SPARSITY AMPLIFIED

Ivan Selesnick

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

ABSTRACT

The L1 norm is often used as a penalty function to obtain a sparse approximate solution to a system of linear equations, but it often underestimates the true values. This paper proposes a different type of penalty that (1) estimates sparse solutions more accurately and (2) maintains the convexity of the cost function. The new penalty is a multivariate generalization of the minimax-concave (MC) penalty. To define the generalized MC (GMC) penalty we first define a multivariate generalized Huber function. The resulting cost function can be minimized by proximal algorithms comprising simple computations. The effectiveness of the GMC penalty is illustrated in a denoising example.

Index Terms— Sparse regularization, sparse-regularized linear least squares, basis pursuit denoising, convex optimization.

1. INTRODUCTION

We consider the problem of sparse-regularized linear least squares, namely, the calculation of a sparse approximate solution to the linear equations \( y = Ax \) by minimizing a cost function \( F : \mathbb{R}^N \to \mathbb{R} \),

\[
F(x) = \frac{1}{2} \| y - Ax \|_2^2 + \lambda \psi(x), \quad \lambda > 0 
\]

where \( \psi : \mathbb{R}^N \to \mathbb{R} \) is a penalty function that induces sparsity of \( x \). It is common to use the \( \ell_1 \) norm for the penalty \( \psi \) because among convex penalty functions it induces sparsity most effectively [14]. It is desirable that \( F \) be a convex function, for then it does not possess non-optimal local minima. Also, setting \( \lambda \) tends to be more straightforward when \( F \) is convex. Unfortunately, the use of the \( \ell_1 \) norm tends to produce solutions that underestimate the true values.

In this work, we propose a new penalty function for the sparse-regularized linear least squares problem that improves upon the \( \ell_1 \) norm. Specifically, we propose a non-convex non-separable penalty that (i) more accurately estimates sparse solutions to \( y \approx Ax \) and (ii) maintains the convexity of the cost function \( F \).

The new penalty function can be considered a generalization of the scalar minimax-concave (MC) penalty [35]. In order to define the new penalty function, we first define a multivariate generalization of the Huber function. It is easy to ensure that the new penalty function maintains the convexity of the sparse-regularized least squares cost function (1), even when the matrix \( A \) is arbitrary. In particular, the matrix \( A \) may be such that \( A^TA \) is singular.

We define the proposed generalized MC (GMC) penalty using tools of convex analysis [1]. In particular, we base the definition on infinitesimal convolution. The proposed GMC penalty does not appear to have a simple explicit expression. Nevertheless, the solution to the sparse-regularized least squares problem with the GMC penalty can be formulated as a saddle-point problem which can then be solved using proximal algorithms (e.g., forward-backward splitting) involving only simple computations [1, 12].

We demonstrate the effectiveness of the proposed GMC penalty for denoising a signal with a sparse Fourier transform. Compared to the \( \ell_1 \)-norm solution which underestimates the true sparse Fourier coefficients, the GMC solution is amplified and more accurate.

The ideas in this paper are also applicable to TV denoising [27].

1.1. Relation to Prior Work

Many prior works have proposed non-convex penalties that strongly promote sparsity or describe algorithms for solving the sparse-regularized linear least squares problem, e.g., [5–8, 11, 15, 16, 20, 21, 23, 26, 36]. However, most of this work (i) uses separable (additive) penalties or (ii) does not seek to maintain the convexity of the cost function. Non-separable non-convex penalties are proposed in Refs. [31, 33], but they are not designed to maintain cost function convexity. The development of convexity-preserving non-convex penalties was pioneered by Blake, Zisserman, and Nikolova [4, 22–24], and further developed in [3, 9, 13, 17–19, 25, 28, 30]. But these are separable penalties, and as such they are fundamentally limited. Specifically, if \( A^TA \) is singular in problem (1), then a separable penalty constrained to maintain cost function convexity can only improve on the \( \ell_1 \) norm to a very limited extent [29].

We recently proposed a bivariate non-separable non-convex penalty to overcome the fundamental limitation of separable non-convex penalties [29]. But that penalty is useful for only a narrow class of problems (deconvolution where the transfer function of the convolution filter has no more than one zero on the unit circle). The new proposed GMC penalty is applicable to much more general linear inverse problems (e.g., deconvolution with arbitrary filters and non-deconvolution problems).

