ℓ_1 Data Fitting with Concave Regularization for Image Recovery

Mila Nikolova

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

Joint work with

Michael K. Ng (HKBU), Hong Kong) and Chi-Pan Tam (HKBU)

SIAM Conference on IMAGING SCIENCE
minisymposium: Advances in nonsmooth and nonconvex minimization for imaging

Philadelphia, USA, May 20 - 22, 2012

1. Problem formulation

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

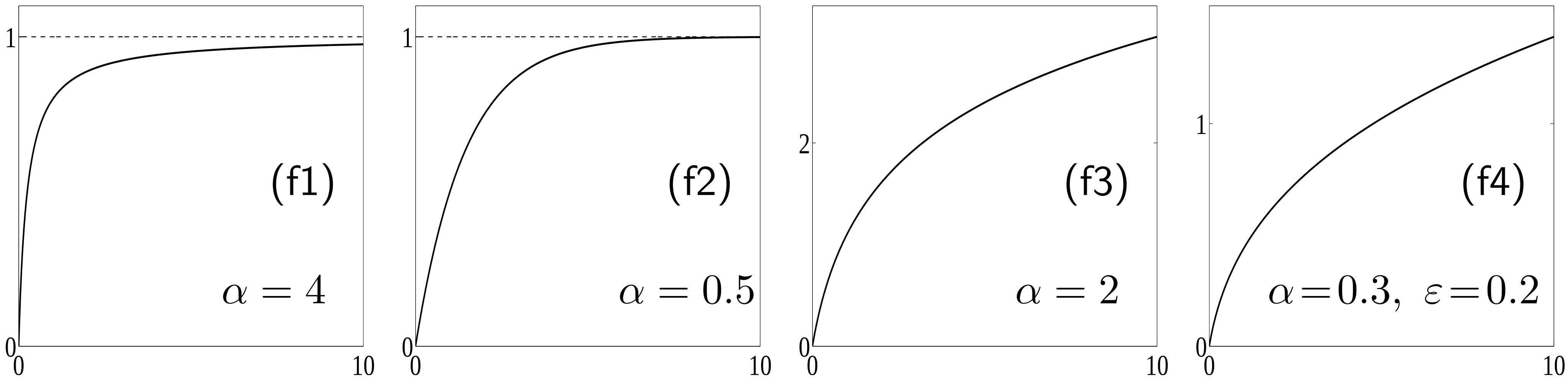
$$\begin{aligned}\mathcal{F}(u) &= \|Au - v\|_1 + \beta \sum_{j \in J} \varphi(\|\mathrm{G}_j u\|_2) \\ &= \sum_{i \in I} \big|a_i u - v[i]\big| + \beta \sum_{j \in J} \varphi(\|\mathrm{G}_j u\|_2) \ , \quad \beta > 0 \ ,\end{aligned}$$

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

- A is a matrix of any rank with rows $a_i \in \mathbb{R}^{1 \times p}$
- G_j are matrices or vectors (e.g. discrete gradient operators)
- φ is concave on $\mathbb{R}_+$

