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 $F\colon \mathbb{R}^N \to \mathbb{R}$,

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

where $\psi\colon \mathbb{R}^N \to \mathbb{R}$ is a penalty function that induces sparsity of x. It is common to use the ℓ_1 norm for the penalty ψ 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 λ tends to be more straightforward when F is convex. Unfortunately, the use of the ℓ_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 ℓ_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 $A^{\mathsf{T}}A$ is singular.

We define the proposed generalized MC (GMC) penalty using tools of convex analysis [1]. In particular, we base the definition on *infimal 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 ℓ_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 $A^{\mathsf{T}}A$ is singular in problem (1), then a separable penalty constrained to maintain cost function convexity can only improve on the ℓ_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 $s\colon \mathbb{R} \to \mathbb{R}$ is defined as*

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

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 $S_B\colon \mathbb{R}^N \to \mathbb{R}$ as*

$$S_B(x) := \min_{v \in \mathbb{R}^N} \big\{ \|v\|_1 + \tfrac{1}{2}\|B(x - v)\|_2^2 \big\}. \tag{3}$$

This material is based upon work supported by the National Science Foundation under Grant No. 1525398 and the Office of Naval Research under grant No. N00014-15-1-2314.

IEEE ICASSP, March 2017

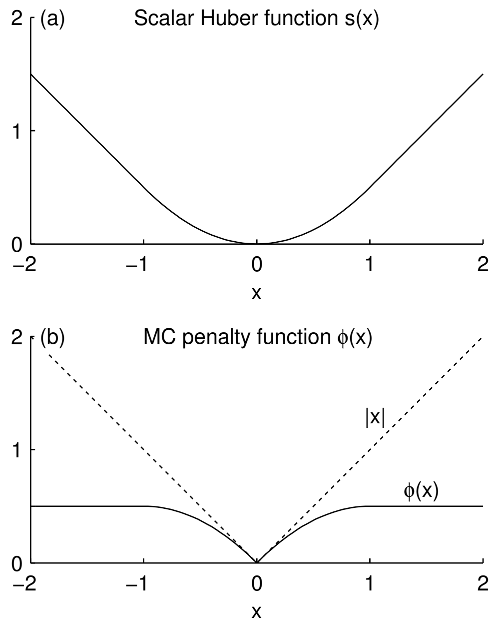

Fig. 1: (a) The scalar Huber function. (b) The MC penalty function.

Although expression (3) does not resemble the formula for the scalar Huber function (2), it does in fact reduce to the scalar Huber function as a special case.

Example 1. *The generalized Huber function reduces to the scalar Huber function when $B = 1$, i.e.,*

$$S_1(x) = s(x), \ \ \forall x \in \mathbb{R}. \tag{4}$$

The generalized Huber function reduces to a scaled scalar Huber function when B is a non-zero scalar. If $B = \alpha \neq 0$, then

$$S_\alpha(x) = \frac{1}{\alpha^2}\, s(\alpha^2 x), \ \ \forall x \in \mathbb{R}. \tag{5}$$

Also,

$$\lim_{\alpha\to\infty} S_\alpha(x) = |x|, \ \ \forall x \in \mathbb{R}. \tag{6}$$

Example 2. *The generalized Huber function reduces to the zero function when B is the zero matrix. If $B = 0 \in \mathbb{R}^{M\times N}$, then*

$$S_0(x) = 0, \ \ \forall x \in \mathbb{R}^N. \tag{7}$$

The value zero is obtained with $v = 0$ in the definition (3)*.*

The generalized Huber function reduces to a separable (additive) function as special case.

Example 3. *If $B^{\mathsf{T}} B = \alpha^2 I$ with $\alpha \neq 0$, then S_B is given by a sum of scaled scalar Huber functions,*

$$B^{\mathsf{T}} B = \alpha^2 I \quad \Longrightarrow \quad S_B(x) = \frac{1}{\alpha^2} \sum_n s(\alpha^2 x_n). \tag{8}$$

The utility of the generalized Huber function will be most apparent when $B^{\mathsf{T}} B$ is a non-diagonal matrix. In this case, the generalized Huber function is non-separable.

