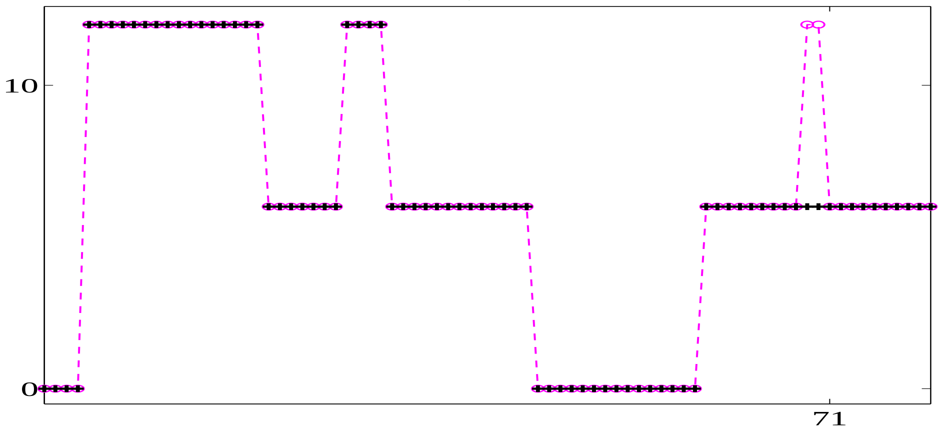
- our $\mathcal{F}$ can be seen as an extension of L1-TV

  ??? **many i, j such that $a_i \hat{\boldsymbol{u}} = v[i]$ and $\mathbf{G}_j \hat{\boldsymbol{u}} = \mathbf{0}$**

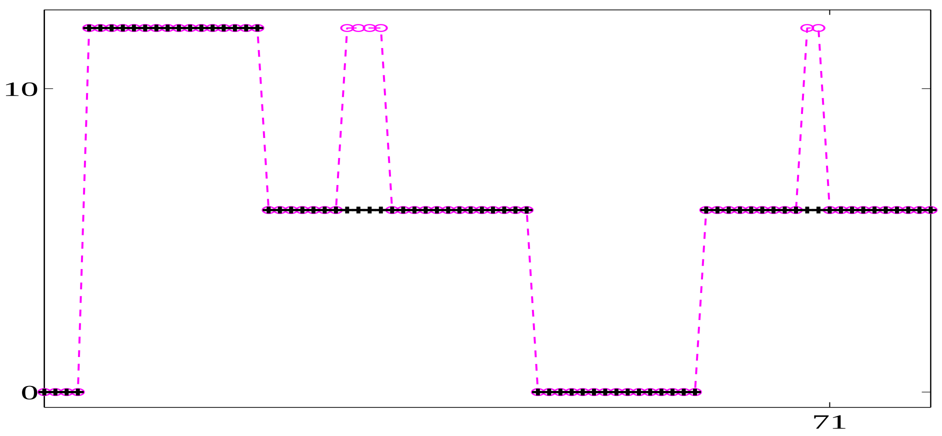
$\boxed{\textbf{2. Peculiar Properties — 1D tests}}$