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

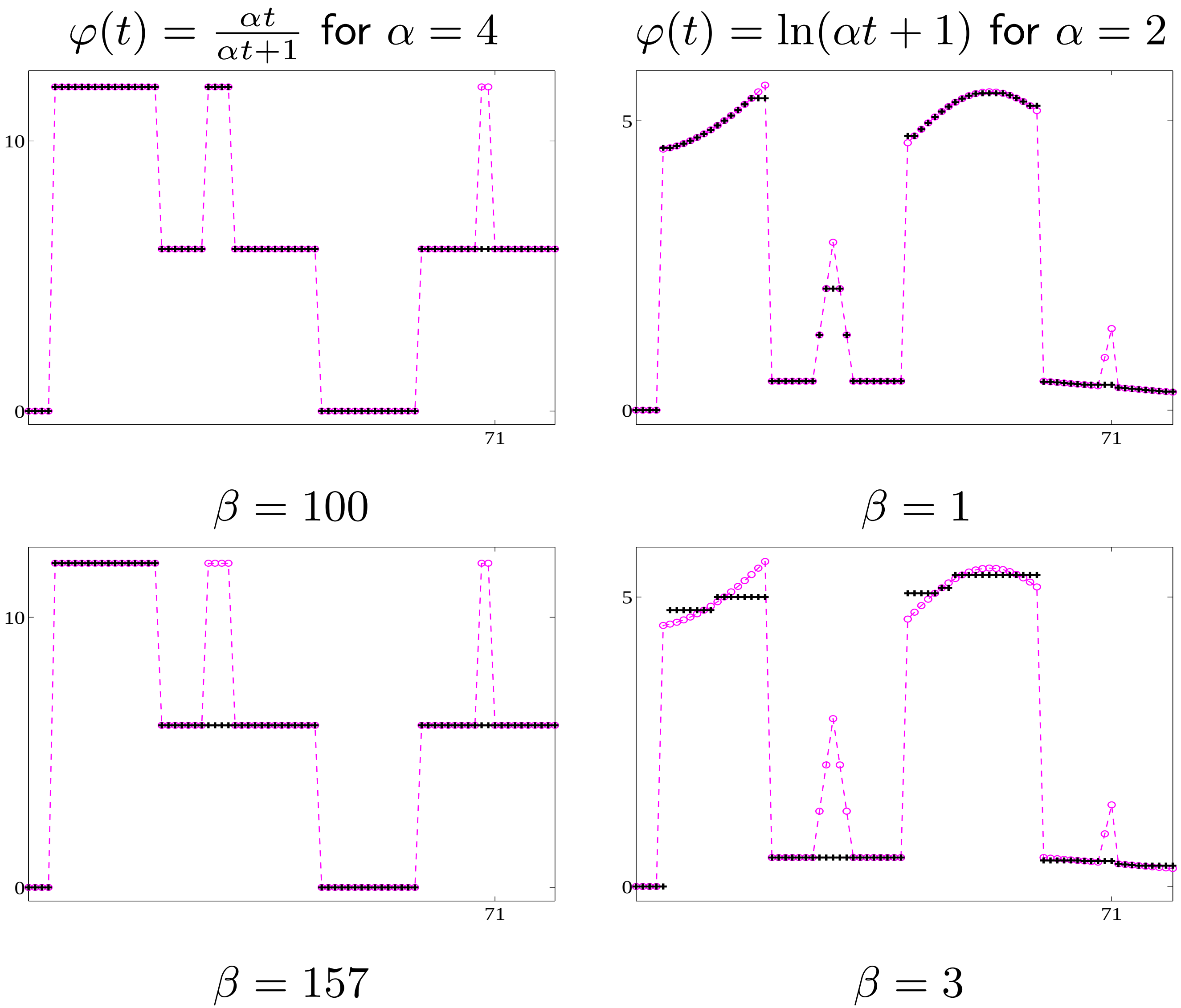
Functions $\boldsymbol{\varphi} : \mathbb{R}_+ \to \mathbb{R}_+$



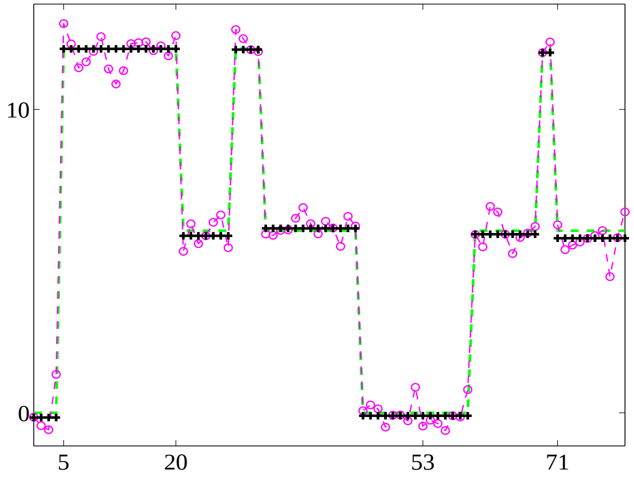
Plots of the PFs φ. Note that (f1) an (f2) are bounded above, (f3) and (f4) are coercive.

2. Peculiar Properties of Minimizers

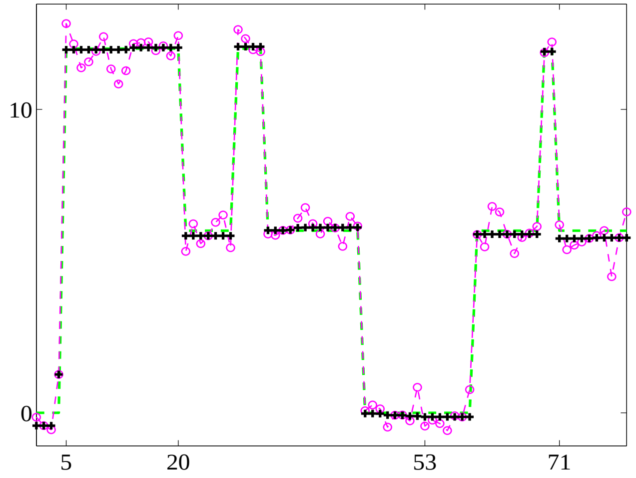
Illustrations by minimizing $\mathcal{F}(u) = \|u - v\|_1 + \beta \sum_{i=1}^{p-1} \varphi(|u[i+1] - u[i]|)$



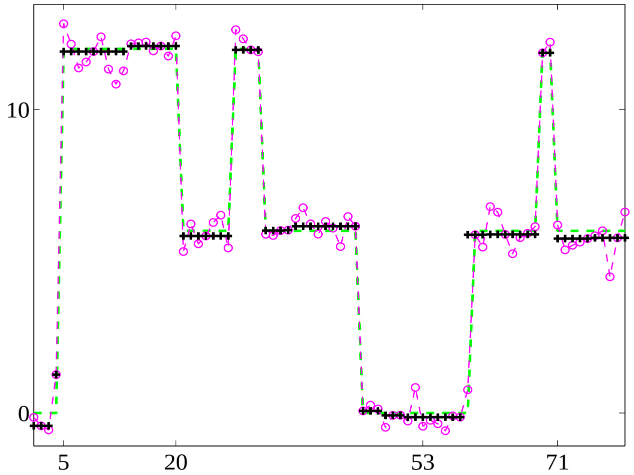
Data samples (ooo), Minimizer samples $\hat{u}[i]$ (+++).



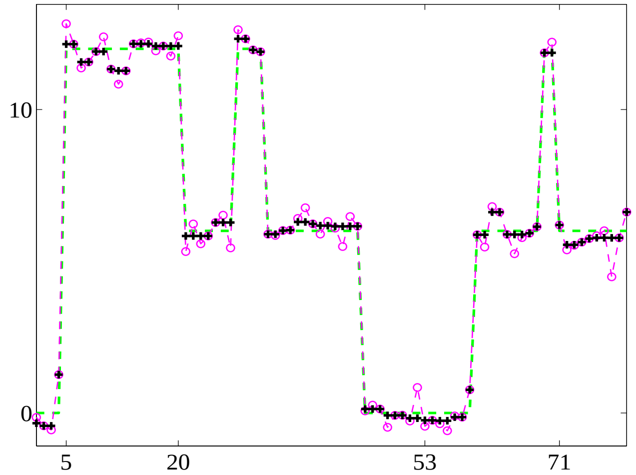
$\varphi(t) = \frac{\alpha\, t}{\alpha\, t+1}$, $\alpha = 4$, $\beta = 3$



