SHARPENING SPARSE REGULARIZERS

Abdullah Al-Shabili and Ivan Selesnick

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

ABSTRACT

Non-convex penalties outperform the convex ℓ_1-norm, but generally sacrifice the cost function convexity. As a middle ground, we propose a framework to design non-convex penalties that induce sparsity more effectively than the ℓ_1-norm, but without sacrificing the cost function convexity. The non-smooth non-convex regularizers are constructed by subtracting from the non-smooth convex penalty its smoothed version. We propose a generalized infimal convolution smoothing smoothing technique to obtain the smoothed version. We call the proposed framework *sharpening sparse regularizers* (SSR) to imply its advantages compared to convex and non-convex regularizers. The SSR framework is applicable to any sparsity regularized ill-posed linear inverse problem. Furthermore, it recovers and generalizes several non-convex penalties in the literature as special cases. The SSR-RLS problem can be formulated as a saddle point problem, and solved by a scalable generalized primal-dual algorithm. The effectiveness of the SSR framework is demonstrated by numerical experiments.

***Index Terms*—** Sparsity, convex analysis, convex optimization, non-convexity, smoothing.

1. INTRODUCTION

Sparse approximation is a prominent theme in numerous signal and image processing applications. This is due to the fact that high-dimensional signals often admit a lower dimension representation [1]. Mathematically, the problem can be stated as finding an approximate solution $x \in \mathbb{R}^n$ to a system of linear equations $Ax = y$, where $A \in \mathbb{R}^{m\times n}$ is the measurement matrix. Sparse approximation is formulated as a regularized least squares (RLS) problem

$$\inf_x F(x) = (1/2)\|y - Ax\|_2^2 + \lambda\,\Psi(Wx) \tag{1}$$

where $\lambda > 0$ is a regularization parameter, Ψ is a sparsity-inducing regularizer, and $W \in \mathbb{R}^{p\times n}$ is an analysis matrix such as the Fourier transform, gradient (total variation, TV), wavelet transform, or data-based transforms. To obtain the sparsest solution, the regularizer should ideally be the ℓ_0 pseudo-norm, but the resultant problem is NP-hard. Often, the ℓ_1-norm is used as a convex surrogate. The ℓ_1-RLS problem is known as LASSO [2]. Nonetheless, the ℓ_1-norm exhibits solution bias, since it underestimates high amplitude components. To rectify this issue, many prior works proposed non-convex penalties and algorithms [3–6]. However, the use of non-convex penalties usually yields a cost function that is not convex, thus global optimality is not guaranteed.

In this paper, we seek a middle ground between convex and non-convex penalties. This is accomplished by proposing a general framework to design non-convex penalties which effectively induce sparsity while preserving convexity of the cost function. The proposed framework is general and flexible, as it is applicable to a wide range of settings such as arbitrary matrices A (does not need to be injective nor surjective) and W. Furthermore, it generates new penalties as well as recovering and generalizing several well known non-convex penalties in the literature such as MC [3], generalized MC (GMC) [7], logarithm [4], and exponential [8].

The proposed framework relies on the synergy between sparsity-inducing convex penalties and their smoothed function. In fact, the proposed non-convex regularizers are formed by subtracting from the non-smooth convex penalty its smoothed version. In order to obtain the smoothed version, we propose a smoothing technique which we call a generalized infimal convolution smoothing (GICS) technique using convex analysis. The GICS technique is a generalization of the classical infimal convolution smoothing [9]. Smoothing techniques are usually motivated for algorithmic purposes [9–11], however, this work takes advantage of the connection between smoothing and sparsity-inducing regularizer. We call proposed the framework *sharpening sparse regularizers* (SSR) to imply its pros compared to sparsity-inducing convex and non-convex regularizers.

The development of convexity-preserving non-convex penalties was pioneered by Blake, Zisserman, and Nikolova [12–14], and further improved in [15–19]. However, these penalties are separable which makes them ineffective in case the Hessian of the data fidelity term, $A^T A$, is singular. Recently, for the case when $W = I$ in (1), we proposed the GMC penalty as a non-separable non-convex regularizer defined using convex analysis [7]. In addition, the GMC concept was partially applied to the total variation denoising problem [20]. Compared to previous work, the SSR framework can generate new non-separable penalties, thus no restrictions on A are imposed. Furthermore, the GMC penalty turns out to be a special case of the SSR framework.

The SSR-RLS problem can be formulated as a saddle point problem and solved by a scalable generalized primal-dual algorithm. Numerical experiments illustrates the effectiveness of the SSR framework on different examples. The proofs are omitted due to space limitations and will be given in a future full-length paper.