Example 4. *For the matrix*

$$B = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 1 \end{bmatrix} \tag{9}$$

the generalized Huber function S_B is non-separable. See Fig. 2(a).

There is not a simple explicit formula for the generalized Huber function for $N > 2$. But, using the definition (3), we can derive several properties regarding the function. The proofs of the following properties will be given in a future full-length paper.

Proposition 1. *The generalized Huber function satisfies*

$$0 \leqslant S_B(x) \leqslant \|x\|_1, \ \ \forall x \in \mathbb{R}^N. \tag{10}$$

Proposition 2. *The generalized Huber function satisfies*

$$S_B(x) = \frac{1}{2}\|Bx\|_2^2, \ \ \forall x \in \mathbb{R}^N \text{ such that } \|B^{\mathsf{T}} B x\|_\infty \leqslant 1. \tag{11}$$

Proposition 3. *The generalized Huber function satisfies*

$$S_B(x) \leqslant \frac{1}{\|B\|_2^2} \sum_n s\big(\|B\|_2^2\, x_n\big), \ \ \forall x \in \mathbb{R}^N \tag{12}$$

where $\|B\|_2^2$ is the maximum eigenvalue of $B^{\mathsf{T}} B$.

The generalized Huber function can be expressed in terms of infimal convolution,

$$S_B = \|\cdot\|_1 \,\square\, \tfrac{1}{2}\|B\,\cdot\|_2^2. \tag{13}$$

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*

$$\|\nabla S_B(x)\|_\infty \leqslant 1, \ \ \forall x \in \mathbb{R}^N. \tag{14}$$

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 $\phi\colon \mathbb{R} \to \mathbb{R}$ is defined as*

$$\phi(x) := \begin{cases} |x| - \frac{1}{2}x^2, & |x| \leqslant 1 \\ \frac{1}{2}, & |x| \geqslant 1. \end{cases} \tag{15}$$

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 ℓ_1 norm and the generalized Huber function.

Definition 4. *Given $B \in \mathbb{R}^{M\times N}$, we define the generalized MC penalty function $\psi_B\colon \mathbb{R}^N \to \mathbb{R}$ as*

$$\psi_B(x) := \|x\|_1 - S_B(x) \tag{17}$$

where S_B is the generalized Huber function (3)*.*

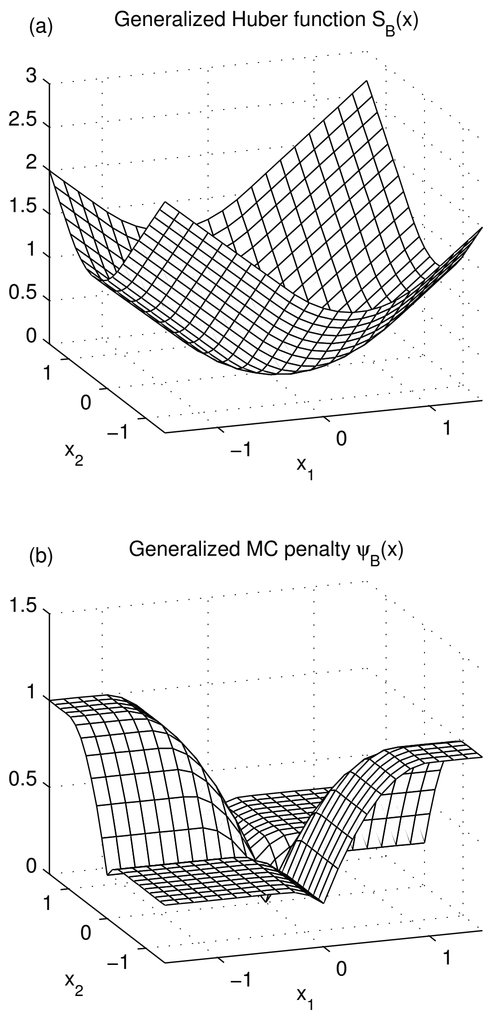

Fig. 2: (a) The generalized Huber function S_B and (b) the generalized MC penalty function ψ_B, both for the matrix B in (9).