2. GENERALIZED HUBER FUNCTION

We recall the definition of the Huber function, illustrated in Fig. 1(a).

**Definition 1.** The scalar Huber function \( s : \mathbb{R} \to \mathbb{R} \) is defined as

\[
s(x) := \begin{cases} 
\frac{1}{2} x^2, & |x| \leq 1 \\
| |x| - \frac{1}{2}, & |x| > 1.
\end{cases} \tag{2}
\]

In this work, we propose a new multivariate generalization of the scalar Huber function. We will use it in Sec. 3.

**Definition 2.** Given a matrix \( B \in \mathbb{R}^{M \times N} \), we define the generalized Huber function \( S_B : \mathbb{R}^N \to \mathbb{R} \) as

\[
S_B(x) := \min_{v \in \mathbb{R}^N} \left\{ \| v \|_1 + \frac{1}{2} \| B(x - v) \|_2^2 \right\}. \tag{3}
\]
The generalized Huber function satisfies
\[ 0 \leq S_B(x) \leq \|x\|_1, \quad \forall x \in \mathbb{R}^N. \tag{10} \]

Proposition 2. The generalized Huber function satisfies
\[ S_B(x) = \frac{1}{2} \|Bx\|_2^2, \quad \forall x \in \mathbb{R}^N \text{ such that } \|B^TB\|_\infty \leq 1. \tag{11} \]

Proposition 3. The generalized Huber function satisfies
\[ S_B(x) \leq \frac{1}{\|B\|_2^2} \sum_n s(\|B\|_2 x_n), \quad \forall x \in \mathbb{R}^N \tag{12} \]
where \( \|B\|_2^2 \) is the maximum eigenvalue of \( B^TB \).

The generalized Huber function can be expressed in terms of infimal convolution,
\[ S_B = \| \cdot \|_1 \square \frac{1}{2} \|B \cdot \|_2^2. \tag{13} \]

Infimal convolution is well studied in convex analysis [1]. Hence, by expressing the generalized Huber function in terms of infimal convolution, we can draw on results in convex analysis to derive further properties of the generalized Huber function.

Proposition 4. The generalized Huber function \( S_B \) is a proper lower semicontinuous convex function.

Lemma 1. The generalized Huber function is differentiable.

Lemma 2. The gradient of the generalized Huber function satisfies
\[ \|\nabla S_B(x)\|_\infty \leq 1, \quad \forall x \in \mathbb{R}^N. \tag{14} \]

3. GENERALIZED MC PENALTY

In this section we use the generalized Huber function to define a new class of multivariate non-convex penalty functions that maintain the convexity of the sparse-regularized least squares cost function (1).

We start with the scalar penalty function illustrated in Fig. 1(b). This is the minimax-concave (MC) penalty [2, 35].

Definition 3. The scalar minimax-concave (MC) penalty function \( \phi: \mathbb{R} \to \mathbb{R} \) is defined as
\[ \phi(x) := \begin{cases} |x| - \frac{1}{2} x^2, & |x| \leq 1 \\ \frac{1}{2}, & |x| > 1. \end{cases} \tag{15} \]
The scalar MC penalty function can be expressed as
\[ \phi(x) = |x| - s(x). \tag{16} \]

where \( s \) is the scalar Huber function (2).

We propose a multivariate generalization of the scalar MC penalty function (15). The basic idea is to generalize (16) using the \( \ell_1 \) norm and the generalized Huber function.

Definition 4. Given \( B \in \mathbb{R}^{M \times N} \), we define the generalized MC penalty function \( \psi_B: \mathbb{R}^N \to \mathbb{R} \) as
\[ \psi_B(x) := \|x\|_1 - S_B(x) \tag{17} \]
where \( S_B \) is the generalized Huber function (3).
The most interesting case (the case that motivates us to define the GMC penalty) is the case where \( B^T B \) is a non-diagonal matrix. If \( B^T B \) is non-diagonal, then the GMC penalty is non-separable.

Example 7. For the matrix \( B \) in (9), the generalized MC penalty \( \psi_B \) is shown in Fig. 2(b).

4. SPARSE-REGULARIZED LEAST SQUARES

We now consider how to set the proposed GMC penalty so as to maintain the convexity of the considered cost function (1).

