$\ell_1$—concave versus $\ell_1$ — 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 \( u \) is stored in a vector in \( \mathbb{R}^p \)
data \( v \in \mathbb{R}^q \)

\[
\mathcal{F}(u) = \|Au - v\|_1 + \beta \sum_{j \in J} \varphi(\|G_ju\|_2)
\]

\[
= \sum_{i \in I} |a_iu - v[i]| + \beta \sum_{j \in J} \varphi(\|G_ju\|_2), \quad \beta > 0,
\]

where

- \( I \) def \( \{1, \cdots, q\} \),
- \( J \) def \( \{1, \cdots, r\} \).

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

\( \varphi \) is concave on \( \mathbb{R}_+ \)

This family of objective functions has never been considered before.
<table>
<thead>
<tr>
<th>φ(t)</th>
<th>(f1)</th>
<th>(f2)</th>
<th>(f3)</th>
<th>(f4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\frac{\alpha t}{\alpha t + 1}$</td>
<td>$1 - \alpha^t$</td>
<td>$\ln(\alpha t + 1)$</td>
<td>$(t + \varepsilon)^\alpha$</td>
<td></td>
</tr>
<tr>
<td>$\alpha &gt; 0$</td>
<td>$0 &lt; \alpha &lt; 1$</td>
<td>$\alpha &gt; 0$</td>
<td>$0 &lt; \alpha &lt; 1, \varepsilon &gt; 0$</td>
<td></td>
</tr>
</tbody>
</table>

Functions $\varphi : \mathbb{R}_+ \rightarrow \mathbb{R}_+$

Plots of the PFs $\varphi$. Note that (f1) and (f2) are bounded above, (f3) and (f4) are coercive.
Motivation

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

- nonsmooth regularization \( \sum_{j \in J} \varphi(\|G_ju\|_2) \) with \( \varphi'(0) > 0 \) (e.g. TV)

\[ \Rightarrow \text{many } j \text{ such that } G_j \hat{u} = 0 \quad [\text{Nikolova 2000, 2004}] \]

- \( \ell_1 \) data fidelity \( \|Au - v\|_1 = \sum_{i \in I} |a_iu - v[i]| \)

\[ \Rightarrow \text{many } i \text{ such that } a_i \hat{u} = v[i] \quad [\text{Nikolova 2002, 2004}] \]

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

???

??? many \( i, j \) such that \( a_i \hat{u} = v[i] \) and \( G_j \hat{u} = 0 \)

2. Peculiar Properties — 1D tests
(Global) minimizers of \( F(u) = \|u - v\|_1 + \beta \sum_{i=1}^{p-1} \varphi(|u[i+1] - u[i]|) \)

\[ \varphi(t) = \frac{\alpha t}{\alpha t + 1} \text{ for } \alpha = 4 \]

\[ \varphi(t) = \ln(\alpha t + 1) \text{ for } \alpha = 2 \]

\[ \beta \in \{78, \cdots, 156\} \]

\[ \beta \in 0.1 \times \{10, \cdots, 14\} \]

\[ \beta \in \{157, \cdots, 400\} \]

\[ \beta \in 0.1 \times \{16, \cdots, 30\} \]

Data samples (ooo), Minimizer samples \( \hat{u}[i] \) (++++).
(a) $\varphi(t) = \frac{\alpha t}{\alpha t+1}$, $\alpha = 4$, $\beta = 3$

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

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

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

**Denoising:** Data samples (○○○○) are corrupted with Gaussian noise. Minimizer samples $\hat{u}[i]$ (++++). Original (-----). $\beta$—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 $\hat{u}$ (◇).
\[ \varphi(t) = t, \beta = 0.8 \quad (\ell_1 - TV) \]

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

closest to \((\ell_1 - TV)\)

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

\[ \| \text{original} - \hat{u} \|_{\infty} = 0.2462 \]

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

original \(\in [0, 12]\), data \(v \in [-0.59, 12.83]\)
Luckily, he same minimizers \( \hat{u} \) were obtained using continuation and Viterbi algorithm (\( 15 \times 10^3 \) 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
- \( \beta \geq \beta_{i+1} \Rightarrow \) the minimizer is simplified

Result proven (under conditions) for the minimizers of \( L_1 - TV \) \[ \text{[Chan, Esedoglu 2005]} \]
Main assumptions

\[ G = [G_1^T, \cdots, G_r^T]^T \]

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

\textbf{H2} \ \varphi : \mathbb{R}_+ \rightarrow \mathbb{R} \text{ in } \mathcal{F} \text{ obeys:}

- \ \varphi : \mathbb{R}_+ \rightarrow \mathbb{R}_+ \text{ is } C^2 \text{ on } \mathbb{R}_+^* \overset{\text{def}}{=} \mathbb{R}_+ \setminus \{0\} \text{ and } \varphi(t) > \varphi(0), \ \forall t > 0; \\
- \ \varphi'(0^+) > 0 \text{ and } \varphi'(t) > 0 \text{ on } \mathbb{R}_+^*.