The generalized MC (GMC) penalty reduces to a separable penalty when $B^{\mathsf{T}}B$ is diagonal.

Example 5. *If $B^{\mathsf{T}}B = \alpha^2 I$ with $\alpha \neq 0$, then ψ_B is given by a sum of scaled scalar MC penalty functions,*

$$B^{\mathsf{T}}B = \alpha^2 I \quad \Longrightarrow \quad \psi_B(x) = \frac{1}{\alpha^2} \sum_n \phi(\alpha^2 x_n). \tag{18}$$

The generalized MC penalty is upper bounded by the ℓ_1 norm, which is recovered as a special case (when B is the zero matrix).

Example 6. *For all $x \in \mathbb{R}^N$, the generalized MC penalty satisfies*

$$\psi_0(x) = \|x\|_1 \tag{19}$$

and

$$0 \leqslant \psi_B(x) \leqslant \|x\|_1 \tag{20}$$

which follow from (7) *and* (10).

Even though the GMC penalty is defined as the difference of two convex functions, inequality (20) ensures the GMC penalty is non-negative. This is important because penalty functions should usually be non-negative.

The most interesting case (the case that motivates us to define the GMC penalty) is the case where $B^{\mathsf{T}}B$ is a non-diagonal matrix. If $B^{\mathsf{T}}B$ is non-diagonal, then the GMC penalty is non-separable.

Example 7. *For the matrix B in* (9)*, the generalized MC penalty ψ_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 $F\colon \mathbb{R}^N \to \mathbb{R}$ as*

$$F(x) = \frac{1}{2}\|y - Ax\|_2^2 + \lambda\, \psi_B(x) \tag{21}$$

where ψ_B is the generalized MC penalty (17)*. If*

$$B^{\mathsf{T}}B \preccurlyeq \frac{1}{\lambda} A^{\mathsf{T}}A \tag{22}$$

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

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

$$\begin{aligned}
F(x) &= \tfrac{1}{2}\|y - Ax\|_2^2 + \lambda \left(\|x\|_1 - S_B(x)\right) && (23)\\
&= \tfrac{1}{2}\|y - Ax\|_2^2 + \lambda \|x\|_1 && (24)\\
&\qquad - \min_{v \in \mathbb{R}^N} \left\{\lambda\|v\|_1 + \tfrac{\lambda}{2}\|B(x-v)\|_2^2\right\} \\
&= \max_{v \in \mathbb{R}^N} \left\{\tfrac{1}{2}\|y - Ax\|_2^2 + \lambda \|x\|_1\right. && (25)\\
&\qquad \left. - \lambda\|v\|_1 - \tfrac{\lambda}{2}\|B(x-v)\|_2^2\right\} \\
&= \max_{v \in \mathbb{R}^N} \left\{\tfrac{1}{2}x^{\mathsf{T}}\big(A^{\mathsf{T}}A - \lambda B^{\mathsf{T}}B\big)x + \lambda\|x\|_1 + g(x,v)\right\} && (26)\\
&= \tfrac{1}{2}x^{\mathsf{T}}\big(A^{\mathsf{T}}A - \lambda B^{\mathsf{T}}B\big)x + \lambda\|x\|_1 + \max_{v \in \mathbb{R}^N} g(x,v) && (27)
\end{aligned}$$

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 $A^{\mathsf{T}}A - \lambda B^{\mathsf{T}}B$ is positive semidefinite. □

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

$$B = \sqrt{\gamma/\lambda}\, A, \quad 0 \leqslant \gamma \leqslant 1. \tag{28}$$

Then $B^{\mathsf{T}}B = (\gamma/\lambda)A^{\mathsf{T}}A$ which satisfies (22) when $\gamma \leqslant 1$. The parameter γ controls the non-convexity of the penalty ψ_B. If $\gamma = 0$, then $B = 0$ and the penalty reduces to the ℓ_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 $0.5 \leqslant \gamma \leqslant 0.8$.