Theorem 1. Let the scalar \( \lambda > 0 \) and the matrix \( A \in \mathbb{R}^{M \times N} \) be given. Define \( F : \mathbb{R}^N \to \mathbb{R} \) as

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

where \( \psi_B \) is the generalized MC penalty (17). If

\[
B^T B \preceq \frac{1}{\lambda} A^T A
\]

then \( F \) is a convex function. (The meaning of expression (22) is that \((1/\lambda) A^T A - B^T B\) is positive semidefinite.)

Proof. Using (3) and (17), we write

\[
F(x) = \frac{1}{2} \| y - Ax \|_2^2 + \lambda \left( \| x \|_1 - S_B(x) \right)
\]

(23)

\[
= \frac{1}{2} \| y - Ax \|_2^2 + \lambda \| x \|_1
- \min_{v \in \mathbb{R}^N} \left\{ \lambda \| v \|_1 + \frac{1}{2} \| B(x - v) \|_2^2 \right\}
\]

(24)

\[
= \max_{v \in \mathbb{R}^N} \left\{ \frac{1}{2} \| y - Ax \|_2^2 + \lambda \| x \|_1
- \lambda \| v \|_1 - \frac{1}{2} \| B(x - v) \|_2^2 \right\}
\]

(25)

\[
= \max_{v \in \mathbb{R}^N} \left\{ \frac{1}{2} x^T \left( A^T A - \lambda B^T B \right)x + \lambda \| x \|_1 + \max_{v \in \mathbb{R}^N} g(x, v) \right\}
\]

(26)

where \( g \) is affine in \( x \). The last term is convex as it is the point-wise supremum of a set of convex functions (Prop. 8.14 in [1]). Hence, \( F \) is convex if \( A^T A - \lambda B^T B \) is positive semidefinite.

The convexity condition (22) is easily satisfied. Given \( A \), we may simply set

\[
B = \sqrt{\gamma / \lambda} A, \quad 0 \leq \gamma \leq 1.
\]

Then \( B^T B = (\gamma / \lambda) A^T A \) which satisfies (22) when \( \gamma \leq 1 \). The parameter \( \gamma \) controls the non-convexity of the penalty \( \psi_B \). If \( \gamma = 0 \), then \( B = 0 \) and the penalty reduces to the \( \ell_1 \) norm. If \( \gamma = 1 \), then (22) is satisfied with equality and the penalty is ‘maximally’ non-convex. We suggest in practice a nominal range of 0.5 \( \leq \gamma \leq 0.8 \).

4.1. Optimization

Even though the GMC penalty does not have a simple explicit formula, a global minimizer of the sparse-regularized cost function (21) can be readily calculated using proximal algorithms. It is not necessary to explicitly evaluate the GMC penalty or its gradient.

To minimize the cost function \( F \) in (21) using proximal algorithms, we rewrite it as a saddle-point problem. We assume \( B^T B = (\gamma / \lambda) A^T A \) with 0 \( \leq \gamma \leq 1 \), cf. (28). Then the problem of minimizing \( F \) in (21) can be written as the saddle-point problem:
Denoising using L1 norm

Denoising using GMC

FFT of noisy signal

Optimized coefficients

Fig. 3: Denoising using the \( \ell_1 \) norm and the proposed GMC penalty. The plot of optimized coefficients shows only the non-zero values.

\[
(x_{\text{opt}}, v_{\text{opt}}) = \arg \min_{x \in \mathbb{R}^N} \max_{v \in \mathbb{R}^N} F(x, v)
\]

\[
F(x, v) = \frac{1}{2} \| y - Ax \|^2 + \lambda \| x \|_1 - \frac{\gamma}{2} \| A(x - v) \|^2 - \lambda \| v \|_1.
\]

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 such a problem; see Theorem 25.8 of Ref. [1]. The FB algorithm involves only simple computational steps (soft-thresholding and the operators \( A \) and \( A^* \)).

\[\text{Fig. 3: Denoising using the } \ell_1 \text{ norm and the proposed GMC penalty.} \]

\[\text{The plot of optimized coefficients shows only the non-zero values.}\]

\[\text{\( (x_{\text{opt}}, v_{\text{opt}}) = \arg \min_{x \in \mathbb{R}^N} \max_{v \in \mathbb{R}^N} F(x, v) \) (29)} \]

\[\text{where}\]

\[F(x, v) = \frac{1}{2} \| y - Ax \|^2 + \lambda \| x \|_1 - \frac{\gamma}{2} \| A(x - v) \|^2 - \lambda \| v \|_1. \]

\[\text{(30)} \]

