$l_1$ Data Fidelity with Concave Regularization: Challenges

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

Joint work with
Michael K. Ng and Chi-Pan Tam — HKBU, Hong Kong

to appear in SIAM J. on Scientific Computing

International Conference on Imaging Science 2012
(In honour of Professor Stanley Osher and his birthday)
December 12-14, 2012, Hong Kong
1. Problem formulation

Image \( u \) is stored in a vector in \( \mathbb{R}^p \) and data \( v \in \mathbb{R}^q \)

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

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

where

\[
I \overset{\text{def}}{=} \{1, \cdots, q\},
\]

\[
J \overset{\text{def}}{=} \{1, \cdots, r\}
\]

• \( A \) 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)

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

This family of objective functions has never been considered before
<table>
<thead>
<tr>
<th>$\varphi(t)$</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 $\varphi : \mathbb{R}_+ \to \mathbb{R}_+$

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

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

- nonsmooth regularization \( \sum_{j \in J} \varphi(\|G_j u\|_2) \) with \( \varphi'(0) > 0 \) (e.g. TV)
  \[ \Rightarrow \text{many } j \text{ such that } G_j \hat{u} = 0 \]  
  \[ \text{[Nikolova 2000, 2004]} \]

- \( \ell_1 \) data fidelity \( \|Au - v\|_1 = \sum_{i \in I} |a_i u - v[i]| \)
  \[ \Rightarrow \text{many } i \text{ such that } a_i \hat{u} = v[i] \]  
  \[ \text{[Nikolova 2002, 2004]} \]

- our \( F \) can be seen as an extension of L1-TV
  \[ \text{[Chan, Esedoglu 2005]} \]
  \[ \text{[Chan, Esedoglu, Nikolova 2006]} \]

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

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

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

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

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

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

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

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

Data samples (○○○) are corrupted with Gaussian noise. Denoising. Minimizer samples $\hat{u}[i]$ (+++). Original (———). $\beta$—the largest value so that the gate at 71 remains.
Constant pieces—solid black line.

Data points $v[\hat{\ell}]$ fitted exactly by the minimizer $\hat{u}$ (♦).
\( \varphi(t) = \alpha \frac{t}{t+1} \), \( \alpha = 4 \), \( \beta = 3 \)

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

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
Main assumptions

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

**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 \( C^2 \) on \( \mathbb{R}_+^* \overset{\text{def}}{=} \mathbb{R}_+ \setminus \{0\} \) and \( \varphi(t) > \varphi(0), \forall t > 0; \)

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

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

\[ \lim_{t \to 0} \varphi''(t) < 0 \] is well defined and finite.
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:

\[
\hat{u} \notin \{0, v\} \implies \begin{cases} 
D F(\hat{u}) = \text{sign}(\hat{u} - v) + \beta \varphi'(|\hat{u}|) \text{sign}(\hat{u}) = 0 \\
D^2 F(\hat{u}) = \beta \varphi''(|\hat{u}|) < 0
\end{cases}
\]

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

\[ \left\{ \hat{u} \in \mathbb{R}^p \mid \mathcal{F}(\hat{u}) = \inf_{u \in \mathbb{R}^p} \mathcal{F}(u) \right\} \neq \emptyset \]

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

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

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

\[ \Rightarrow \quad \hat{u} \text{ is the unique solution of the linear system} \]
\[ \begin{cases} a_i u = v[i] & \forall i \in \hat{I}_0 \\ G_j u = 0 & \forall j \in \hat{J}_0 \end{cases} \]

\[ \Rightarrow \quad (\star) \text{ the matrix with rows } (a_i, \forall i \in \hat{I}_0 \text{ and } G_j, \forall j \in \hat{J}_0) \text{ has full column rank} \]

\[ (\star) \text{ is a necessary condition for a (local) minimizer} \]

\[ \Rightarrow \quad \text{“contrast invariance” of (local) minimizers since } \hat{u} \text{ is linear in } v \]
Example

The data vector \( \nu \) is of length 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 with rows \((a_i, \forall i \in \widehat{I}_0 \text{ and } G_j, \forall j \in \widehat{J}_0)\) is of size 149 \(\times\) 80.

Its rank is 80.
• Let \( \hat{u} \in \mathbb{R}^p \) be a (local) minimizer of \( F \). Then

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

where \( G_j(k) \) is the \( k \)-th column of the linear operator \( G_j \)

• \( \implies \) each pixel of a (local) minimizer \( \hat{u} \) of \( F \) 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.
3. Numerical scheme

Continuation approach

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

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

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

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

where \( \Psi_\varepsilon(u) = \sum_{j \in J} \psi_\varepsilon(\|G_j u\|_2) . \)
For each $\varepsilon$ fixed—variable splitting and penalty decomposition techniques:

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

Alternate optimization:

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

Then

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

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

$u^{(k)}$ solves

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

where Quasi Newton method with preconditioning is used

$\Rightarrow$ fast algorithm
MR Image Reconstruction from Highly Undersampled Data

0-filling Fourier

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

$\| \cdot \|_1 + TV$

Our method

restored — original

Reconstructed images from 7% noisy randomly selected samples in the $k$-space.
MR Image Reconstruction from Highly Undersampled Data

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

Observed

\ell_1\text{-}TV

Our method, \( \varphi(t) = \frac{\alpha t}{\alpha t + 1} \)
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 $f_1$ 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.

Warmest wishes to Stan!

Thanks to Raymond for the invitation.