Sparse Regularization via Convex Analysis

Ivan Selesnick

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

2019

Convex or non-convex: Which is better for inverse problems?

Benefits of convex optimization:

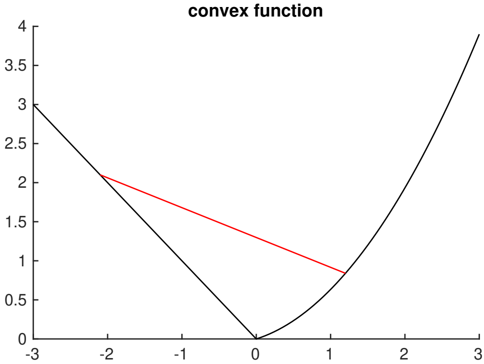
1. Absence of suboptimal local minima
2. Continuity of solution as a function of input data
3. Algorithms guaranteed to converge to a global optimum
4. Regularization parameters easier to set

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

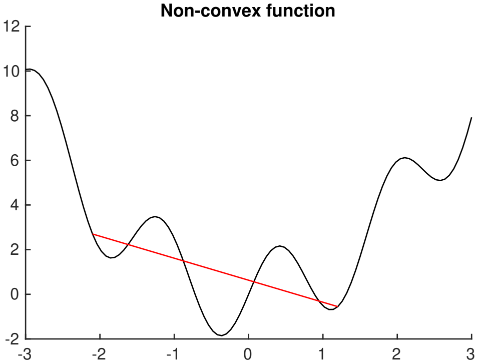
Non-convex regularization often performs better!

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

Convex function



Non-Convex function



Goal

Goal: Find a sparse approximate solution to a linear system $y = Ax$.

Minimize a cost function:

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

or

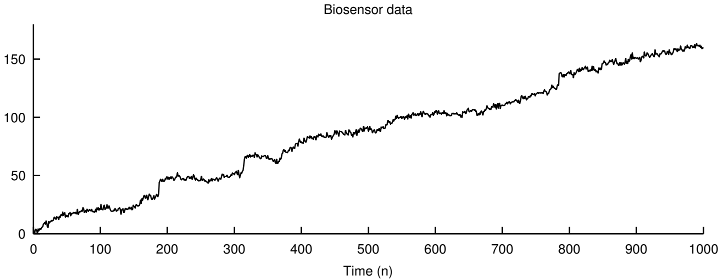
$$F(x) = \frac{1}{2}\|y - Ax\|_2^2 + \lambda\,\psi(x)$$

Question: How to define ψ ?

Let us allow ψ to be non-convex such that F is convex.

This is the *Convex Non-Convex* (CNC) approach.

Biosensor Signal



Linear Filter

Given noisy data $y \in \mathbb{R}^N$, perform smoothing via:

$$\hat{x} = \arg\min_{x \in \mathbb{R}^N} \left\{ \sum_{n=0}^{N-1} |y(n) - x(n)|^2 + \lambda \sum_{n=1}^{N-1} |x(n) - x(n-1)|^2 \right\}$$

which can be written

$$\hat{x} = \arg\min_{x \in \mathbb{R}^N} \left\{ \|y - x\|_2^2 + \lambda \|Dx\|_2^2 \right\}$$

where

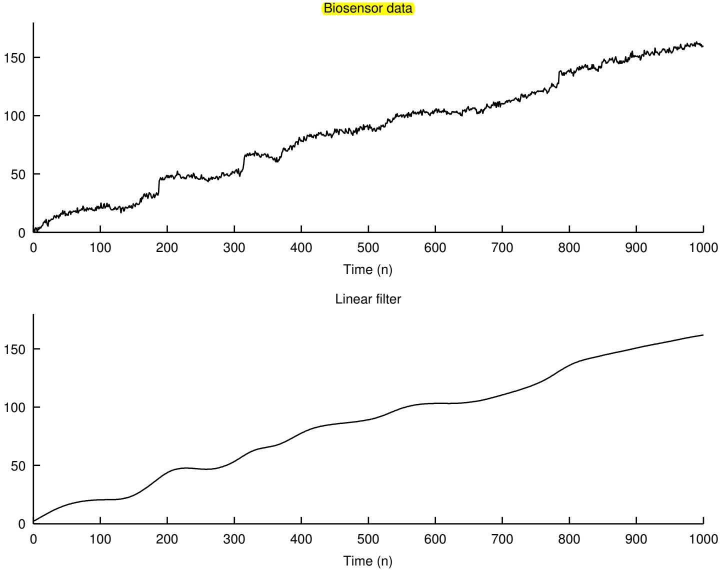
$$\|x\|_2^2 := \sum_n |x(n)|^2$$

$$D = \begin{bmatrix} -1 & 1 & & & \\ & -1 & 1 & & \\ & & \ddots & \ddots & \\ & & & -1 & 1 \end{bmatrix}$$

Solution:

$$\hat{x} = (I + \lambda D^{\mathsf{T}} D)^{-1} y$$

Biosensor Signal



Total Variation Denoising (Nonlinear Filter)

Given noisy data $y \in \mathbb{R}^N$, perform smoothing via:

$$\hat{x} = \arg\min_{x \in \mathbb{R}^N} \left\{ \frac{1}{2} \sum_{n=0}^{N-1} |y(n) - x(n)|^2 + \lambda \sum_{n=1}^{N-1} |x(n) - x(n-1)| \right\}$$

which can be written

$$\hat{x} = \arg\min_{x \in \mathbb{R}^N} \left\{ \frac{1}{2} \|y - x\|_2^2 + \lambda \|Dx\|_1 \right\}$$

where

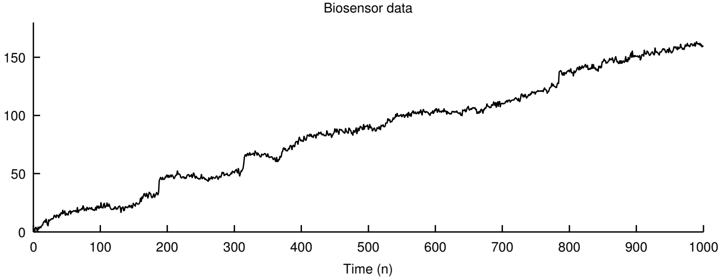
$$\|x\|_2^2 := \sum_n |x(n)|^2, \quad \|x\|_1 := \sum_n |x(n)|$$

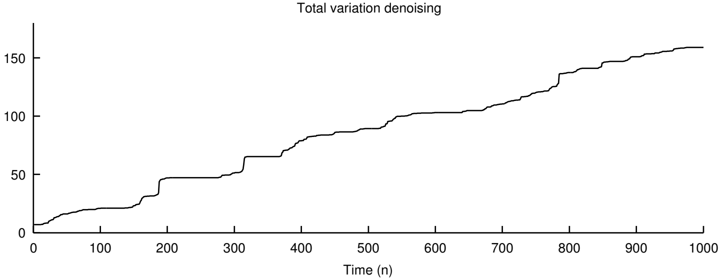
$$D = \begin{bmatrix} -1 & 1 & & & \\ & -1 & 1 & & \\ & & \ddots & \ddots & \\ & & & -1 & 1 \end{bmatrix}$$

Solution? *No closed form solution.*

Use iterative algorithm ... but note the cost function is not differentiable!