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

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

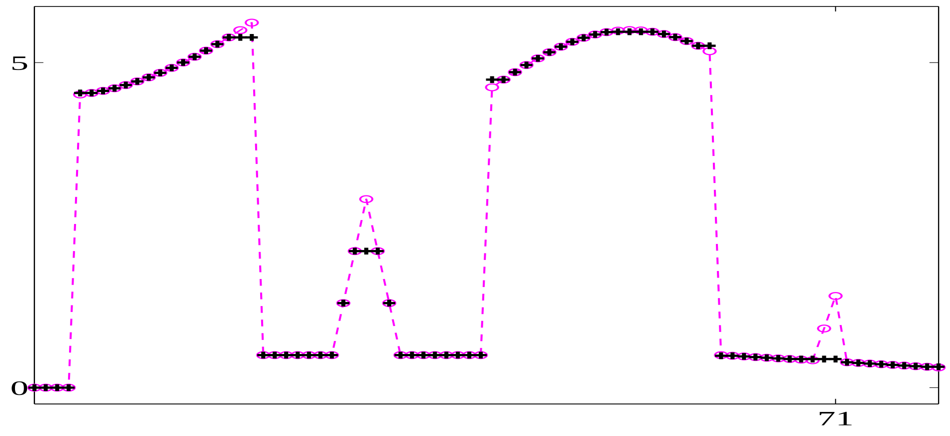


$\beta \in \{78, \cdots, 156\}$

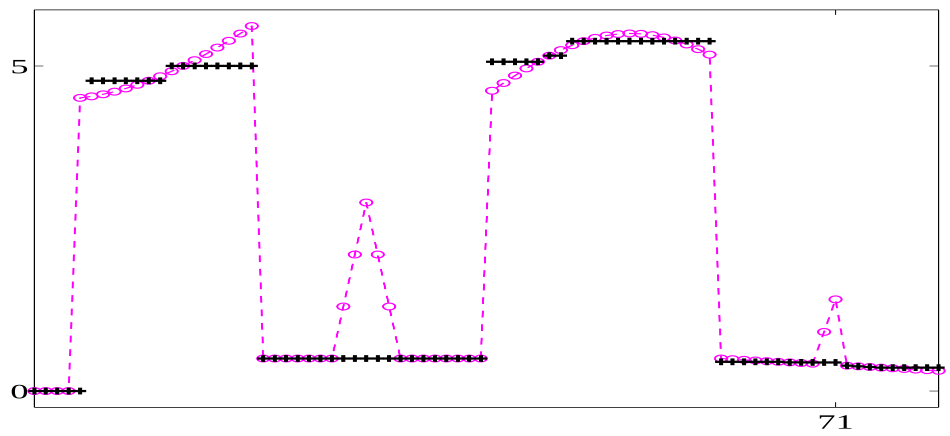


$\beta \in \{157, \cdots, 400\}$

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

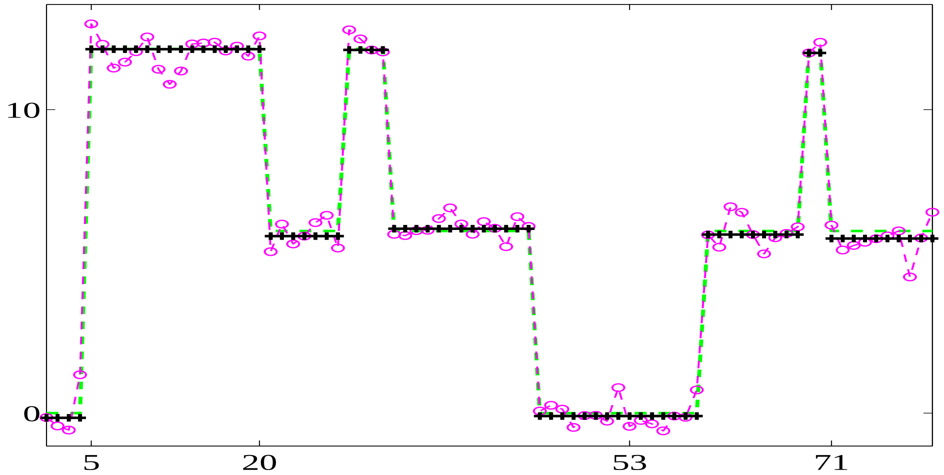


$\beta \in 0.1 \times \{10, \cdots, 14\}$

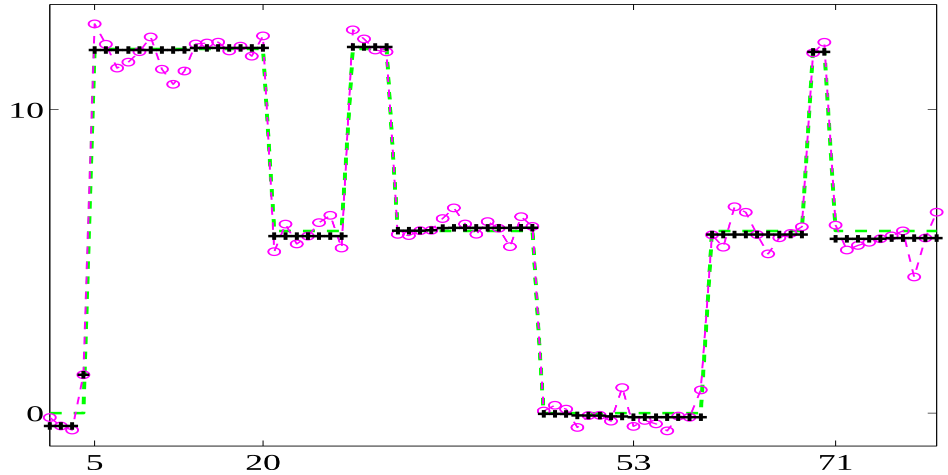


$\beta \in 0.1 \times \{16, \cdots, 30\}$

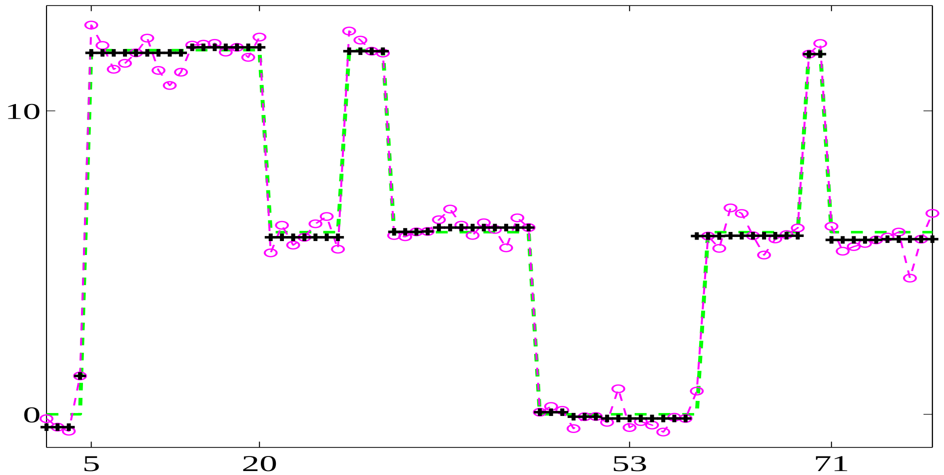
Data samples (ooo), Minimizer samples $\hat{u}[i]$ (+++).