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 $B^{\mathsf{T}}B = (\gamma/\lambda)A^{\mathsf{T}}A$ with $0 \leqslant \gamma < 1$, cf. (28). Then the problem of minimizing F in (21) can be written as the saddle-point problem:

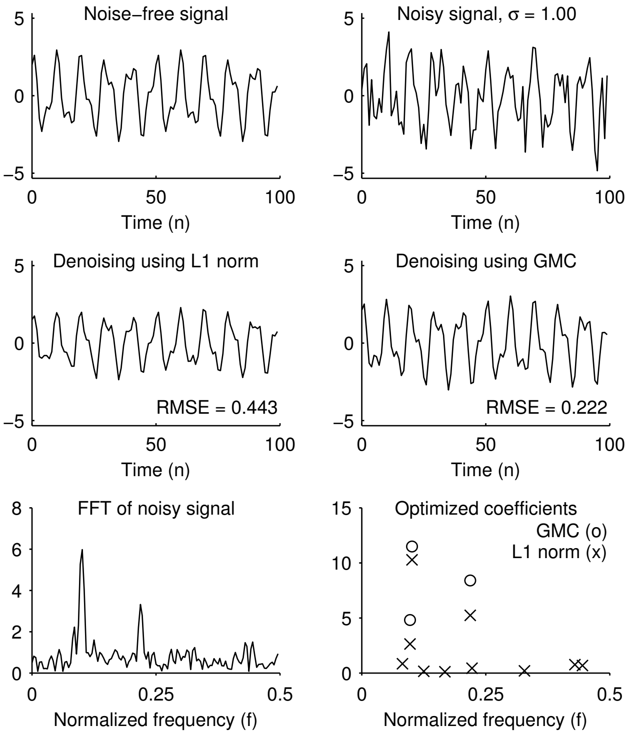

Fig. 3: Denoising using the ℓ_1 norm and the proposed GMC penalty. The plot of optimized coefficients shows only the non-zero values.

$$(x^{\mathsf{opt}},\, v^{\mathsf{opt}}) = \arg \min_{x\in\mathbb{R}^N} \max_{v\in\mathbb{R}^N} F(x,v) \tag{29}$$

where

$$F(x,v) = \frac{1}{2}\|y - Ax\|_2^2 + \lambda\,\|x\|_1 - \frac{\gamma}{2}\|A(x-v)\|_2^2 - \lambda\|v\|_1. \tag{30}$$

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 problems; see Theorem 25.8 of Ref. [1]. The FB algorithm involves only simple computational steps (soft-thresholding and the operators A and A^{T}).

5. NUMERICAL EXPERIMENT

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

$$g(m) = 2\cos(2\pi f_1 m) + \sin(2\pi f_2 m), \;\; m = 0,\dots,M-1 \tag{31}$$

of length $M = 100$ with frequencies $f_1 = 0.1$ and $f_2 = 0.22$. The signal g is sparse in the frequency domain, so we model the signal as $g = Ax$ where A is an over-sampled inverse discrete Fourier transform and $x \in \mathbb{C}^N$ is a sparse vector of Fourier coefficients with $N \geqslant M$. Specifically, we define the matrix $A \in \mathbb{C}^{M\times N}$ as

$$A_{m,n} = \big(1/\sqrt{N}\,\big)\exp(\mathrm{j}(2\pi/N)mn), \qquad m = 0,\dots,M-1,\; n = 0,\dots,N-1. \tag{32}$$

The columns of A form a normalized tight frame, i.e., $AA^{\mathsf{H}} = I$ 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.