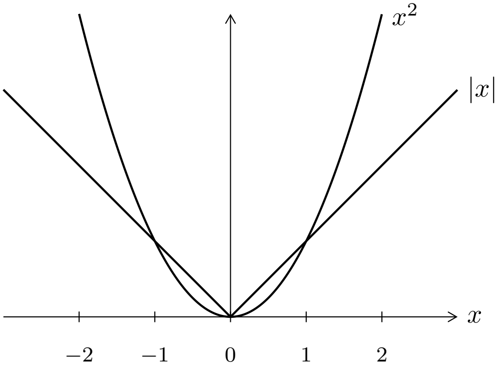
Biosensor Signal





Two Penalties

The ℓ_1 norm induces sparsity unlike the the sum of squares.



Combine Quadratic and Sparse Regularization

$$\arg \min_{u,v \in \mathbb{R}^N} \left\{ \frac{1}{2} \|y - u - v\|_2^2 + \lambda_1 \|Du\|_1 + \frac{\lambda_2}{2} \|Dv\|_2^2 \right\}$$

$$\hat{x} = u + v$$

Combine Quadratic and Sparse Regularization

$$\arg\min_{u,v\in\mathbb{R}^N}\left\{\frac{1}{2}\|y-u-v\|_2^2+\lambda_1\|Du\|_1+\frac{\lambda_2}{2}\|Dv\|_2^2\right\}$$

Solving for v gives

$$v=(I+\lambda_2 D^{\mathsf{T}}D)^{-1}(y-u)$$

$$x=v+u=(I+\lambda_2 D^{\mathsf{T}}D)^{-1}(y+\lambda_2 D^{\mathsf{T}}Du)$$

Substituting v back in to the cost function:

$$J(u)=\frac{\lambda_2}{2}(y-u)^{\mathsf{T}}D^{\mathsf{T}}(I+\lambda_2 DD^{\mathsf{T}})^{-1}D(y-u)+\lambda_1\|Du\|_1$$

or

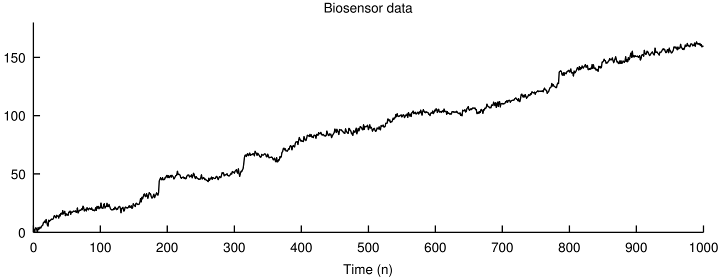
$$J(u)=\frac{\lambda_2}{2}\|R^{-1}D(y-u)\|_2^2+\lambda_1\|Du\|_1$$

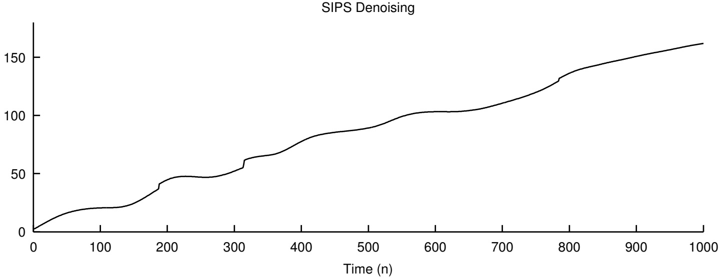
$$RR^{\mathsf{T}}=I+\lambda_2 DD^{\mathsf{T}}\qquad (R\text{ is a banded matrix})$$

Since x depends on Du, not u directly, define $g=Du$. So we need to minimize

$$F(g)=\frac{\lambda_2}{2}\|R^{-1}Dy-R^{-1}g\|_2^2+\lambda_1\|g\|_1$$

Biosensor Signal

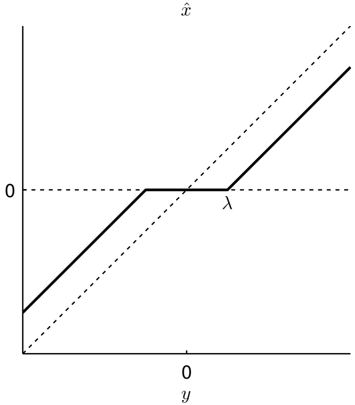




Scalar case

$$\hat{x} = \arg\min_{x} \left\{ \frac{1}{2}(y - x)^2 + \lambda |x| \right\}$$

$\Longrightarrow$



Non-convex scalar penalty functions (alternatives to ℓ_1 norm)

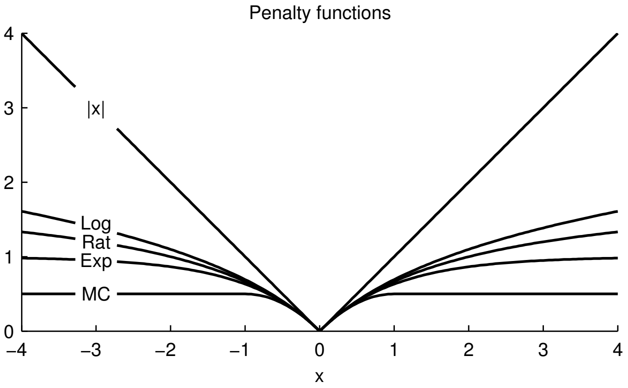
Log	$\phi_a(x) = \dfrac{1}{a}\log(1 + a\|x\|)$
Rat	$\phi_a(x) = \dfrac{\|x\|}{1 + a\|x\|/2}$
Exp	$\phi_a(x) = \dfrac{1}{a}\left(1 - \mathrm{e}^{-a\|x\|}\right)$
MC	$\phi_a(x) = \begin{cases} \|x\| - \frac{a}{2}x^2, & \|x\| \leqslant 1/a \\ \frac{1}{2a}, & \|x\| \geqslant 1/a \end{cases}$

The penalties are parameterized such that

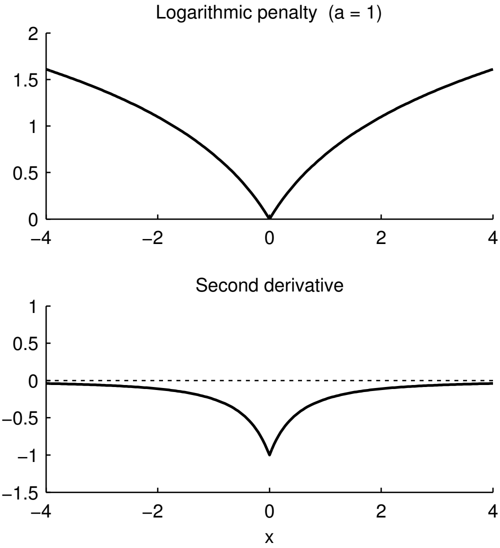
$$\phi_a'(0^+) = 1$$
$$\phi_a''(0^+) = -a.$$

Non-convex scalar penalty functions

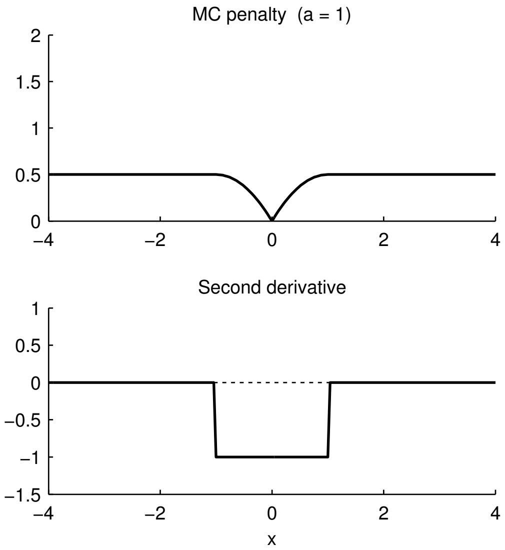
Penalty functions with $a = 1$.