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

Given \( v \neq 0 \), consider the function

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

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

\[
\hat{u} \not\in \{0, v\} \quad \Rightarrow \quad \begin{cases}
DF(\hat{u}) = \text{sign}(\hat{u} - v) + \beta \varphi'(|\hat{u}|)\text{sign}(\hat{u}) = 0 \\
D^2F(\hat{u}) = \beta \varphi''(|\hat{u}|) < 0
\end{cases}
\]

\( F \) does have minimizers \( \Rightarrow \) \( \hat{u} \in \{0, v\} \).
3. Main theoretical results

- \( \{ \hat{u} \in \mathbb{R}^p \mid \mathcal{F}(\hat{u}) = \inf_{u \in \mathbb{R}^p} \mathcal{F}(u) \} \neq \emptyset \)

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

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

  \[
  \hat{I}_0 = \{ i \in I : a_i \hat{u} = v[i] \} \\
  \hat{J}_0 = \{ j \in J : G_j \hat{u} = 0 \}
  \]

  \( \Rightarrow \) \( \hat{u} \) is the unique solution of the liner system

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

  \( \Rightarrow \)

  \( (\star) \) the matrix \( H_{\hat{I}_0, \hat{J}_0} \) with rows \((a_i, \forall i \in \hat{I}_0 \text{ and } G_j, \forall j \in \hat{J}_0)\)

  has full column rank \( \text{rank}(H_{\hat{I}_0, \hat{J}_0}) = p \)

  \( (\star) \) is a necessary condition for a (local) minimizer
The data vector $v$ is of length $p = 80$.

One checks that the minimizer meets

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

The matrix $H_{\hat{I}_0, \hat{J}_0}$ is of size $149 \times 80$ and $\text{rank} \ H_{\hat{I}_0, \hat{J}_0} = p = 80$.

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

Is there another way to design / learn the matrix $H_{\hat{I}_0, \hat{J}_0}$???
• Let \( \hat{u} \in \mathbb{R}^p \) be a (local) minimizer of \( \mathcal{F} \). Then

\[
1 \leq k \leq p \Rightarrow \begin{cases} 
\exists i \text{ obeying } a_i \hat{u} = v[i] \text{ such that } a_i[k] \neq 0 \\
\text{or} \\
\exists j \text{ obeying } G_j \hat{u} = 0 \text{ such that } G_j(k) \neq 0
\end{cases}
\]

where \( G_j(k) \) 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 \( \|G_j \hat{u}\|_2 = 0 \), or in both types of equations.

• If \( A = \text{Id} \) 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 $\ell_1-\text{TV}$

\[
F(u) = \sum_{i \in I} |a_i u - v[i]| + \beta \sum_{j \in J} \|G_j u\|_2, \quad \beta > 0.
\]

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

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

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

⇒ the matrix with rows $(a_i, \forall i \in \hat{I}_0$ and $G_j, \forall j \in \hat{J}_0)$ typically does not have full column rank

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

\[
G\hat{u}_1 \propto G\hat{u}_2
\]

i.e. $\hat{u}_1$ and $\hat{u}_2$ share the same level lines. [Durand, Nikolova 2007]
5. Numerical scheme

Continuation approach

\[ \varphi_\epsilon, \epsilon \in [0, 1] \text{ where } \varphi_0(t) = t \text{ and } \varphi_1 = \varphi \]

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

\( \varphi_\epsilon \) for \( \epsilon \in (0, 1] \) satisfies H2.

\[
\mathcal{F}_\epsilon(u) = \|Au - v\|_1 + \beta \alpha_\epsilon \sum_{j \in J} \|G_ju\|_2 + \beta \Psi_\epsilon(u),
\]

where \( \Psi_\epsilon(u) = \sum_{j \in J} \psi_\epsilon(\|G_ju\|_2) \).

For \( \epsilon = 0 \): \( \mathcal{F}_0(u) = \|Au - v\|_1 + \beta \alpha_\epsilon \text{TV}(u) \)
For each $\varepsilon$ fixed—variable splitting and penalty decomposition techniques:

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

for $\gamma \to \infty$

Alternate optimization:

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

Then

$$
\begin{align*}
  z_j^{(k)} &= \frac{G_j u^{(k-1)}}{\| G_j u^{(k-1)} \|_2} \max \left\{ \| G_j u^{(k-1)} \|_2 - \frac{\beta \alpha_\varepsilon}{2\gamma}, 0 \right\}, \quad \forall j \in J .
\end{align*}
$$

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

$u^{(k)}$ solves

$$
\begin{align*}
  \arg \min_{u \in \mathbb{R}^p} \left\{ \gamma \| Au - w^{(k)} \|^2_2 + \gamma \| Gu - z^{(k)} \|^2_2 + \beta \Psi_\varepsilon(u) \right\}
\end{align*}
$$

where Quasi Newton method with preconditioning is used

$\Rightarrow$ fast algorithm
6. Numerical tests

MR Image Reconstruction from Highly Undersampled Data

Reconstructed images from 7% noisy randomly selected samples in the $k$-space.
Reconstructed images from 5% noisy randomly selected samples in the $k$-space.

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

\( \ell_1\text{-TV} \)

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

Cartoon
7. Concluding remarks

- The (local) minimizers of the proposed objectives inherit some features of $L_1-\text{TV}$ (e.g. “scale-invariance”) but in a much sharper way.

- In practice, they neatly outperform $L_1-\text{TV}$.

- All (local) minimizers are strict.

- Bounded above functions (like $f_1$ and $f_2$) yield much better numerical results than coercive functions (like $f_3$ and $f_4$).

  We do not have a theoretical explanation.

- The regularization parameter $\beta$ is not involved in the computation of a local minimizer.

  Implicitly, $\beta$ helps the selection of the subsets $\hat{I}_0$ and $\hat{J}_0$.

  The ordering of the (local) minimizers $\hat{u}$ of $\mathcal{F}$ according to their value $\mathcal{F}(\hat{u})$ is determined by $\beta$. 
Thank you for your attention!

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

Main reference:


More details: