Sparsity Amplified

Ivan Selesnick

Electrical and Computer Engineering
Tandon School of Engineering
New York University
Brooklyn, New York

March 2017

Convex or non-convex: which is better?

(for sparse-regularized linear inverse problems)

Benefits of convex optimization:

1. Absence of suboptimal local minima
2. Continuity of solution as a function of input data
3. Algorithms guaranteed to converge to a global optimum
4. Regularization parameters easier to set

But convex regularization tends to **under-estimate** signal values.

Non-convex regularization often performs better!

Can we design non-convex sparsity-inducing penalties that maintain the convexity of the cost function to be minimized?

An alternative

Conventional sparse-regularized least squares:

$$J(x) = \frac{1}{2}\|y - Ax\|_2^2 + \lambda\|x\|_1$$

Alternatively, use NON-CONVEX penalty ψ that maintains CONVEXITY of cost function F

$$F(x) = \frac{1}{2}\|y - Ax\|_2^2 + \lambda\psi(x)$$

I. Selesnick. Total Variation Denoising via the Moreau Envelope. *IEEE Signal Processing Letters*, 24(2):216–220, February 2017.

I. W. Selesnick and I. Bayram. Enhanced Sparsity by Non-Separable Regularization. *IEEE Trans. Signal Process.*, 64(9):2298–2313, May 2016.

A. Parekh and I. W. Selesnick. Enhanced low-rank matrix approximation. *IEEE Signal Processing Letters*, 23(4):493–497, April 2016.

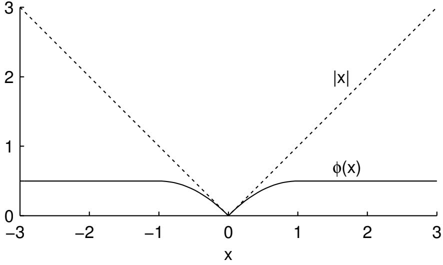
Y. Ding and I. W. Selesnick. Artifact-Free Wavelet Denoising: Non-convex Sparse Regularization, Convex Optimization. *IEEE Signal Processing Letters*, 22(9):1364–1368, September 2015.

I. W. Selesnick, A. Parekh, and I. Bayram. Convex 1-D Total Variation Denoising with Non-convex Regularization. *IEEE Signal Processing Letters*, 22(2):141–144, February 2015.

MC penalty

The minimax-concave (MC) penalty $\phi\colon \mathbb{R} \to \mathbb{R}$ is defined as

$$\phi(x) := \begin{cases} |x| - \frac{1}{2}x^2, & |x| \leqslant 1 \\ \frac{1}{2}, & |x| \geqslant 1 \end{cases}$$



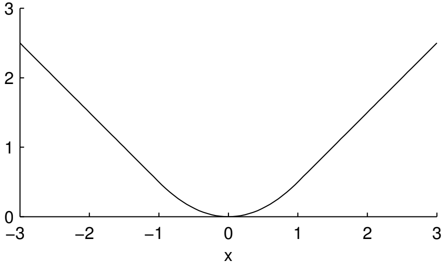
C.-H. Zhang. Nearly unbiased variable selection under minimax concave penalty. *The Annals of Statistics*:894–942, 2010.

Huber function

The MC penalty can be expressed as

$$\phi(x) = |x| - s(x)$$

where s is the Huber function.



The Huber function.

Huber function

The Huber function $s\colon \mathbb{R} \to \mathbb{R}$ is defined as

$$s(x) := \begin{cases} \frac{1}{2}x^2, & |x| \leqslant 1 \\ |x| - \frac{1}{2}, & |x| \geqslant 1. \end{cases}$$

The Huber function can be written as

$$s(x) = \min_{v \in \mathbb{R}} \{ |v| + \tfrac{1}{2}(x - v)^2 \},$$

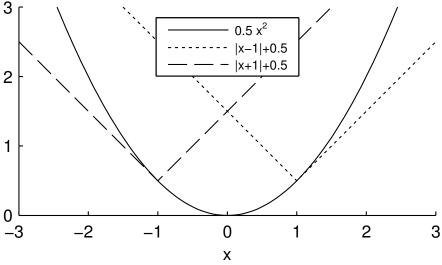
equivalently

$$s = |\cdot| \,\square\, \tfrac{1}{2}(\,\cdot\,)^2$$

where $\square$ denotes infimal convolution.

H. H. Bauschke and P. L. Combettes. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. Springer, 2011.

Huber function



The Huber function as the pointwise minimum of three functions.