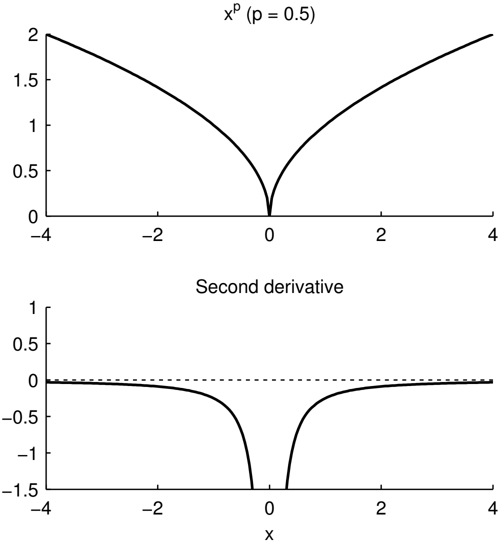
Logarithmic penalty



MC penalty


MC penalty (a = 1)
2
1.5
1
0.5
0
–4
–2
0
2
4
Second derivative
1
0.5
0
–0.5
–1
–1.5
–4
–2
0
2
4
x

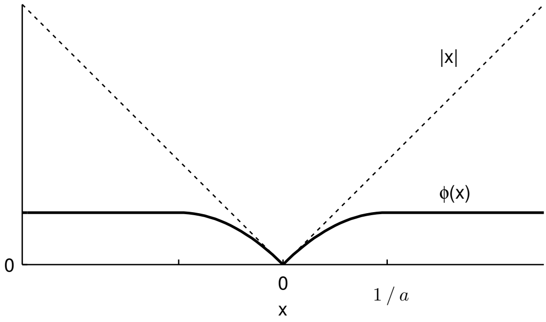
ℓ_p penalty, $0 < p < 1$ *(precluded)*



Scalar MC penalty

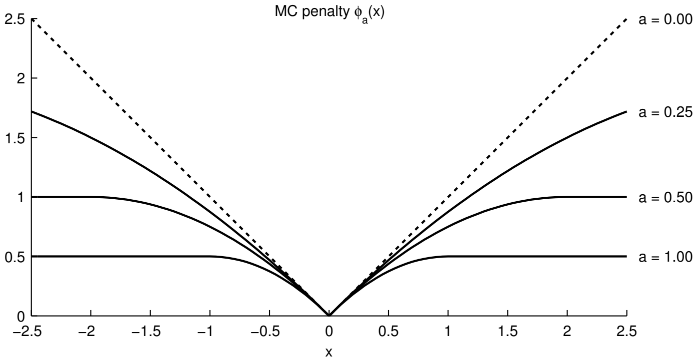
We consider henceforth only the minimax-concave (MC) penalty function

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



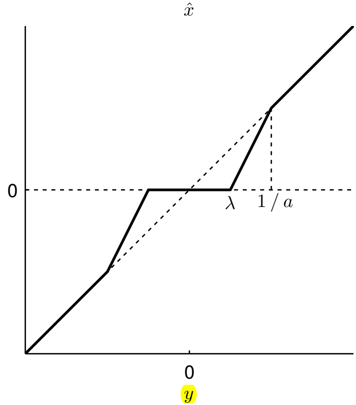
Scalar MC penalty

The parameter $a \geqslant 0$ controls the non-convexity of ϕ_a.



Scalar case using MC penalty

$$\hat{x} = \arg\min_{x} \left\{ \frac{1}{2}(y - x)^2 + \lambda\phi_a(x) \right\}$$



$\hat{x}$ is a continuous function of y when $a < \lambda$.

Scalar case using MC penalty

Consider

$$f(x) = \frac{1}{2}(y - x)^2 + \lambda\phi_a(x).$$

For what values 'a' is f a convex function?

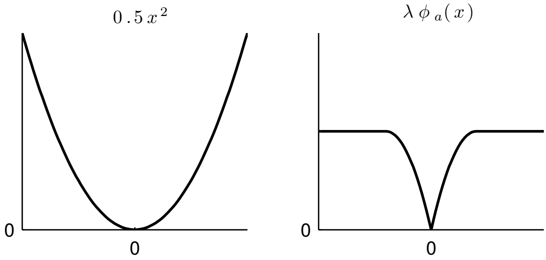
$$f(x) = \underbrace{\frac{1}{2}x^2 + \lambda\phi_a(x)}_{f_0(x)} + \underbrace{\left[\frac{1}{2}y^2 - yx\right]}_{\text{convex in } x}.$$

It is sufficient to consider the convexity of

$$f_0(x) = \frac{1}{2}x^2 + \lambda\phi_a(x).$$

Scalar case using MC penalty

$$f_0(x) = \frac{1}{2}x^2 + \lambda\phi_a(x)$$



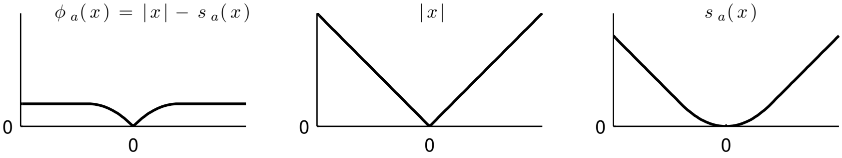
Is f_0 convex?

ϕ_a is *not differentiable*. So we can not simply check that the second derivative of f_0 is positive . . .

Scalar case using MC penalty

Let us write

$$\phi_a(x) = |x| - s_a(x)$$



We see the Huber function:

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

Scalar case using MC penalty

Writing ϕ_a as

$$\phi_a(x) = |x| - s_a(x),$$

we have

$$\begin{aligned} f(x) &= \frac{1}{2}(y-x)^2 + \lambda\phi_a(x) \\ &= \frac{1}{2}(y-x)^2 + \lambda\Big[|x| - s_a(x)\Big] \\ &= \underbrace{\frac{1}{2}x^2 - \lambda s_a(x)}_{g(x)} + \underbrace{\Big[\lambda|x| + \frac{1}{2}y^2 - yx\Big]}_{\text{convex in } x} \end{aligned}$$

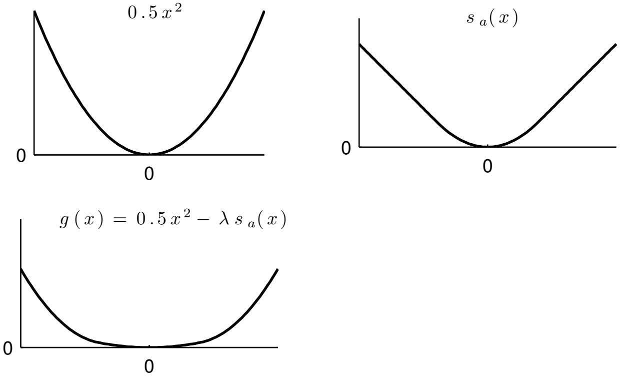
g convex $\implies$ f convex

Note that g is *differentiable* unlike f.