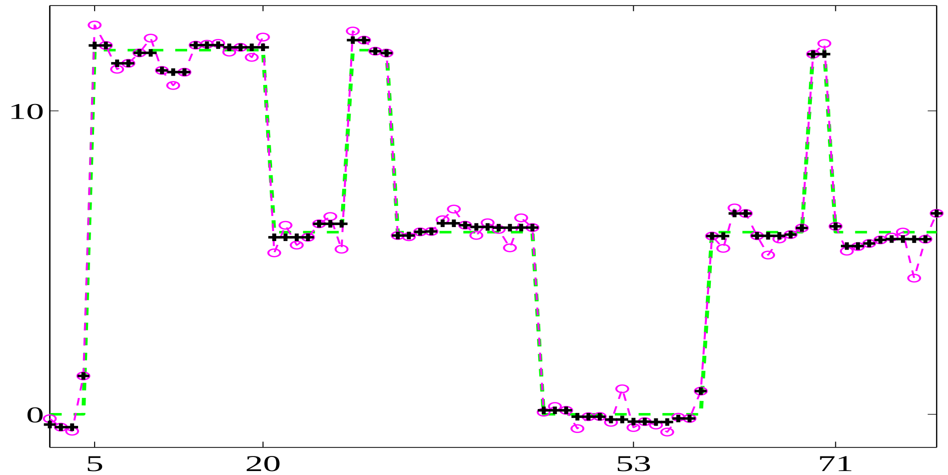
(a) $\varphi(t) = \frac{\alpha\, t}{\alpha\, t+1}$, $\alpha = 4$, $\beta = 3$