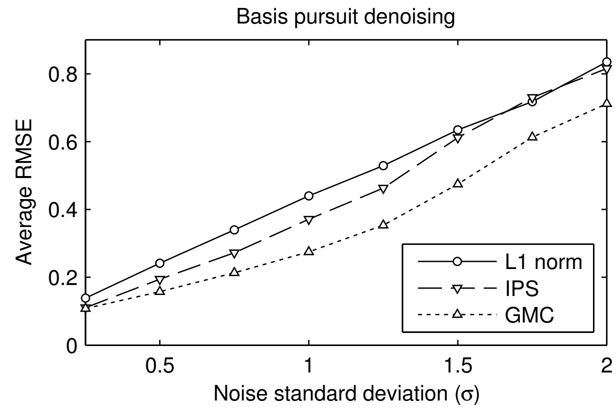

Fig. 4: Average RMSE for three denoising methods.

The ℓ_1-norm and GMC solutions are shown in Fig. 3. We set λ 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 $B^{\mathsf{H}}B$ is not diagonal, the GMC penalty is non-separable.

Like the ℓ_1-norm solution, the GMC solution minimizes a *convex* cost function; yet, the GMC solution is significantly more accurate than the ℓ_1-norm solution, as shown in Fig. 3. We observe that the ℓ_1-norm solution is not as sparse in the frequency domain as the GMC solution. In addition, the ℓ_1-norm solution underestimates the significant (large-amplitude) Fourier coefficients. In comparison with the ℓ_1-norm solution, the GMC solution is *amplified*.

Neither increasing nor decreasing the regularization parameter λ helps the ℓ_1-norm solution here. A larger value of λ would make the ℓ_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 λ would increase the significant Fourier coefficients of the ℓ_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 σ for $0.25 \leqslant \sigma \leqslant 2.0$. We calculate the average RMSE here using 150 noise realizations for each σ. The value of λ must be adjusted according to the noise level σ. In this experiment, for each noise level σ, we scale λ proportional to σ. That is, for the ℓ_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 *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 λ so as to minimize the average RMSE at $\sigma = 1$ and then scale it linearly for other σ (as we did for the ℓ_1-norm and GMC methods). (This strategy for setting λ is not necessarily optimal for any of the methods, but it is straightforward and indicative of the sensitivity of each method to λ.) 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 ℓ_1 norm while maintaining the convexity of the cost function to be minimized.

7. REFERENCES

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

[2] İ. Bayram. Penalty functions derived from monotone mappings. *IEEE Signal Processing Letters*, 22(3):265–269, March 2015.

[3] I. Bayram. On the convergence of the iterative shrinkage/thresholding algorithm with a weakly convex penalty. *IEEE Trans. Signal Process.*, 64(6):1597–1608, March 2016.

[4] A. Blake and A. Zisserman. *Visual Reconstruction*. MIT Press, 1987.

[5] E. J. Candès, M. B. Wakin, and S. Boyd. Enhancing sparsity by reweighted l1 minimization. *J. Fourier Anal. Appl.*, 14(5):877–905, December 2008.

[6] M. Castella and J.-C. Pesquet. Optimization of a Geman-McClure like criterion for sparse signal deconvolution. In *IEEE Int. Workshop Comput. Adv. Multi-Sensor Adaptive Proc. (CAMSAP)*, pages 309–312, December 2015.

[7] R. Chartrand. Shrinkage mappings and their induced penalty functions. In *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP)*, pages 1026–1029, May 2014.

[8] L. Chen and Y. Gu. The convergence guarantees of a nonconvex approach for sparse recovery. *IEEE Trans. Signal Process.*, 62(15):3754–3767, August 2014.

[9] P.-Y. Chen and I. W. Selesnick. Group-sparse signal denoising: Non-convex regularization, convex optimization. *IEEE Trans. Signal Process.*, 62(13):3464–3478, July 2014.

[10] S. Chen, D. L. Donoho, and M. A. Saunders. Atomic decomposition by basis pursuit. *SIAM J. Sci. Comput.*, 20(1):33–61, 1998.

[11] E. Chouzenoux, A. Jezierska, J. Pesquet, and H. Talbot. A majorize-minimize subspace approach for $\ell_2 - \ell_0$ image regularization. *SIAM J. Imag. Sci.*, 6(1):563–591, 2013.

[12] P. L. Combettes and J.-C. Pesquet. Proximal splitting methods in signal processing. In H. H. Bauschke et al., editors, *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, pages 185–212. Springer-Verlag, 2011.