Is g *convex*? It depends on a and λ.

Scalar case using MC penalty

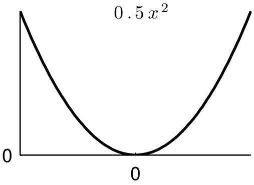
$$g(x) = \frac{1}{2}x^2 - \lambda s_a(x)$$

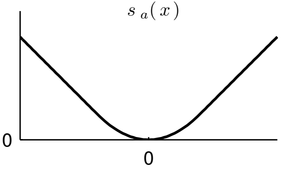


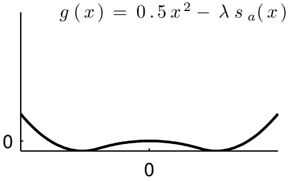
$a = 1$, $\lambda = 0.8$, g *is* convex.

Scalar case using MC penalty

$$g(x) = \frac{1}{2}x^2 - \lambda s_a(x)$$



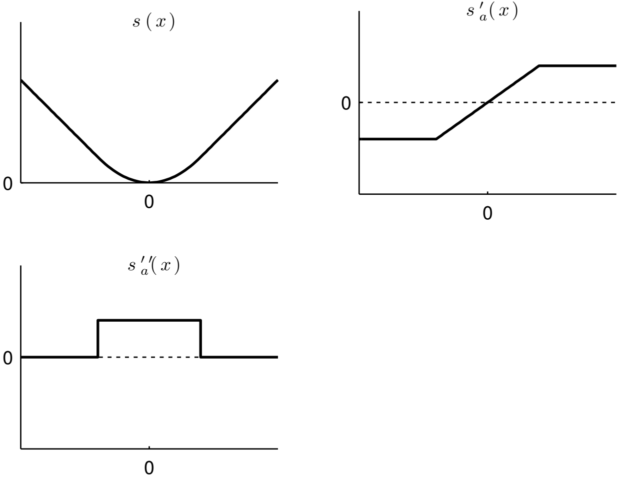




$a = 1$, $\lambda = 1.3$, g *is not* convex.

Scalar case using MC penalty

The Huber function is differentiable. But not twice differentiable.



Scalar case using MC penalty

$$g(x) = \frac{1}{2}x^2 - \lambda s_a(x)$$

When is g convex?

We can not check the second derivative of g because it is not twice differentiable (see previous page).

How can we ensure g (and hence f) is *convex*?

Huber function as an infimal convolution

The Huber function can be written as

$$s_a(x) = \min_v \left\{ \frac{a}{2}(x - v)^2 + |v| \right\}.$$

As infimal convolution

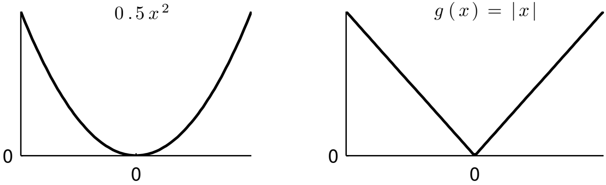
$$s_a(x) = \left(\frac{a}{2}(\,\cdot\,)^2 \,\square\, |\cdot| \right)(x)$$

where infimal convolution (Moreau-Yosida regularization) is defined as

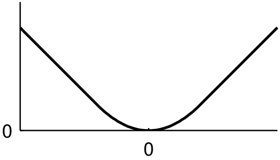
$$(f \,\square\, g)(x) := \min_v \{ f(v) + g(x - v) \}$$

Huber function as an infimal convolution

$$\left(\frac{a}{2}(\cdot)^2 \,\square\, |\cdot|\right)(x)$$

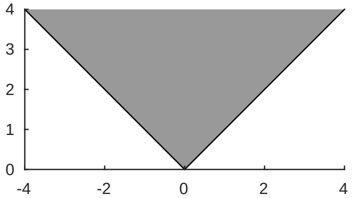


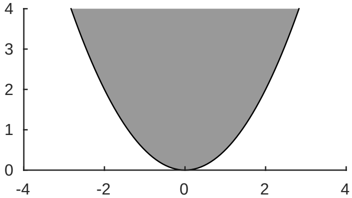
Huber function



Epigraph

The epigraph of a function is a set comprising points on and above the graph.





The epigraph of

$$\mathrm{epi}\{f \,\square\, g\} = \mathrm{epi}\{f\} + \mathrm{epi}\{g\}$$

Scalar case using MC penalty

The Huber function can be written as

$$s_a(x) = \min_v \left\{ \frac{a}{2}(x - v)^2 + |v| \right\}.$$

When is g convex?

$$\begin{aligned}
g(x) &= \frac{1}{2}x^2 - \lambda s_a(x) \\
&= \frac{1}{2}x^2 - \lambda \min_v \left\{ \frac{a}{2}(x - v)^2 + |v| \right\} \\
&= \frac{1}{2}x^2 - \lambda \min_v \left\{ \frac{a}{2}(x^2 - 2xv + v^2) + |v| \right\} \\
&= \frac{1}{2}x^2 - \lambda \frac{a}{2}x^2 + \underbrace{\lambda \max_v \Big\{ \underbrace{\frac{a}{2}(2xv - v^2) - |v|}_{\text{affine in } x} \Big\}}_{\text{convex in } x} \\
&= \frac{1}{2}(1 - a\lambda)\, x^2 + \text{convex function}
\end{aligned}$$

The function g is *convex* if $1 - a\lambda$ is non-negative, i.e.,

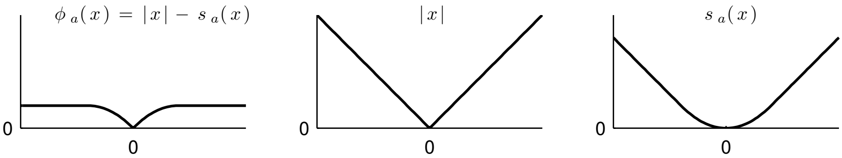
$$\boxed{a \leqslant 1/\lambda}$$

We do *not* need derivatives!

Scalar MC penalty

The MC penalty can be written as

$$\begin{aligned}\phi_a(x) &= |x| - s_a(x)\\ &= |x| - \underbrace{\min_v \left\{ \frac{a}{2}(x-v)^2 + |v| \right\}}_{\text{Huber function}}\\ &= |x| - \left(\frac{a}{2}(\,\cdot\,)^2 \,\Box\, |\cdot| \right)(x)\end{aligned}$$



Multivariate case

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

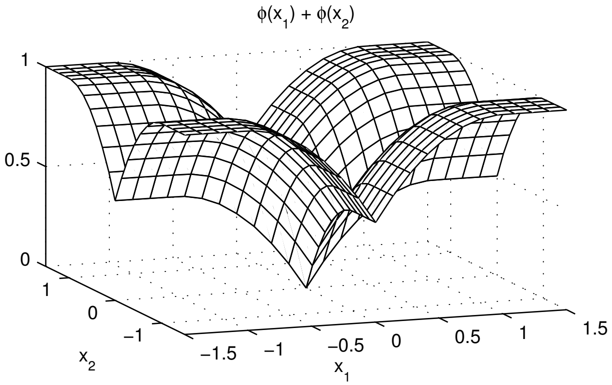
How can we set ψ so that F is convex and promotes sparsity of x?