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

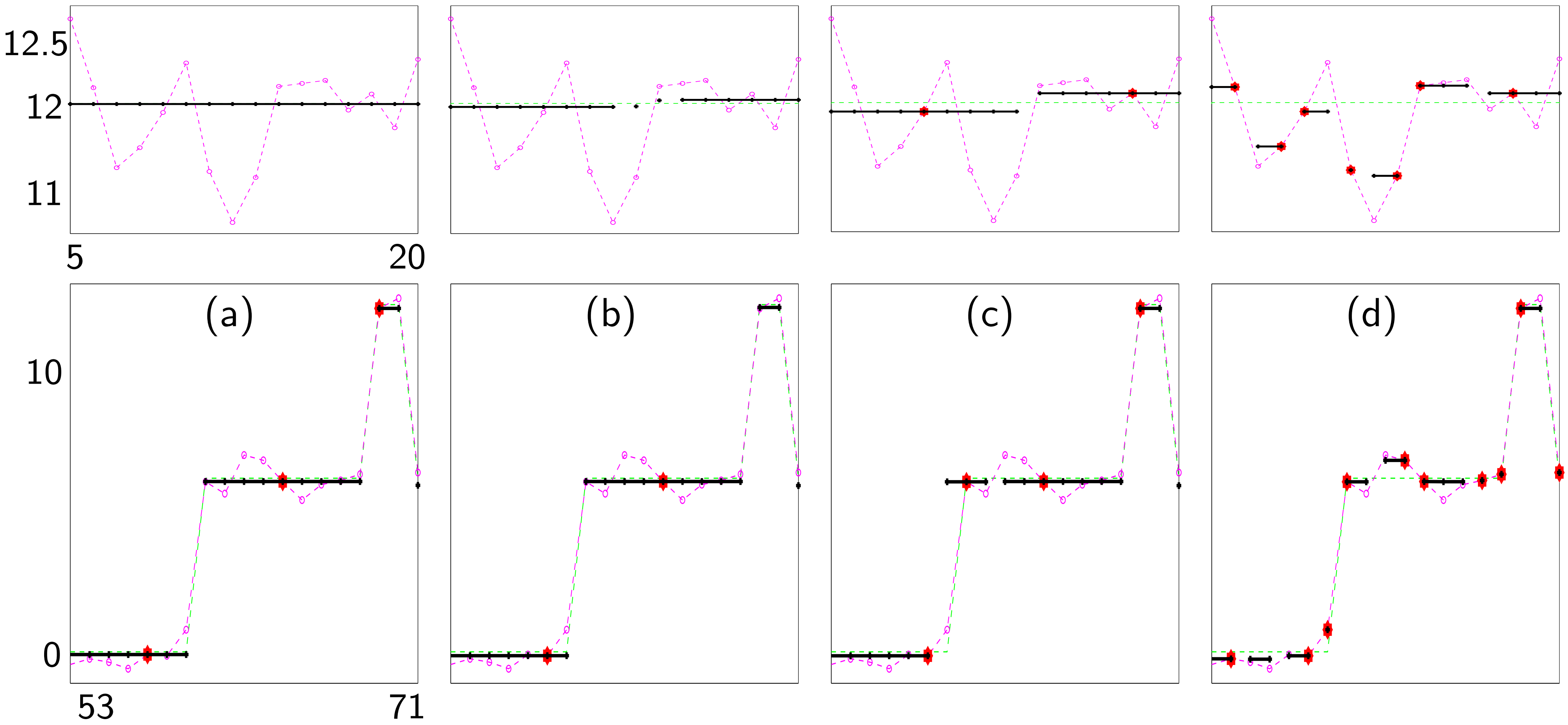


(c) $\varphi(t) = \ln(\alpha t + 1)$, $\alpha = 2$, $\beta = 1.3$



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

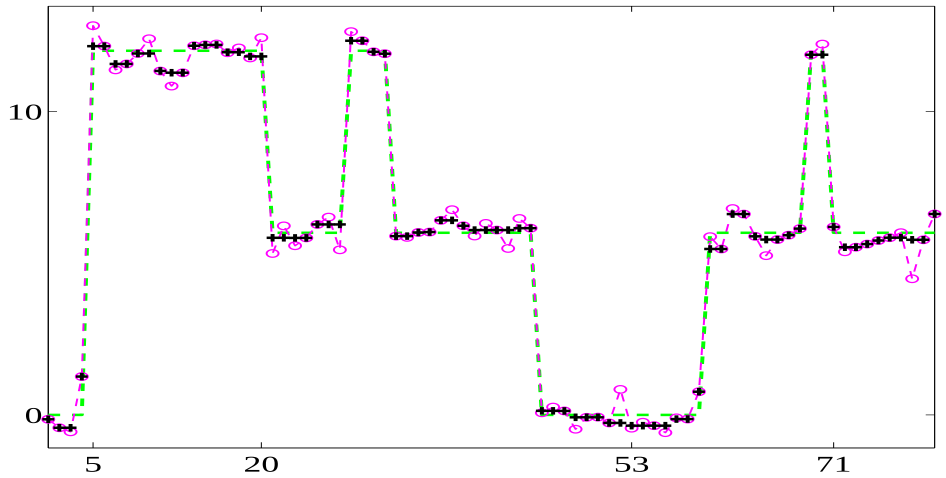
Denoising: Data samples (ooo) are corrupted with Gaussian noise. Minimizer samples $\hat{u}[i]$ (+++). Original (– – –). $\boldsymbol{\beta}$—the largest value so that the gate at 71 survives.



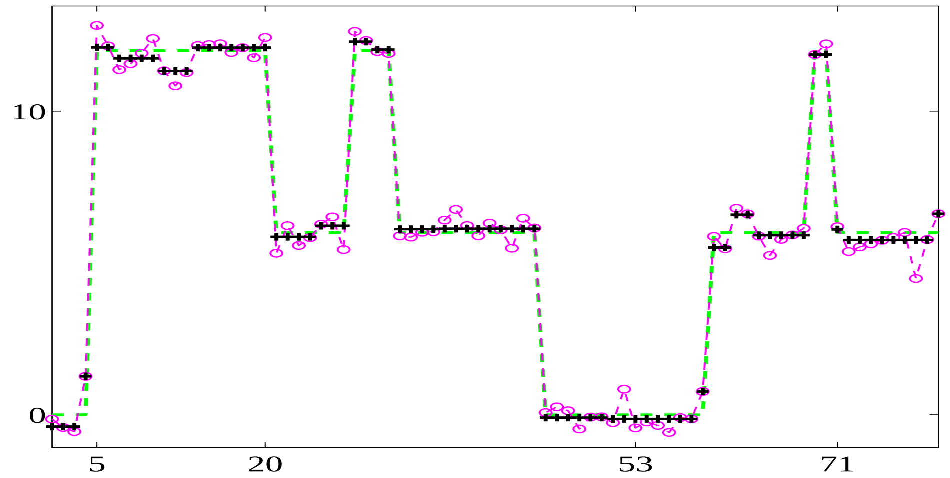
Zooms.

Constant pieces—solid black line.

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

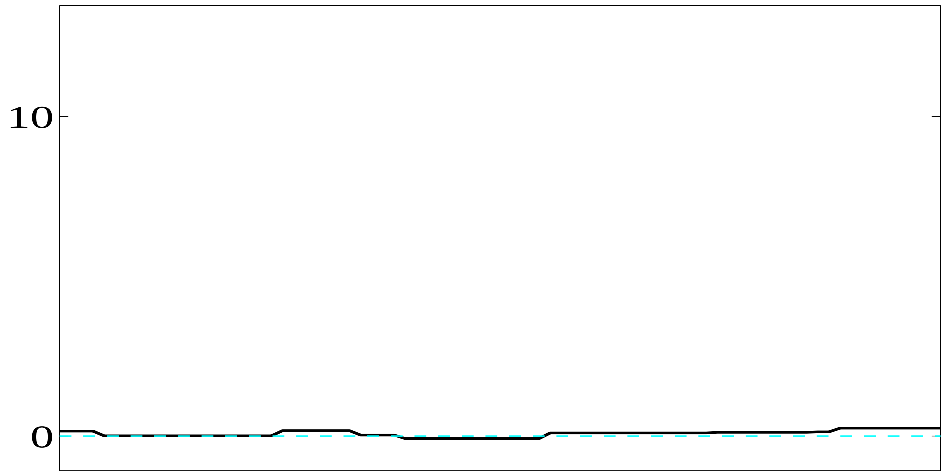


$\varphi(t) = t,\ \beta = 0.8 \quad (\ell_1 - \mathrm{TV})$



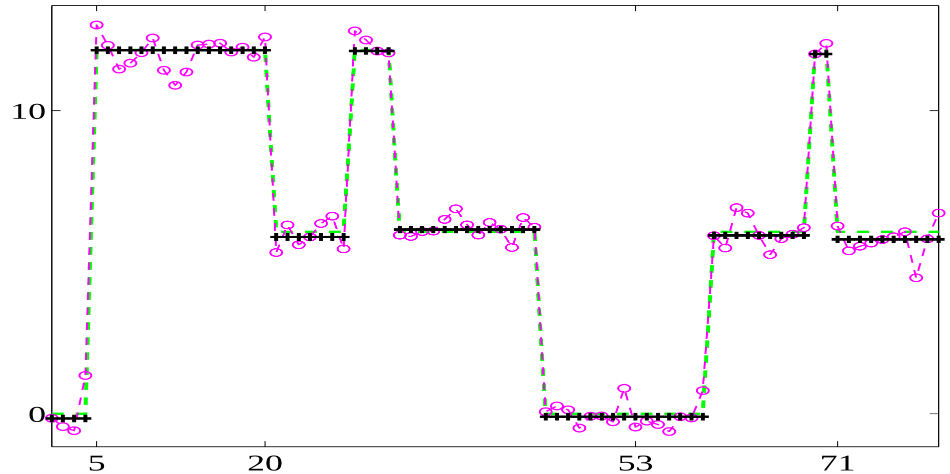
the minimizer for $\varphi(t) = \frac{\alpha\ t}{\alpha\ t+1}$, $\alpha = 4$, $\beta = 3$

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



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

$\|$original $- \hat{\boldsymbol{u}}\|_\infty = 0.2462$



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

original $\in [0, 12]$, data $v \in [-0.59, 12.83]$

Luckily, he same minimizers $\hat{u}$ were obtained using continuation and Viterbi algorithm (15×10^3 states) which yields a global minimizer.