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

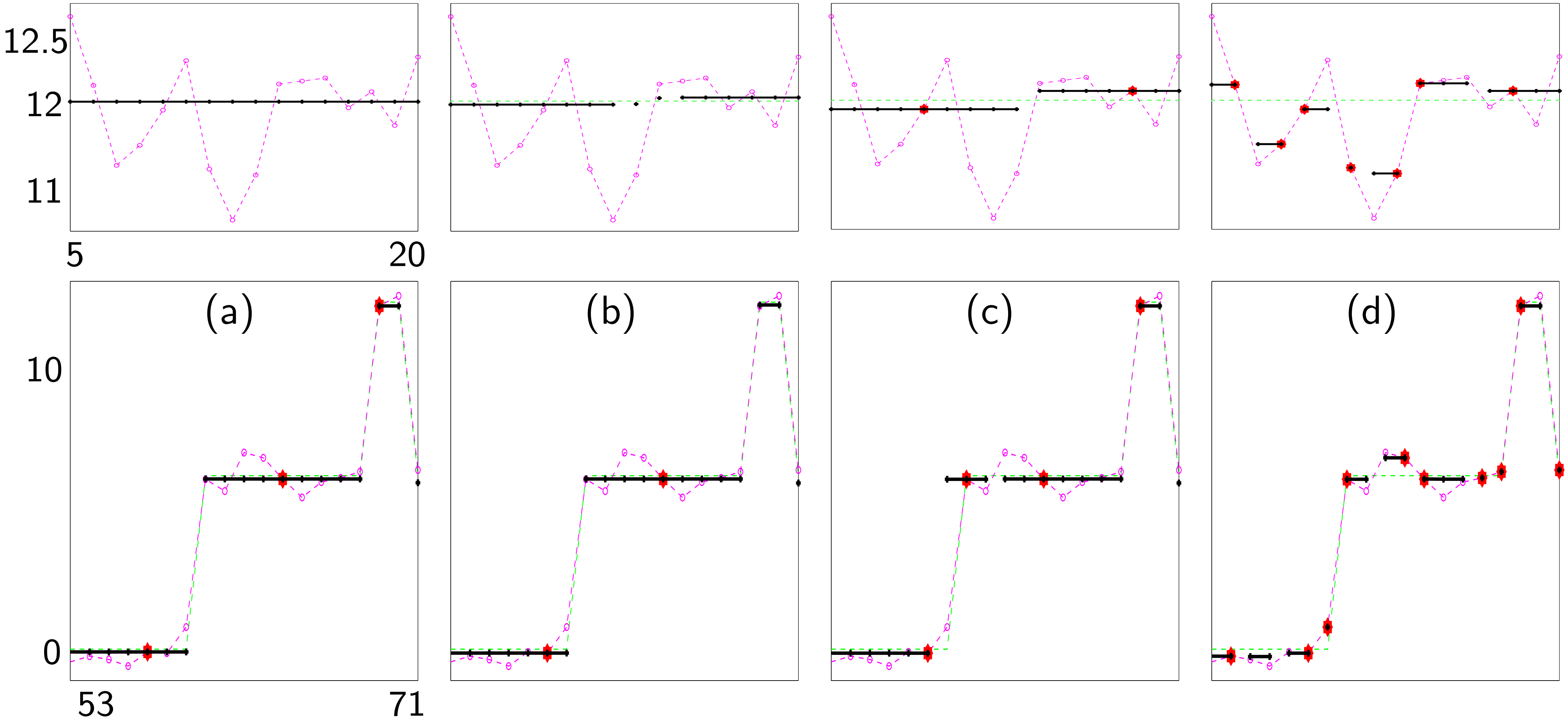


$\varphi(t) = \ln(\alpha t + 1)$, $\alpha = 2$, $\beta = 1.3$



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

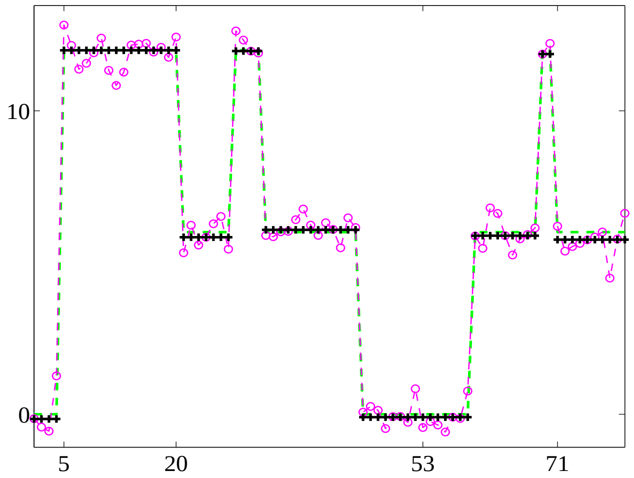
Data samples (ooo) are corrupted with Gaussian noise. Denoising. Minimizer samples $\hat{u}[i]$ (+++). Original (– – –). β—the largest value so that the gate at 71 remains.



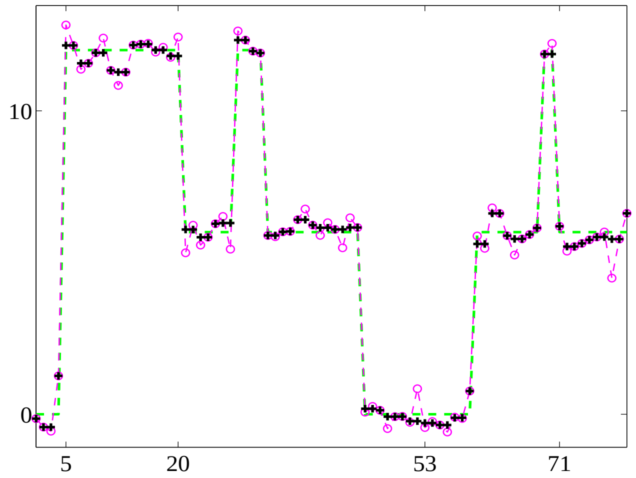
Zooms.

Constant pieces—solid black line.