We can generalize the scalar case . . .

Separable penalty *(precluded)*

Conventional penalty: additive (separable)

$$\phi(x) = \phi(x_1) + \phi(x_2)$$



Generalized Huber function

Let $B \in \mathbb{R}^{M \times N}$. We define the *generalized Huber function*

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

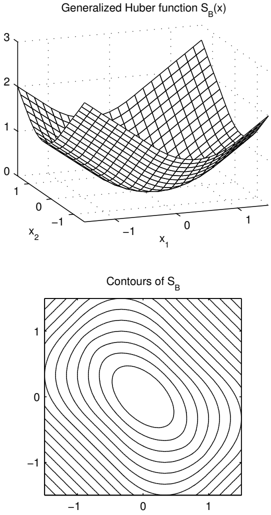
In the notation of infimal convolution, we have

$$S_B(x) := \left(\frac{1}{2} \|B \cdot\|_2^2 \,\Box\, \| \cdot \|_1 \right)(x).$$

Example 1. Generalized Huber function

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

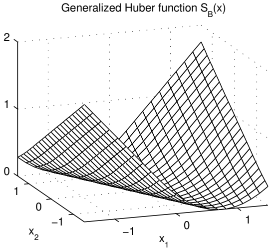
$$S_B(x) := \min_v \left\{ \tfrac{1}{2} \|B(x-v)\|_2^2 + \|v\|_1 \right\}$$

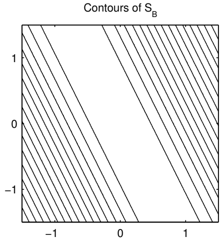


Example 2. Generalized Huber function

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

$$S_B(x) := \min_v \left\{ \tfrac{1}{2}\|B(x-v)\|_2^2 + \|v\|_1 \right\}$$



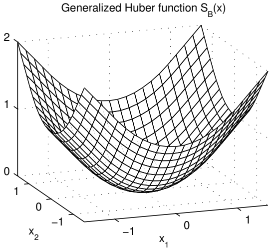


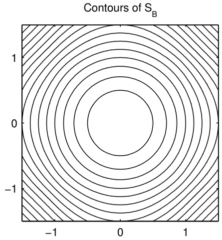
Example 3. Generalized Huber function

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

$$S_B(x) := \min_v \left\{ \tfrac{1}{2} \|B(x - v)\|_2^2 + \|v\|_1 \right\}$$

If B is diagonal, then S_B is separable.





Generalized Huber function

The generalized Huber function is differentiable.

Its gradient is given by

$$\nabla S_B(x) = B^{\mathsf{T}}B\Big(x - \arg\min_{v \in \mathbb{R}^N} \Big\{\frac{1}{2}\|B(x-v)\|_2^2 + \|v\|_1\Big\}\Big).$$

Neither the generalized Huber function nor its gradient have simple closed form expressions. But we will still be able to use them . . .

When $B = I$ we recover a well known identity

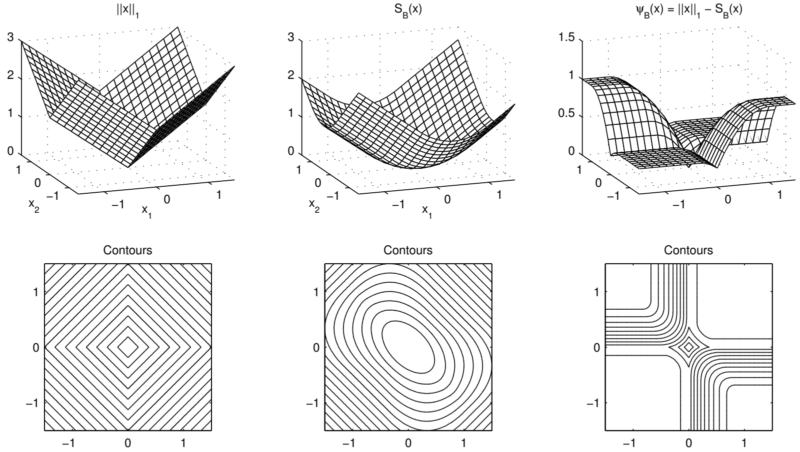
$$\nabla S_I(x) = x - \arg\min_{v \in \mathbb{R}^N} \Big\{\frac{1}{2}\|x-v\|_2^2 + \|v\|_1\Big\}$$

Generalized MC penalty

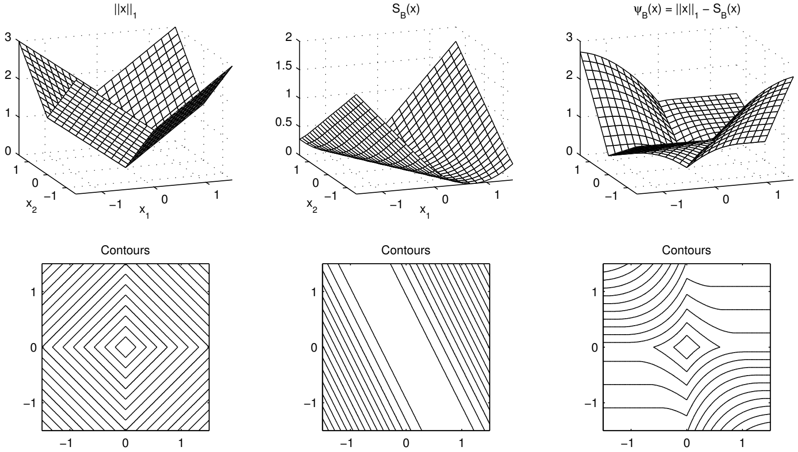
We define the *generalized MC (GMC)* penalty

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

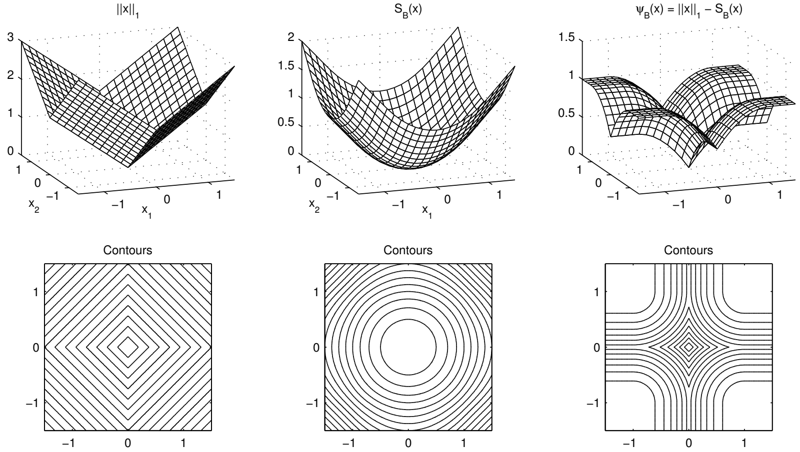
Example 1. Generalized MC penalty



Example 2. Generalized MC penalty



Example 3. Generalized MC penalty



Convexity

Theorem. The function

$$
\begin{aligned}
F(x) &= \frac{1}{2}\|y - Ax\|_2^2 + \lambda\,\psi_B(x) \\
&= \frac{1}{2}\|y - Ax\|_2^2 + \lambda\big[\|x\|_1 - S_B(x)\big] \\
&= \frac{1}{2}\|y - Ax\|_2^2 - \lambda\|x\|_1 - \lambda \min_{v \in \mathbb{R}^N} \big\{ \tfrac{1}{2}\|B(x - v)\|_2^2 + \|v\|_1 \big\}.
\end{aligned}
$$

is *convex* if

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

even when ψ_B is *non-convex*.