Numerical evidence:

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

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

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

Main assumptions

$$\mathrm{G} = [\mathrm{G}_1^T, \cdots, \mathrm{G}_r^T]^T$$

H1 $\ker A \cap \ker G = \{0\}$.

H2 $\varphi : \mathbb{R}_+ \to \mathbb{R}$ *in* $\mathcal{F}$ *obeys:*

- $\varphi : \mathbb{R}_+ \to \mathbb{R}_+$ *is* $\mathcal{C}^2$ *on* $\mathbb{R}_+^* \stackrel{\text{def}}{=} \mathbb{R}_+ \setminus \{0\}$ *and* $\varphi(t) > \varphi(0), \ \forall t > 0$*;*

- $\varphi'(0^+) > 0$ *and* $\varphi'(t) > 0$ *on* $\mathbb{R}_+^*$*.*

- φ'' *is increasing on* $\mathbb{R}_+^*$*,* $\varphi''(t) < 0$*,* $\forall t > 0$ *and* $-\infty < \lim_{t \searrow 0} \varphi''(t) < 0$

Example

Given $v \neq 0$, consider the function

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

The necessary conditions for $\mathcal{F}$ to have a (local) minimum at $\hat{u} \neq 0$ and $\hat{u} \neq v$ fail:

$$\hat{u} \notin \{0, v\} \quad \Rightarrow \quad \left\{ \begin{array}{rcl} D\mathcal{F}(\hat{u}) = \text{sign}(\hat{u} - v) + \beta\varphi'(|\hat{u}|)\text{sign}(\hat{u}) & = & 0 \\ \\ D^2\mathcal{F}(\hat{u}) = \beta\varphi''(|\hat{u}|) & < & 0 \end{array} \right.$$

$\mathcal{F}$ does have minimizers $\quad \Rightarrow \quad \boldsymbol{\hat{u} \in \{0, v\}}$.

3. Main theoretical results

- $\left\{\hat{u} \in \mathbb{R}^p \;\mid\; \mathcal{F}(\hat{u}) = \inf_{u \in \mathbb{R}^p} \mathcal{F}(u)\right\} \neq \varnothing$

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

- Let $\hat{\boldsymbol{u}} \in \mathbb{R}^p$ be a (local) minimizer of $\mathcal{F}$. Set

$$\begin{aligned} \widehat{\boldsymbol{I}}_{\mathbf{0}} &= \{\boldsymbol{i} \in \boldsymbol{I} \;:\; \boldsymbol{a}_i \hat{\boldsymbol{u}} = \boldsymbol{v}[\boldsymbol{i}]\} \\ \widehat{\boldsymbol{J}}_{\mathbf{0}} &= \{\boldsymbol{j} \in \boldsymbol{J} \;:\; \mathbf{G}_j \hat{\boldsymbol{u}} = \mathbf{0}\} \end{aligned}$$

$\Rightarrow$ $\hat{\boldsymbol{u}}$ is the **unique** solution of the liner system

$$\begin{cases} \boldsymbol{a}_i \boldsymbol{u} = \boldsymbol{v}[\boldsymbol{i}] & \forall \boldsymbol{i} \in \widehat{\boldsymbol{I}}_{\mathbf{0}} \\ \boldsymbol{G}_j \boldsymbol{u} = \mathbf{0} & \forall \boldsymbol{j} \in \widehat{\boldsymbol{J}}_{\mathbf{0}} \end{cases}$$

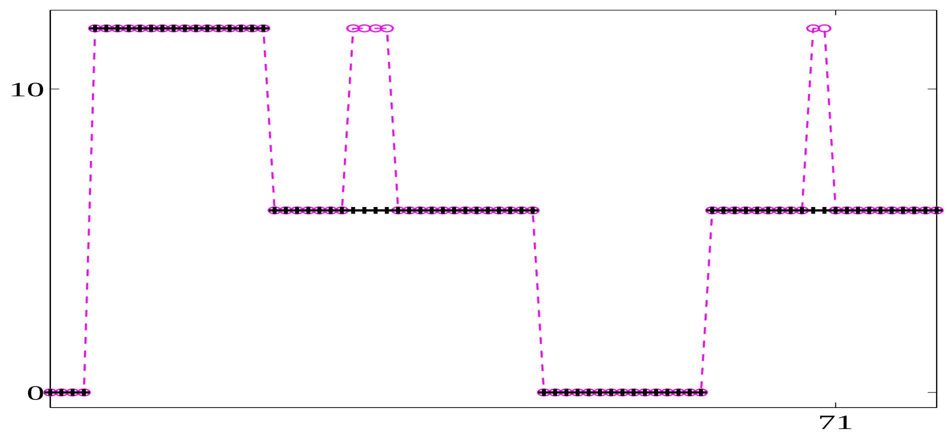
$\Rightarrow$

$(\star)$ the matrix $\boldsymbol{H}_{\widehat{I}_0, \widehat{J}_0}$ with rows $\left(\boldsymbol{a}_i, \forall \boldsymbol{i} \in \widehat{\boldsymbol{I}}_{\mathbf{0}}\right.$ and $\left.\mathbf{G}_j, \forall \boldsymbol{j} \in \widehat{\boldsymbol{J}}_{\mathbf{0}}\right)$

has **full column rank** $\quad \left(\operatorname{rank}\left(H_{\widehat{I}_0, \widehat{J}_0}\right) = p\right)$

$(\star)$ is a necessary condition for a (local) minimizer

Example



The data vector v is of length $\boldsymbol{p = 80}$.