[13] 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.

[14] M. Elad. *Sparse and redundant representations: from theory to applications in signal and image processing*. Springer, 2010.

[15] J. Fan and R. Li. Variable selection via nonconcave penalized likelihood and its oracle properties. *J. Amer. Statist. Assoc.*, 96(456):1348–1360, 2001.

[16] A. Gholami and S. M. Hosseini. A general framework for sparsity-based denoising and inversion. *IEEE Trans. Signal Process.*, 59(11):5202–5211, November 2011.

[17] W. He, Y. Ding, Y. Zi, and I. W. Selesnick. Sparsity-based algorithm for detecting faults in rotating machines. *Mechanical Systems and Signal Processing*, 72-73:46–64, May 2016.

[18] A. Lanza, S. Morigi, and F. Sgallari. Convex image denoising via non-convex regularization with parameter selection. *J. Math. Imaging and Vision*, pages 1–26, 2016.

[19] M. Malek-Mohammadi, C. R. Rojas, and B. Wahlberg. A class of nonconvex penalties preserving overall convexity in optimization-based mean filtering. *IEEE Trans. Signal Process.*, 64(24):6650–6664, December 2016.

[20] Y. Marnissi, A. Benazza-Benyahia, E. Chouzenoux, and J.-C. Pesquet. Generalized multivariate exponential power prior for wavelet-based multichannel image restoration. In *Proc. IEEE Int. Conf. Image Processing (ICIP)*, 2013.

[21] H. Mohimani, M. Babaie-Zadeh, and C. Jutten. A fast approach for overcomplete sparse decomposition based on smoothed l0 norm. *IEEE Trans. Signal Process.*, 57(1):289–301, January 2009.

[22] M. Nikolova. Estimation of binary images by minimizing convex criteria. In *Proc. IEEE Int. Conf. Image Processing (ICIP)*, pages 108–112 vol. 2, 1998.

[23] M. Nikolova. Energy minimization methods. In O. Scherzer, editor, *Handbook of Mathematical Methods in Imaging*, chapter 5, pages 138–186. Springer, 2011.

[24] M. Nikolova, M. K. Ng, and C.-P. Tam. Fast nonconvex nonsmooth minimization methods for image restoration and reconstruction. *IEEE Trans. Image Process.*, 19(12):3073–3088, December 2010.

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

[26] J. Portilla and L. Mancera. L0-based sparse approximation: two alternative methods and some applications. In *Proceedings of SPIE*, volume 6701 (Wavelets XII), San Diego, CA, USA, 2007.

[27] I. W. Selesnick. Total variation denoising via the Moreau envelope. *IEEE Signal Processing Letters*, 2017. To appear.

[28] I. W. Selesnick and I. Bayram. Sparse signal estimation by maximally sparse convex optimization. *IEEE Trans. Signal Process.*, 62(5):1078–1092, March 2014.

[29] I. W. Selesnick and I. Bayram. Enhanced sparsity by non-separable regularization. *IEEE Trans. Signal Process.*, 64(9):2298–2313, May 2016.

[30] 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.

[31] M. E. Tipping. Sparse Bayesian learning and the relevance vector machine. *J. Machine Learning Research*, 1:211–244, 2001.

[32] S. Voronin and R. Chartrand. A new generalized thresholding algorithm for inverse problems with sparsity constraints. In *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP)*, pages 1636–1640, May 2013.

[33] D. P. Wipf, B. D. Rao, and S. Nagarajan. Latent variable Bayesian models for promoting sparsity. *IEEE Trans. Inform. Theory*, 57(9):6236–6255, September 2011.

[34] J. Woodworth and R. Chartrand. Compressed sensing recovery via nonconvex shrinkage penalties. *Inverse Problems*, 32(7):75004–75028, July 2016.

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

[36] H. Zou and R. Li. One-step sparse estimates in nonconcave penalized likelihood models. *Ann. Statist.*, 36(4):1509–1533, 2008.