Convexity condition – proof

Write F as

$$\begin{aligned} F(x) &= \tfrac{1}{2}\|y - Ax\|_2^2 + \lambda\,\psi_B(x) \\ &= \tfrac{1}{2}\|y - Ax\|_2^2 + \lambda\big[\|x\|_1 - S_B(x)\big] \\ &= \Big[\tfrac{1}{2}\|Ax\|_2^2 + \tfrac{1}{2}\|y\|_2^2 - y^{\mathsf{T}}Ax\Big] + \lambda\|x\|_1 - \lambda S_B(x) \\ &= \underbrace{\tfrac{1}{2}\|Ax\|_2^2 - \lambda S_B(x)}_{G(x)} + \underbrace{\tfrac{1}{2}\|y\|_2^2 - y^{\mathsf{T}}Ax + \lambda\|x\|_1}_{\text{convex in } x} \end{aligned}$$

G convex $\implies$ F convex

Convexity condition – proof

Write G as

$$
\begin{aligned}
G(x) &= \tfrac{1}{2}\|Ax\|_2^2 - \lambda S_B(x) \\
&= \tfrac{1}{2}\|Ax\|_2^2 - \lambda \min_v \left\{ \tfrac{1}{2}\|B(x-v)\|_2^2 + \|v\|_1 \right\} \\
&= \tfrac{1}{2}\|Ax\|_2^2 - \lambda \min_v \left\{ \tfrac{1}{2}\|Bx\|_2^2 + \tfrac{1}{2}\|Bv\|_2^2 - v^{\mathsf{T}} B^{\mathsf{T}} Bx + \|v\|_1 \right\} \\
&= \tfrac{1}{2}\|Ax\|_2^2 - \tfrac{\lambda}{2}\|Bx\|_2^2 + \underbrace{\lambda \max_v \Big\{ \underbrace{v^{\mathsf{T}} B^{\mathsf{T}} Bx - \tfrac{1}{2}\|Bv\|_2^2 - \|v\|_1}_{\text{affine in } x} \Big\}}_{\text{convex in } x} \\
&= \tfrac{1}{2} x^{\mathsf{T}} \left(A^{\mathsf{T}} A - \lambda B^{\mathsf{T}} B \right) x + \text{ convex function}
\end{aligned}
$$

The function G (hence F) is *convex* if $A^{\mathsf{T}}A - \lambda B^{\mathsf{T}}B$ is positive semidefinite, i.e.

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

Convexity condition

A straightforward choice of B to satisfy

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

is

$$B = \sqrt{\frac{\gamma}{\lambda}}A$$

for some γ with $0 \leqslant \gamma \leqslant 1$.

Optimization Algorithm

We can use the Forward-Backward Splitting (FBS) algorithm

$$\begin{aligned}
F(x) &= \tfrac{1}{2}\|y - Ax\|_2^2 + \lambda\,\psi_B(x) \\
&= \tfrac{1}{2}\|y - Ax\|_2^2 + \lambda\|x\|_1 - \lambda S_B(x) \\
&= \underbrace{\tfrac{1}{2}\|y - Ax\|_2^2 - \lambda S_B(x)}_{\text{convex, differentiable}} + \underbrace{\lambda\|x\|_1}_{\text{convex}} \\
&= f_1(x) + f_2(x)
\end{aligned}$$

The FBS algorithm is given by:

$$\begin{aligned}
w^{(i)} &= x^{(i)} - \mu\left[\nabla f_1(x^{(i)})\right] \\
x^{(i+1)} &= \arg\min_x \left\{\tfrac{1}{2}\|w^{(i)} - x\|_2^2 + \mu f_2(x)\right\} \\
&= \operatorname{prox}_{\mu f_2}(w^{(i)})
\end{aligned}$$

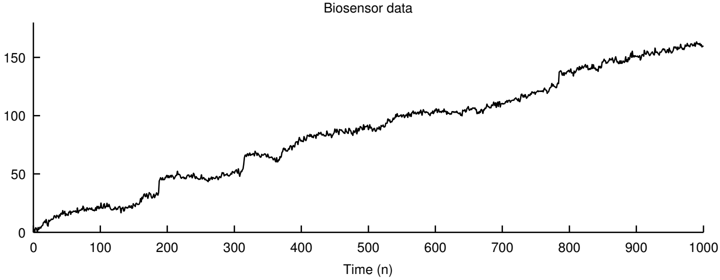
Optimization Algorithm

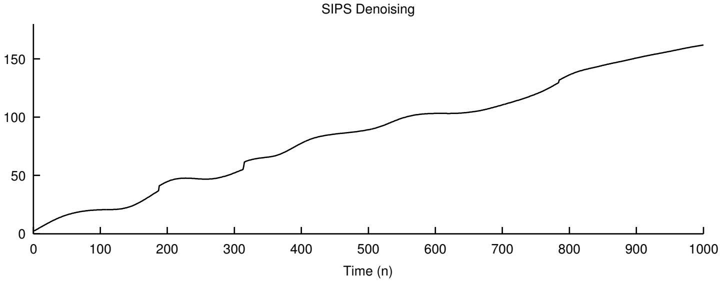
Let $0 \leqslant \gamma \leqslant 1$ and $B = \sqrt{\gamma/\lambda}\, A$. Then a minimizer of the objective function F is obtained by the iteration:

$$
\begin{aligned}
v^{(i)} &= \arg\min_{v \in \mathbb{R}^N} \left\{ \frac{\gamma}{2} \|A(x^{(i)} - v)\|_2^2 + \lambda \|v\|_1 \right\} \\
z^{(i)} &= \gamma A(x^{(i)} - v^{(i)}) \\
x^{(i+1)} &= \arg\min_{x \in \mathbb{R}^N} \left\{ \frac{1}{2} \|y + z^{(i)} - Ax\|_2^2 + \lambda \|x\|_1 \right\}
\end{aligned}
$$

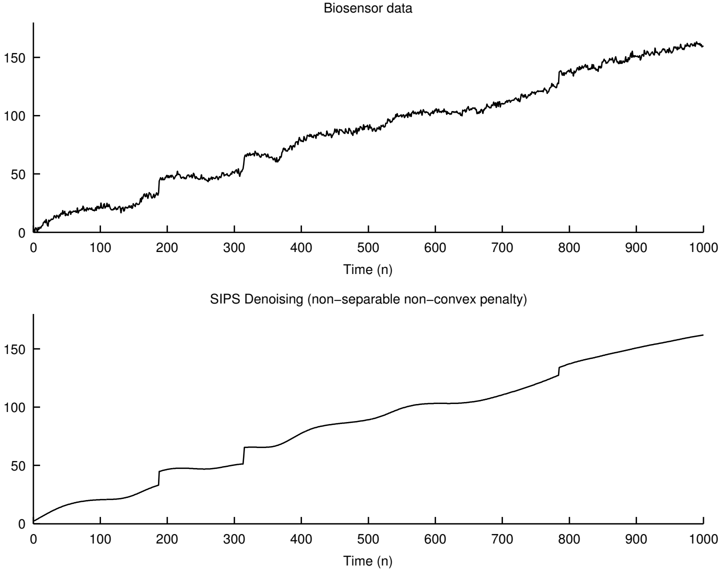
Interpretation: iteratively adjusted additive data perturbation of ℓ_1 norm regularized problem ...