Scaled functions

Let $b \in \mathbb{R}$. The scaled Huber function $s_b \colon \mathbb{R} \to \mathbb{R}$ is defined as

$$s_b(x) := s(b^2 x)/b^2, \quad b \neq 0.$$

For $b = 0$, the function is defined as

$$s_0(x) := 0.$$

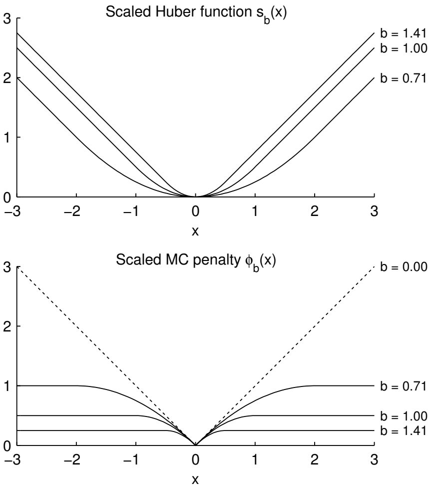
The scaled MC penalty function $\phi_b \colon \mathbb{R} \to \mathbb{R}$ is defined as

$$\phi_b(x) := |x| - s_b(x)$$

where s_b is the scaled Huber function.

Scaled functions

Scaled Huber function and MC penalty for several values of the scaling parameter.



Convexity condition

Let $\lambda > 0$ and $a \in \mathbb{R}$.

Define $f : \mathbb{R} \to \mathbb{R}$,

$$f(x) = \frac{1}{2}(y - ax)^2 + \lambda\, \phi_b(x)$$

where ϕ_b is the scaled MC penalty.

If

$$b^2 \leqslant a^2/\lambda,$$

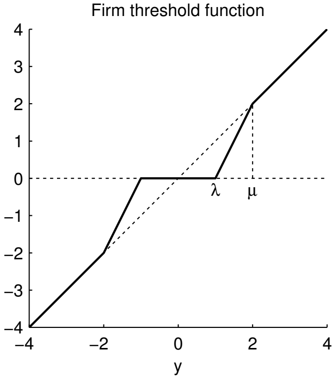
then

1. f is convex
2. the minimizer is given by firm thresholding,

$$x^{\mathrm{opt}} = \mathrm{firm}(y/a; \lambda/a^2, 1/b^2).$$

Firm threshold function

$$\mathsf{firm}(y; \lambda, \mu)$$



Aims

The minimizer of the scalar function f is easily obtained via firm thresholding.

However, the situation in the multivariate case is more complicated.

To generalize this process to the multivariate case, we aim to:

1. define a multivariate MC penalty (non-convex),
2. define a sparse-regularized least squares cost function (convex),
3. generalize the convexity condition,
4. provide a method to calculate a minimizer.

Generalized Huber function

Let $B \in \mathbb{R}^{M \times N}$.

We define the generalized Huber function $S_B \colon \mathbb{R}^N \to \mathbb{R}$ as

$$S_B(x) := \min_{v \in \mathbb{R}^N} \left\{ \|v\|_1 + \tfrac{1}{2} \|B(x - v)\|_2^2 \right\}.$$

In the notation of infimal convolution, we have

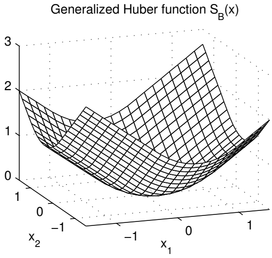
$$S_B = \| \cdot \|_1 \,\square\, \tfrac{1}{2} \|B \cdot \|_2^2.$$

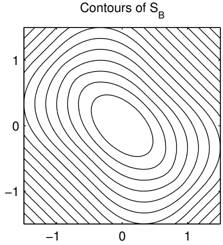
The generalized Huber function satisfies

$$0 \leqslant S_B(x) \leqslant \|x\|_1, \quad \forall x \in \mathbb{R}^N.$$

Generalized Huber function

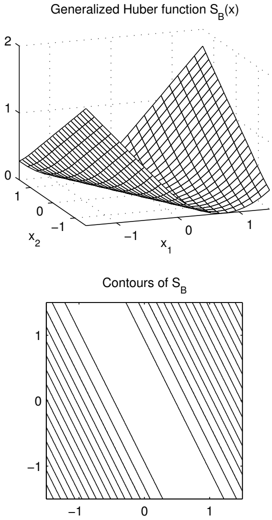
$$B = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 1 \end{bmatrix}$$





Generalized Huber function

$B = [1 \;\; 0.5]$



Generalized MC penalty

Let $B \in \mathbb{R}^{M \times N}$.

We define the generalized MC (GMC) penalty function $\psi_B \colon \mathbb{R}^N \to \mathbb{R}$ as

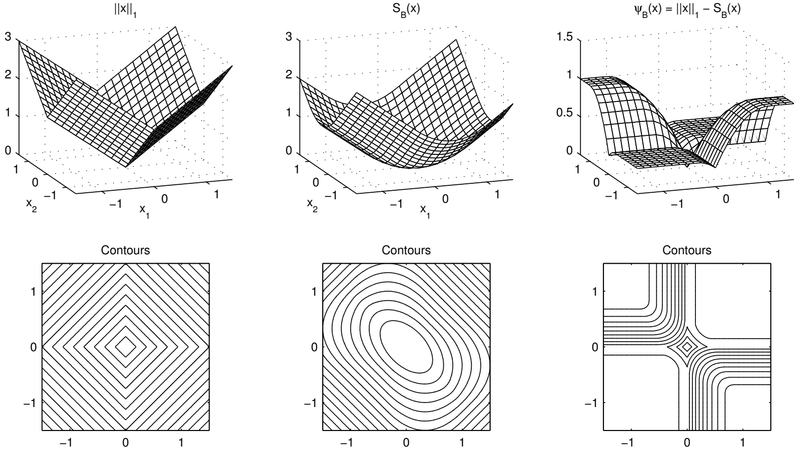
$$\psi_B(x) := \|x\|_1 - S_B(x)$$

where S_B is the generalized Huber function.

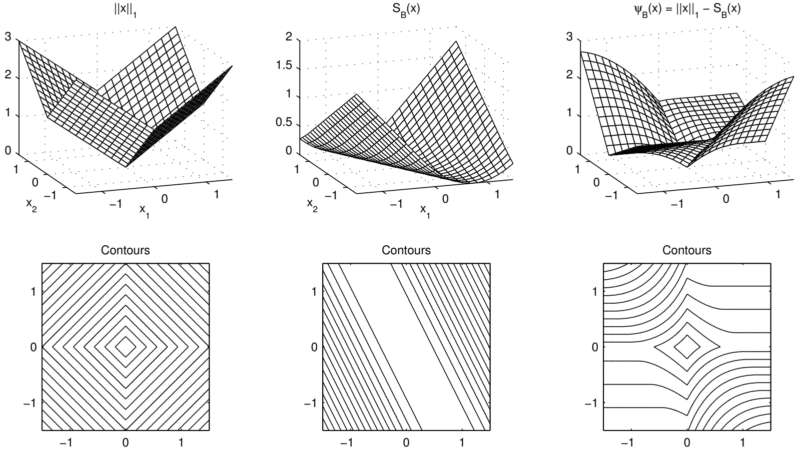
The generalized MC penalty satisfies

$$0 \leqslant \psi_B(x) \leqslant \|x\|_1 \quad \text{for all } x \in \mathbb{R}^N.$$

Generalized MC penalty



Generalized MC penalty



Convexity condition

Let $y \in \mathbb{R}^M$, $A \in \mathbb{R}^{M \times N}$, and $\lambda > 0$.

Define $F \colon \mathbb{R}^N \to \mathbb{R}$ as

$$F(x) = \frac{1}{2}\|y - Ax\|_2^2 + \lambda\, \psi_B(x)$$

where ψ_B is the generalized MC penalty.

If

$$B^{\mathsf{T}} B \preccurlyeq \frac{1}{\lambda} A^{\mathsf{T}} A,$$

then F is a convex function.

Hence, for convexity of F, we may simply set

$$B = \sqrt{\gamma/\lambda}\, A, \quad 0 \leqslant \gamma \leqslant 1. \tag{1}$$

Proximal algorithm

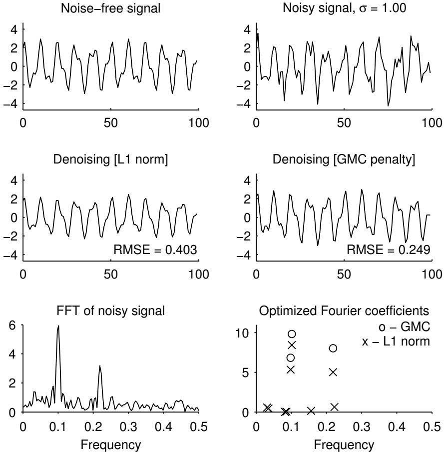
Using forward-backward splitting, we have:

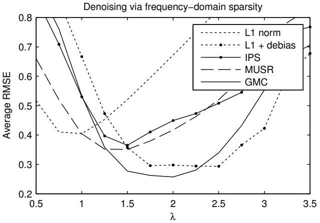
Let $\lambda > 0$ and $0 \leqslant \gamma < 1$. Let $y \in \mathbb{R}^N$ and $A \in \mathbb{R}^{M \times N}$. Then a saddle-point $(x^{\text{opt}}, v^{\text{opt}})$ of F can be obtained by the iterative algorithm:

$$
\begin{aligned}
&\text{Set } \rho = \max\{1, \gamma/(1-\gamma)\} \, \|A^{\mathsf{T}} A\|_2 \\
&\text{Set } \mu : 0 < \mu < 2/\rho \\
&\text{For } i = 0, 1, 2, \dots \\
&\quad w^{(i)} = x^{(i)} - \mu A^{\mathsf{T}} \big(A(x^{(i)} + \gamma(v^{(i)} - x^{(i)})) - y \big) \\
&\quad u^{(i)} = v^{(i)} - \mu \gamma A^{\mathsf{T}} A (v^{(i)} - x^{(i)}) \\
&\quad x^{(i+1)} = \operatorname{soft}(w^{(i)}, \mu\lambda) \\
&\quad v^{(i+1)} = \operatorname{soft}(u^{(i)}, \mu\lambda) \\
&\text{end}
\end{aligned}
$$

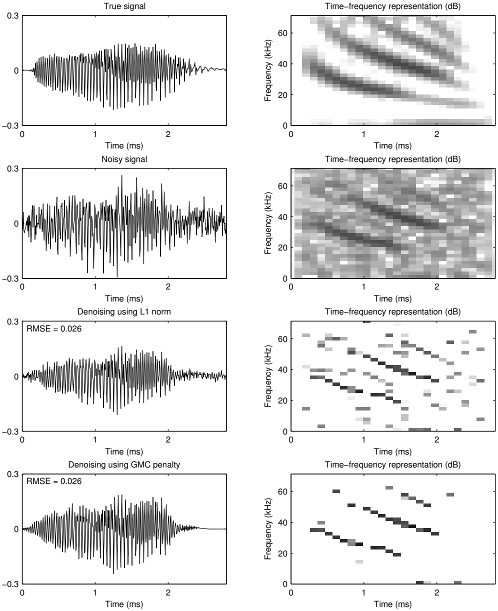
where i is the iteration counter.

Example: frequency-domain sparsity

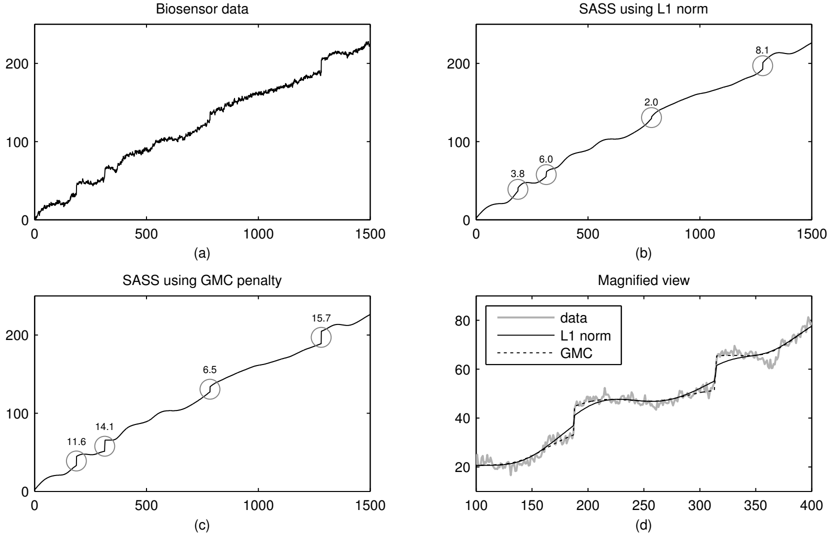




Example: time-frequency domain sparsity



Example: sparsity-assisted signal smoothing (SASS)



Conclusion

Convex and the non-convex methods for sparse-regularized least squares are usually mutually exclusive and incompatible.

To bridge these two approaches, we introduce a non-convex alternative to the L1 norm that preserves the convexity of the cost function to be minimized.

The proposed penalty leads to optimization problems with no extraneous suboptimal local minima and allows the use of globally convergent, computationally efficient, scalable convex optimization algorithms.

The advantages compared to L1 norm regularization are
(i) more accurate estimation of high-amplitude components of sparse solutions,
(ii) a higher level of sparsity in a sparse approximation problem.