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



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



$\varphi(t) = t$, $\beta = 0.8$ (TV)

Numerical evidence for $\mathcal{F}$: critical values $\beta_1, \cdots, \beta_n$ such that

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

Example

Given $v \neq 0$, consider the pair of functions given below

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

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

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

Since $\mathcal{F}$ does have minimizers, they necessarily meet

$$\hat{u} \in \{0, v\} \ .$$

Essential 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* $\mathbb{R}_+^* \stackrel{\mathrm{def}}{=} \mathbb{R}_+ \setminus \{0\}$ *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*

  $\lim_{t \searrow 0} \varphi''(t) < 0$ *is well defined and finite.*

Main theoretical results

- $\hat{U} \overset{\text{def}}{=} \left\{ \hat{u} \in \mathbb{R}^p \;\middle|\; \mathcal{F}(\hat{u}) = \inf_{u \in \mathbb{R}^p} \mathcal{F}(u) \right\} \neq \varnothing$

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

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

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

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

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

  $\Rightarrow$ The matrix with rows $\left(\boldsymbol{a_i}, \forall \boldsymbol{i} \in \boldsymbol{\hat{I}_0} \text{ and } \mathbf{G}_{\boldsymbol{j}}, \forall \boldsymbol{j} \in \boldsymbol{\hat{J}_0}\right)$ has **full column rank**

- $\Rightarrow$ "contrast invariance" of (local) minimizers since $\hat{u}$ is linear in v

- $\Rightarrow$ Necessary condition for a (local) minimizer

- 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 differential 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.

3. Numerical scheme

Continuation approach

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

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

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

$$\mathcal{F}_\varepsilon(u) = \|Au - v\|_1 + \beta\alpha_\varepsilon \sum_{j\in J} \|\mathrm{G}_j u\|_2 + \beta\Psi_\varepsilon(u) \ ,$$

$$\text{where} \quad \Psi_\varepsilon(u) = \sum_{j\in J} \psi_\varepsilon(\|\mathrm{G}_j u\|_2) \ .$$

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

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

Alternate optimization: $$\begin{cases} z^{(k)} &= \arg\min_z \mathcal{J}_{\varepsilon,\gamma}(u^{(k-1)}, w^{(k-1)}, z^{(k-1)}) \\ w^{(k)} &= \arg\min_w \mathcal{J}_{\varepsilon,\gamma}(u^{(k-1)}, w^{(k-1)}, z^{(k)}) \\ u^{(k)} &= \arg\min_u \mathcal{J}_{\varepsilon,\gamma}(u, w^{(k)}, z^{(k)}) \end{cases}$$

Then

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

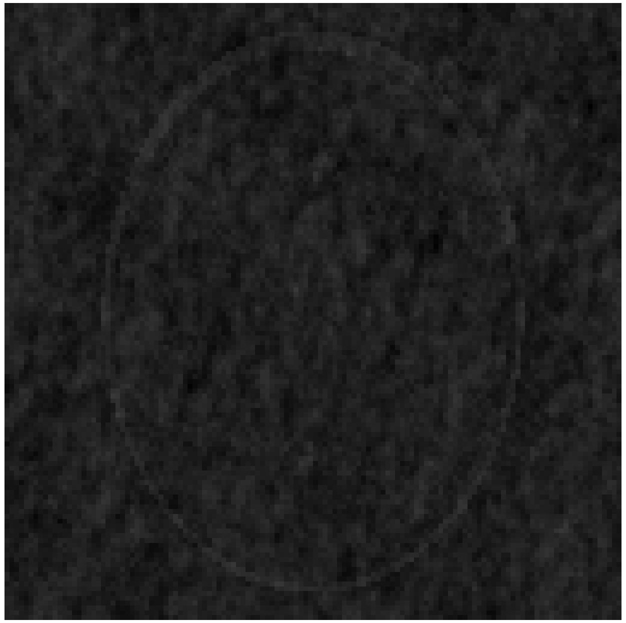
$$w_i^{(k)} = \frac{Au^{(k-1)} - v}{\|Au^{(k-1)} - v\|_2} \max\left\{ \|Au^{(k-1)} - v\|_2 - \frac{1}{2\gamma},\ 0 \right\}, \quad \forall i \in I \ .$$

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

where Quasi Newton method with preconditioning is used

$\Rightarrow$ fast algorithm

4. MR Image Reconstruction from Highly Undersampled Data



Zero-filling Fourier recovery

ℓ_2-TV

Our method

Reconstructed images from **5%** noisy randomly selected samples in the k-space using $\varphi(t) = \frac{\alpha t}{\alpha t+1}$.

5. 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.