Biosensor Signal

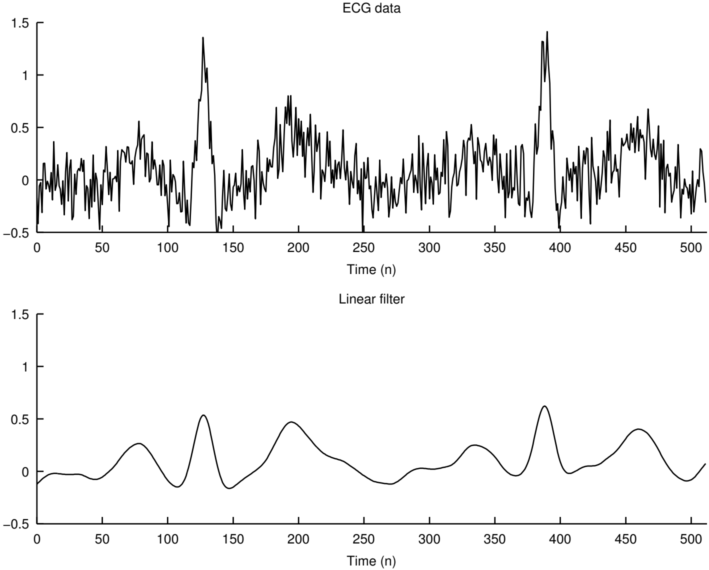




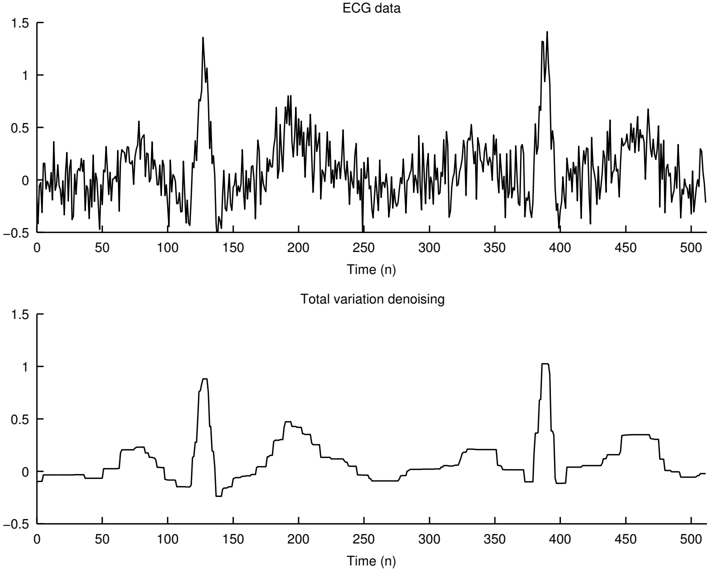
Biosensor Signal



ECG Signal

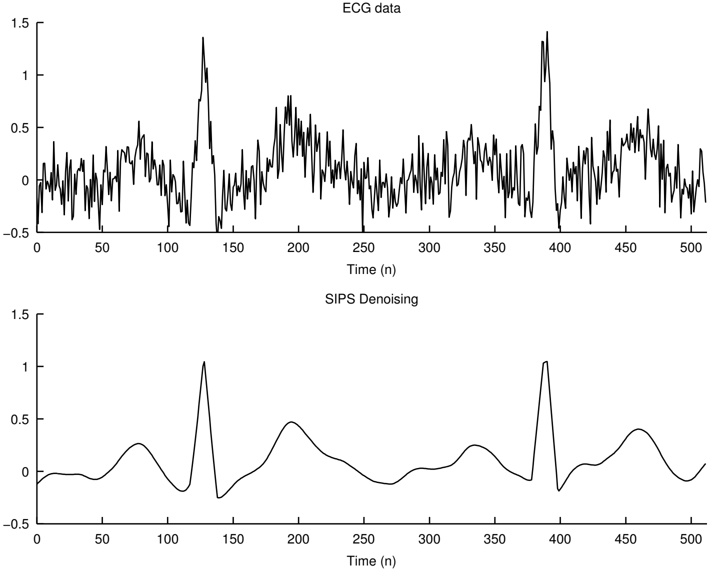

ECG data
1.5
1
0.5
0
−0.5
0 50 100 150 200 250 300 350 400 450 500
Time (n)
Linear filter
1.5
1
0.5
0
−0.5
0 50 100 150 200 250 300 350 400 450 500
Time (n)

ECG Signal

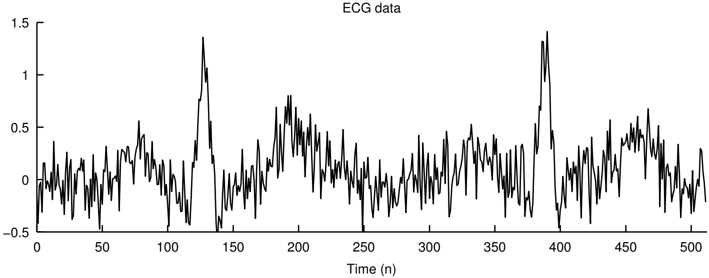


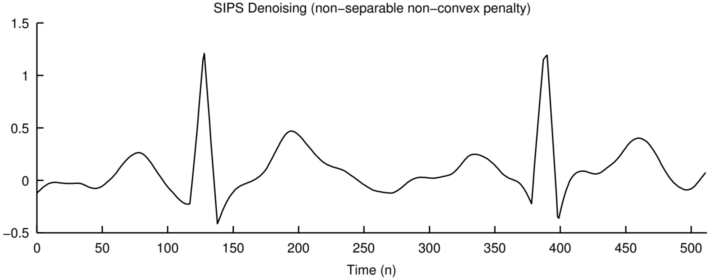
ECG data
1.5
1
0.5
0
−0.5
0
50
100
150
200
250
300
350
400
450
500
Time (n)
Total variation denoising
1.5
1
0.5
0
−0.5
0
50
100
150
200
250
300
350
400
450
500
Time (n)

ECG Signal


ECG data
1.5
1
0.5
0
−0.5
0
50
100
150
200
250
300
350
400
450
500
Time (n)
SIPS Denoising
1.5
1
0.5
0
−0.5
0
50
100
150
200
250
300
350
400
450
500
Time (n)

ECG Signal


ECG data
1.5
1
0.5
0
−0.5
0
50
100
150
200
250
300
350
400
450
500
Time (n)


SIPS Denoising (non−separable non−convex penalty)
1.5
1
0.5
0
−0.5
0
50
100
150
200
250
300
350
400
450
500
Time (n)

Generalizations (TV, nuclear norm, etc)

Consider the function

$$\mathcal{J}(x) := \frac{1}{2}\|Ax - b\|_2^2 + \lambda\,\mathcal{R}(x) \qquad (1)$$

where $\mathcal{R}$ is a convex regularizer of the form

$$\mathcal{R}(x) = \Phi(G(Lx))$$

where L is a linear operator, G is possibly nonlinear, and Φ promotes sparsity.