5. NUMERICAL EXPERIMENT

This example illustrates the use of the generalized MC penalty for denoising [10]. We consider the discrete-time signal

\[g(m) = 2 \cos(2\pi f_1 m) + \sin(2\pi f_2 m), \quad m = 0, \ldots, M - 1 \] (31)

doing experiment, we corrupt the signal with additive white Gaussian noise (AWGN) with standard deviation \( \sigma = 1.0 \). See Fig. 3.

\[\text{The columns of } A \text{ form a normalized tight frame, i.e., } A A^H = I \]

where \( A^H \) is the complex conjugate transpose of \( A \). For the denoising experiment, we corrupt the signal with additive white Gaussian noise (AWGN) with standard deviation \( \sigma = 1.0 \). See Fig. 3.

\[\text{In the sparse-regularized linear least squares problem, the proposed penalty promotes sparsity more effectively than the } \ell_1 \text{ norm while maintaining the convexity of the cost function to be minimized.}\]

The \( \ell_1 \)-norm and GMC solutions are shown in Fig. 3. We set \( \lambda \) in each case so as to minimize the average root-mean-square error (RMSE). This leads to the values \( \lambda = 1.8 \) and \( \lambda = 2.6 \), respectively. For the GMC method, we set \( B \) using (28) with \( \gamma = 0.8 \). Since \( B^HB \) is not diagonal, the GMC penalty is non-separable.

Like the \( \ell_1 \)-norm solution, the GMC solution minimizes a convex cost function; yet, the GMC solution is significantly more accurate than the \( \ell_1 \)-norm solution, as shown in Fig. 3. We observe that the \( \ell_1 \)-norm solution is not as sparse in the frequency domain as the GMC solution. In addition, the \( \ell_1 \)-norm solution underestimates the significant (large-amplitude) Fourier coefficients. In comparison with the \( \ell_1 \)-norm solution, the GMC solution is amplified.

Neither increasing nor decreasing the regularization parameter \( \lambda \) helps the \( \ell_1 \)-norm solution here. A larger value of \( \lambda \) would make the \( \ell_1 \)-norm solution sparser, but would reduce the amplitudes of the significant Fourier coefficients, i.e., the significant coefficients would be even more underestimated. A smaller value of \( \lambda \) would increase the significant Fourier coefficients of the \( \ell_1 \)-norm solution, but would make the solution less sparse and more noisy.

The favorable behavior of GMC is maintained over a range of noise levels, as shown in Fig. 4. The figure shows the average RMSE as a function of \( \sigma \) for \( 0.25 \leq \sigma \leq 2.0 \). We calculate the average RMSE here using 150 noise realizations for each \( \sigma \). The value of \( \lambda \) must be adjusted according to the noise level \( \sigma \). In this experiment, for each noise level \( \sigma \), we scale \( \lambda \) proportional to \( \sigma \). That is, for the \( \ell_1 \) norm we set \( \lambda = 1.8\sigma \); for the GMC penalty we set \( \lambda = 2.6\sigma \).

The GMC approach also compares favorably with the \textit{iterative p-shrinkage} (IPS) algorithm [32,34]. We compare GMC to this algorithm because, in a detailed comparison of several algorithms [29], we found the IPS algorithm performed particularly well. In contrast to GMC, the IPS algorithm aims to minimize a non-convex cost function (both algorithms use non-convex penalties). For IPS we set \( \lambda \) so as to minimize the average RMSE at \( \sigma = 1 \) and then scale it linearly for other \( \sigma \) (as we did for the \( \ell_1 \)-norm and GMC methods). (This strategy for setting \( \lambda \) is not necessarily optimal for any of the methods, but it is straightforward and indicative of the sensitivity of each method to \( \lambda \).)

As shown in Fig. 4, the GMC method compares favorably to the IPS algorithm in this experiment.

6. CONCLUSION

In the sparse-regularized linear least squares problem, the proposed penalty promotes sparsity more effectively than the \( \ell_1 \) norm while maintaining the convexity of the cost function to be minimized.
7. REFERENCES