One checks that the minimizer meets

$$\widehat{I}_0^c = (28\,,\ 29\,,\ 30\,,\ 31\,,\ 69\,,\ 70) \quad \text{and} \quad \widehat{J}_0^c = (4\,,\ 20\,,\ 44\,,\ 59)\ .$$

The matrix $\boldsymbol{H_{\widehat{I}_0,\widehat{J}_0}}$ is of size $\mathbf{149} \times \mathbf{80}$ and $\textbf{rank}\,\boldsymbol{H_{\widehat{I}_0,\widehat{J}_0} = p = 80}$.

$\Rightarrow$ "contrast invariance" of (local) minimizers $\hat{u}$ w.r.t v (like $\ell_1 - \mathrm{TV}$)

Is there another way to design / learn the matrix $\boldsymbol{H_{\widehat{I}_0,\widehat{J}_0}}$???

- Let $\hat{u} \in \mathbb{R}^p$ be a (local) minimizer of $\mathcal{F}$. Then

$$1 \leqslant k \leqslant p \ \Rightarrow \begin{cases} \exists i \text{ obeying } a_i\, \hat{u} = v[i] & \text{such that } a_i[k] \neq 0 \\ & \text{or} \\ \exists j \text{ obeying } \mathrm{G}_j \hat{u} = 0 & \text{such that } \mathrm{G}_j(k) \neq 0 \end{cases}$$

  where $\mathrm{G}_j(k)$ is the k-th column of the linear operator G_j

- $\Rightarrow$ each pixel of a (local) minimizer $\hat{u}$ of $\mathcal{F}$ is involved in (at least) one data equation that is fitted exactly $a_i\hat{u} = v[i]$, or in (at least) one vanishing operator $\|\mathrm{G}_j\hat{u}\|_2 = 0$, or in both types of equations.

- If $A = \mathrm{Id}$ and G_j yield discrete gradients or first-order finite differences between adjacent samples, a (local) minimizer is composed partly of constant patches, partly of pixels that fit data samples exactly, remind the figure.

4. Comparison with $\ell_1 - \mathrm{TV}$

$$\boldsymbol{F(u) \;=\; \sum_{i \in I} \big| a_i u - v[i] \big| + \beta \sum_{j \in J} \|\mathbf{G}_j u\|_2 \;, \quad \beta > 0 \;.}$$

$\mathbf{H1}$ $\ker A \cap \ker G = \{0\}$.

The set of minimizers: $\widehat{U} = \{\hat{u} \mid F(\hat{u}) = \min_{u \in \mathbb{R}^p} F(u)\}$

Typically, $\widehat{U}$ is not a singleton.

$\Rightarrow$ the matrix with rows $\big(\boldsymbol{a_i}, \forall \boldsymbol{i} \in \widehat{\boldsymbol{I}}_\mathbf{0}$ and $\mathbf{G}_{\boldsymbol{j}}, \forall \boldsymbol{j} \in \widehat{\boldsymbol{J}}_\mathbf{0}\big)$ typically

does not have full column rank

$(\star)$ If $\hat{u}_1 \in \widehat{U}$ and $\hat{u}_2 \in \widehat{U}$, $\hat{u}_1 \neq \hat{u}_2$ then

$$\mathbf{G}\hat{\boldsymbol{u}}_\mathbf{1} \propto \mathbf{G}\hat{\boldsymbol{u}}_\mathbf{2}$$

i.e. $\hat{\boldsymbol{u}}_\mathbf{1}$ and $\hat{\boldsymbol{u}}_\mathbf{2}$ share the same level lines. *[Durand, Nikolova 2007]*

$\boxed{\textbf{5. Numerical scheme}}$

Continuation approach

φ_ε, $\varepsilon \in [0,1]$ where $\varphi_0(t) = t$ and $\varphi_1 = \varphi$

$$\varphi_\varepsilon(t) = \psi_\varepsilon(t) + \alpha_\varepsilon t \quad \text{where} \quad \alpha_\varepsilon = \varphi'_\varepsilon(0^+).$$

φ_ε for $\varepsilon \in (0,1]$ satisfies H2.

$$\begin{aligned} \mathcal{F}_\varepsilon(u) &= \|Au - v\|_1 + \beta\alpha_\varepsilon \sum_{j\in J} \|\mathrm{G}_j u\|_2 + \beta\Psi_\varepsilon(u) \ , \\ \text{where} \quad \Psi_\varepsilon(u) &= \sum_{j\in J} \psi_\varepsilon(\|\mathrm{G}_j u\|_2) \ . \end{aligned}$$

For $\varepsilon = 0$: $\quad \mathcal{F}_0(u) = \|Au - v\|_1 + \beta\alpha_\varepsilon \mathrm{TV}(u)$

For each ε fixed—variable splitting and penalty decomposition techniques:

$$\mathcal{J}_{\varepsilon,\gamma}(u,w,z) = \gamma\|Au-w\|_2^2+\|w-v\|_1+\beta\Psi_\varepsilon(u)+\gamma\|\mathrm{G}u-z\|_2^2+\beta\alpha_\varepsilon\sum_{j\in J}\|z_j\|_2\ , \ \text{ for } \gamma\to\infty$$