For example, total variation regularization is expressed as

$$\mathcal{R}(x) = \|G(Lx)\|_1 = \sum_i |g_i(Lx)|$$

with

$$g_i(Lx) = \begin{cases} \left(\begin{bmatrix} D_h \\ D_v \end{bmatrix} x\right)_i & \text{anisotropic TV} \\ \sqrt{(D_h x)_i^2 + (D_v x)_i^2} & \text{isotropic TV} \end{cases}$$

Generalizations (TV, nuclear norm, etc)

Suppose $\mathcal{R}$ satisfies

A1) $\mathcal{R}(\cdot) = \Phi(G(L\,\cdot))$ is convex and bounded from below by zero;

A2) $\Phi(G(\cdot))$ is a proper, lower semicontinuous and coercive function.

Consider the function

$$\mathcal{J}_B(x) := \frac{1}{2}\|Ax - b\|_2^2 + \lambda\,\mathcal{R}_B(x) \tag{2}$$

with non-convex regularizer

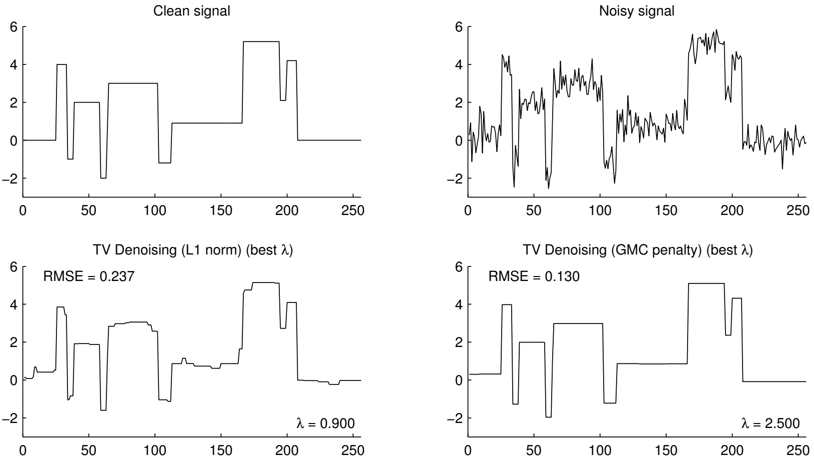
$$\mathcal{R}_B(x) := \mathcal{R}(x) - \min_{v\in\mathbb{R}^N}\left\{\tfrac{1}{2}\|B(x-v)\|_2^2 + \mathcal{R}(v)\right\}.$$

The function $\mathcal{J}_B(x)$ is convex (strongly convex) if

$$A^{\mathsf{T}}A - \lambda\,B^{\mathsf{T}}B \succcurlyeq 0 \;\; (\succ\; 0)\,.$$

Reference: A. Lanza, S. Morigi, I. Selesnick, and F. Sgallari. Sparsity-inducing Non-convex Non-separable Regularization for Convex Image Processing. 2017. Submitted.

Numerical Examples – 1D Total Variation denoising



Reference: I. Selesnick, A. Lanza, S. Morigi, and F. Sgallari. Total Variation Signal Denoising via Convex Analysis. 2017. Submitted.

Summary

- We show how to construct non-convex regularizers that preserve the convexity of functionals for sparse-regularized linear least-squares.
- Generalizes the ℓ_1 norm.
- Can be used in conjunction with other convex non-smooth regularizers (TV, nuclear norm, etc).
- Resulting regularizers are non-separable.
- Optimization implementable using proximal algorithms (as for ℓ_1 norm).

References

A. Lanza, S. Morigi, I. Selesnick, and F. Sgallari. Sparsity-inducing Non-convex Non-separable Regularization for Convex Image Processing. 2017. Submitted.

I. Selesnick, A. Lanza, S. Morigi, and F. Sgallari. Total Variation Signal Denoising via Convex Analysis. 2017. Submitted.

A. Lanza, S. Morigi, I. Selesnick, and F. Sgallari. Nonconvex nonsmooth optimization via convex–nonconvex majorization–minimization. *Numerische Mathematik*, 136(2):343–381, 2017.

I. Selesnick. Sparse Regularization via Convex Analysis. *IEEE Trans. Signal Process.*, 65(17):4481–4494, September 2017.

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

A. Lanza, S. Morigi, and F. Sgallari. Convex Image Denoising via Non-convex Regularization with Parameter Selection. *J. Math. Imaging and Vision*, 56(2):195–220, 2016.

A. Lanza, S. Morigi, and F. Sgallari. Convex Image Denoising via Non-Convex Regularization. In *Scale Space and Variational Methods in Computer Vision*. Volume 9087. 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.

Optimization Algorithm (saddle point algorithm)

With

$$\mathcal{J}_B(x) = \tfrac{1}{2}\|Ax - b\|_2^2 + \lambda\,\mathcal{R}_B(x)$$
$$\mathcal{R}_B(x) = \mathcal{R}(x) - \min_v \left\{ \tfrac{1}{2}\|B(x - v)\|_2^2 + \mathcal{R}(v) \right\}$$

we have

$$\begin{aligned}
\mathcal{J}_B(x) &= \tfrac{1}{2}\|Ax - b\|_2^2 + \lambda\mathcal{R}(x) - \lambda \min_v \left\{ \tfrac{1}{2}\|B(x - v)\|_2^2 + \mathcal{R}(v) \right\} \\
&= \tfrac{1}{2}\|Ax - b\|_2^2 + \lambda\mathcal{R}(x) + \lambda \max_v \left\{ -\tfrac{1}{2}\|B(x - v)\|_2^2 - \mathcal{R}(v) \right\}
\end{aligned}$$

and

$$\begin{aligned}
\hat{x} &= \arg\min_x \mathcal{J}_B(x) \\
&= \arg\min_x \left\{ \tfrac{1}{2}\|Ax - b\|_2^2 + \lambda\mathcal{R}(x) + \lambda \max_v \left\{ -\tfrac{1}{2}\|B(x - v)\|_2^2 - \mathcal{R}(v) \right\} \right\}
\end{aligned}$$

or a *saddle point* problem

$$(\hat{x}, \hat{v}) = \arg\min_x \max_v \left\{ \tfrac{1}{2}\|Ax - b\|_2^2 + \lambda\mathcal{R}(x) - \tfrac{\lambda}{2}\|B(x - v)\|_2^2 - \lambda\mathcal{R}(v) \right\}$$

Optimization Algorithm (saddle point algorithm)

This saddle-point problem is an instance of a monotone inclusion problem. Hence, the solution can be obtained using the forward-backward (FB) algorithm.

for $k = 0, 1, 2, \ldots$

$$w_k = x_k - \mu \left[A^{\mathsf{T}}(Ax_k - b) + \lambda\, B^{\mathsf{T}}B\,(v_k - x_k) \right]$$

$$u_k = v_k - \mu\lambda B^{\mathsf{T}}B(v_k - x_k)$$

$$x_{k+1} = \arg\min_{x \in \mathbb{R}^n} \left\{ \mathcal{R}(x) + \frac{1}{2\mu\lambda} \|x - w_k\|_2^2 \right\}$$

$$v_{k+1} = \arg\min_{v \in \mathbb{R}^n} \left\{ \mathcal{R}(v) + \frac{1}{2\mu\lambda} \|v - u_k\|_2^2 \right\}$$

end

This algorithm reduces to ISTA for $B = 0$ and $\mathcal{R}(x) = \|x\|_1$.