Alternate optimization: $\left\{\begin{array}{lcl} z^{(k)} & = & \arg\min\limits_{z} \mathcal{J}_{\varepsilon,\gamma}(u^{(k-1)},w^{(k-1)},z^{(k-1)}) \\ w^{(k)} & = & \arg\min\limits_{w} \mathcal{J}_{\varepsilon,\gamma}(u^{(k-1)},w^{(k-1)},z^{(k)}) \\ u^{(k)} & = & \arg\min\limits_{u} \mathcal{J}_{\varepsilon,\gamma}(u,w^{(k)},z^{(k)}) \end{array}\right.$

Then

$$z_j^{(k)} = \frac{\mathrm{G}_j u^{(k-1)}}{\|\mathrm{G}_j u^{(k-1)}\|_2}\ \max\left\{\|\mathrm{G}_j u^{(k-1)}\|_2 - \frac{\beta\alpha_\varepsilon}{2\gamma},\ 0\right\},\quad \forall j\in J\ .$$

$$w_i^{(k)} = \frac{Au^{(k-1)}-v}{\|Au^{(k-1)}-v\|_2}\ \max\left\{\|Au^{(k-1)}-v\|_2 - \frac{1}{2\gamma},\ 0\right\},\quad \forall i\in I\ .$$

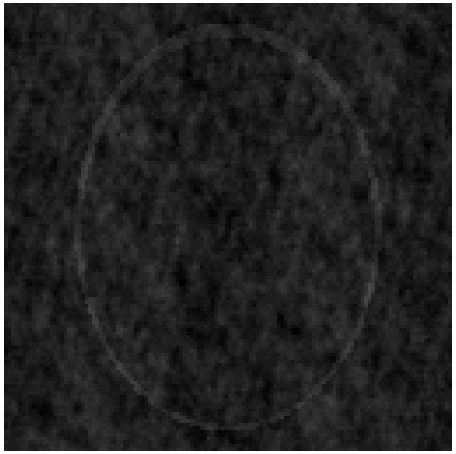
$$u^{(k)} \text{ solves } \arg\min_{u\in\mathbb{R}^p}\left\{\gamma\|Au-w^{(k)}\|_2^2+\gamma\|\mathrm{G}u-z^{(k)}\|_2^2+\beta\Psi_\varepsilon(u)\right\}$$

where Quasi Newton method with preconditioning is used

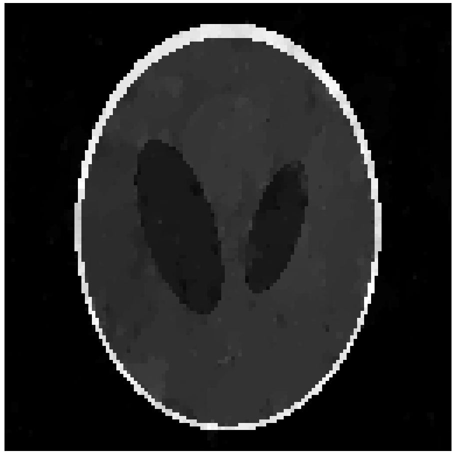
$\Rightarrow$ fast algorithm

6. Numerical tests

MR Image Reconstruction from Highly Undersampled Data



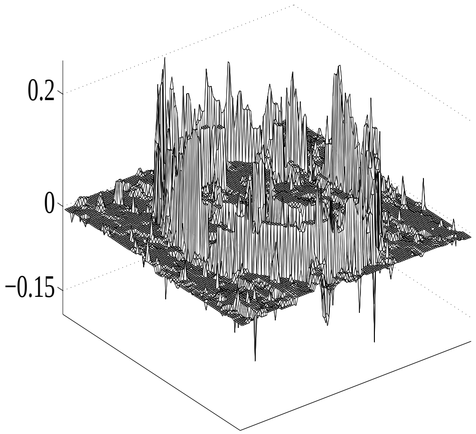
0-filling Fourier

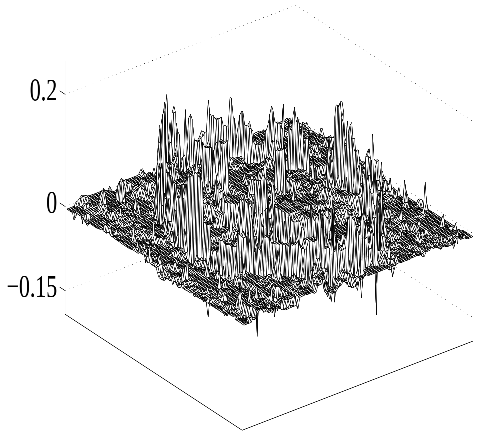


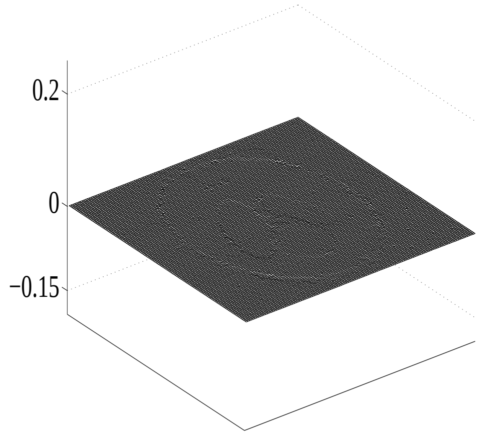
$\|\cdot\|_2^2$+TV

$\|\cdot\|_1$+TV

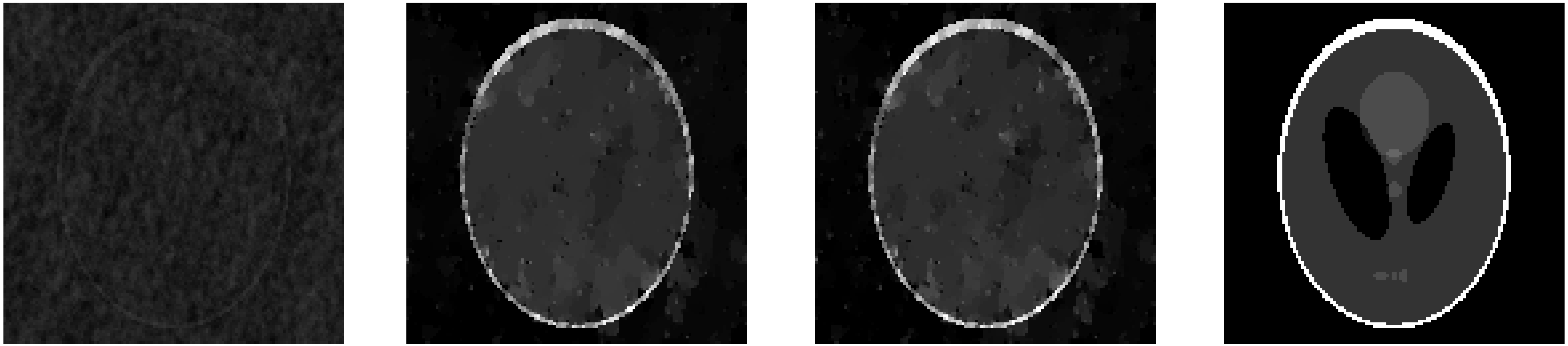
Our method



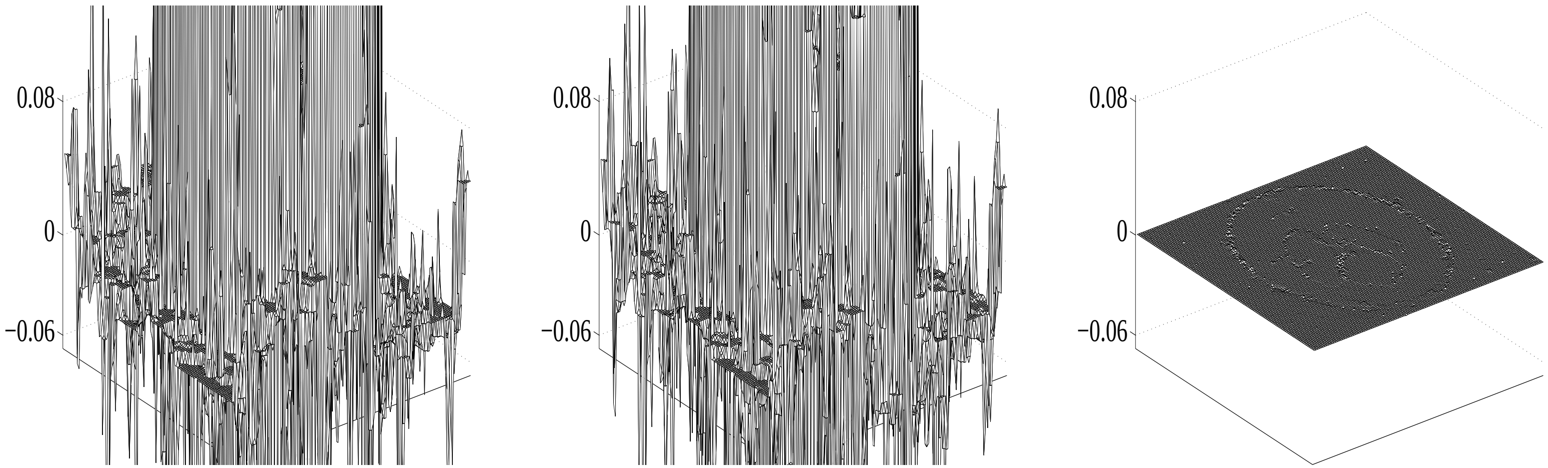




Reconstructed images from **7% noisy** randomly selected samples in the k-space.



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



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

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

Cartoon

Observed

ℓ_1-TV

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

7. Concluding remarks

- The (local) minimizers of the proposed objectives inherit some features of L_1-TV (e.g. "scale-invariance") but in a much sharper way.
- In practice, they neatly outperform L_1-TV.
- All (local) minimizers are strict.
- Bounded above functions (like f1 and f2) yield much better numerical results than coercive functions (like f3 and f4).

  We do not have a theoretical explanation.
- The regularization parameter β is not involved in the computation of a local minimizer.

  Implicitly, β helps the selection of the subsets $\widehat{I_0}$ and $\widehat{J_0}$.

  The ordering of the (local) minimizers $\hat{u}$ of $\mathcal{F}$ according to their value $\mathcal{F}(\hat{u})$ is determined by β.

Thank you for your attention!

Thanks to the Organizers for the invitation and for the excellent conference!

Main reference:

M. Nikolova, M. Ng and C. P. Tam, *On ℓ_1 Data Fitting and Concave Regularization for Image Recovery*, SIAM J. on Scientific Computing, vol. 35, No. 1, pp. A397–A430, online 24 Jan 2013

More details:

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