2. BACKGROUND

2.1. Convex Analysis Preliminaries

For readers' convenience, we provide a brief summary of convex analysis, for more detailed discussion see [21–23]. In this work, we focus on the class of extended-real-valued function that are proper lower semicontinuous (lsc) convex donated as $\Gamma_0(\mathbb{R}^n) = \{f : \mathbb{R}^n \to \mathbb{R} \cup \{\pm\infty\} \mid f \text{ proper, lsc and convex}\}$. The Legendre-Fenchel transform, or conjugate function f^* is an

Thanks to XYZ agency for funding.

important tool in convex analysis that is defined as

$$f^*(u) = \sup_x \left\{ x^T u - f(x) \right\} \tag{2}$$

note that $f = f^{**}$ if and only if $f \in \Gamma_0(\mathbb{R}^n)$ [23]. The epi-multiplication operation and its conjugate are defined as

$$f_\mu(\cdot) = \mu f(\cdot/\mu), \quad (f_\mu)^* = \mu f^* \tag{3}$$

where $\mu > 0$. Furthermore, the infimal convolution of two proper functions f_1 and f_2 is defined as

$$(f_1 \,\square\, f_2)(x) = \inf_v \left\{ f_1(v) + f_2(x - v) \right\}. \tag{4}$$

It is insightful to consider of infimal convolution and Fenchel conjugate as an analogy of the convolution integral and Fourier transform in signal processing [24] . In fact, the conjugate of the infimal convolution is given as

$$(f_1 \,\square\, f_2)^* = f_1^* + f_2^*. \tag{5}$$

2.2. Smoothing

Smoothing of non-smooth functions is usually motivated for algorithmic purposes when solving non-smooth optimization problems [9–11]. In this paper, contrary to the classical motivation, smoothing is employed to generate non-convex sparse penalties. Before discussing the relation between sparsity-inducing convex penalties and smoothing, we provide a concise summary of smoothing techniques.

The *Moreau envelope* or *Moreau-Yosida regularization* is a traditional smoothing approach defined as

$$h_\mu^M(x) \triangleq \inf_v \left\{ h(v) + 1/(2\mu)\|x - v\|_2^2 \right\}. \tag{6}$$

The Moreau envelope has many nice properties, *e.g.*, if $f \in \Gamma_0(\mathbb{R}^n)$ then f_μ^M is convex continuous, finite-valued, differentiable with continuous Lipschitz gradient, and $f_\mu^M \in \Gamma_0(\mathbb{R}^n)$ [23]. As a classical example, the Moreau envelope of the absolute value function, $f(x) = |x|$, is the Huber function [25] which is given as

$$h_\mu^M(x) = H_\mu(x) = \begin{cases} x^2/(2\mu) & |x| \leqslant \mu \\ |x| - \mu/2 & |x| \geqslant \mu. \end{cases} \tag{7}$$

Several smoothing approaches that generalize Moreau envelope were proposed in the literature such as Nesterov's [11] and infimal convolution smoothing [9]. Towards our end goal of designing sparsity-inducing non-convex penalties, we propose a generalized smoothing technique.

3. SHARPENING SPARSE REGULARIZERS

The SSR framework depends on the synergy between sparsity-inducing convex penalties and smoothing. The designed penalties are defined by subtracting from the sparse non-smooth convex penalty its smoothed version. As an example, the minimax-concave (MC) penalty can be written as

$$\Psi_\mu^{\mathrm{MC}}(x) = \|x\|_1 - \sum_{i=1}^n H_\mu(x_i). \tag{8}$$

where H_μ is the Huber function (7).

We propose a generalized infimal convolution smoothing (GICS) technique that includes the Huber function as a special case. The proposed smoothing technique an extension to the classical infimal convolution smoothing [9]. Then, we define the SSR framework penalty as

$$\Psi_\mu^B(Wx) \triangleq \|Wx\|_1 - \tilde{h}_{\mu,B}(x) \tag{9}$$

where $\tilde{h}_{\mu,B}$ is defined in (11) as the GICS smoothing of $\|Wx\|_1$. Next, we formularize this concept and provide various examples and interpretations.

Definition 1 *Let the convex regularizer be*

$$h(x) = \|Wx\|_1 \tag{10}$$

and $B \in \mathbb{R}^{q\times n}$ be the steering matrix, $\phi : \mathbb{R} \to \mathbb{R}$ be a smoothing kernel that is a convex function continuously differentiable with Lipschitz gradient constant $1/\sigma$ ($\sigma > 0$), and without loss of generality let $\min_z \phi^*(z) = 0$. Then, we define the generalized infimal convolution smoothing (GICS) as

$$\tilde{h}_{\mu,B}(x) \triangleq \inf_v \left\{ h(v) + \Phi_\mu\left(B(x - v)\right) \right\} \tag{11}$$

$$\triangleq \left(h \,\square\, (\Phi_\mu \circ B)\right)(x) \tag{12}$$

where $\Phi(Bx) \triangleq \sum_{i=1}^q \phi(b_i^T x)$, and Φ_μ is the epi-multiplication operation defined in (3)

It is straight forward to see that if $B = I$, the GICS reduces to the plain infimal convolution smoothing [9]. Our generalization is non-trivial in the context of generating non-seperable non-convex penalties that preserve convexity even if the data fidelity term Hessian, $A^T A$, is singular. Furthermore, if $\phi(x) = (1/2)\, x^2$, then GICS gives the Moreau envelope (6), and more interestingly, the GICS recovers the generalized Huber function as a special case [7].

3.1. Properties

Here we show properties of GICS smoothing (11), and consequently the SSR penalty (9). The following proposition addresses the conditions under which the GISC inherits the interesting properties of Moreau envelope.

Proposition 1 *Let the smoothing kernel ϕ be coercive and $\mathcal{N}(W) \cap \mathcal{N}(B) = \{0\}$. Then, the infimal convolution of GICS* (11) *is exact (i.e., for any x the infimum is attained for some v), and $\tilde{h}_{\mu,B}$ is convex continuous, finite-valued, which implies that $\tilde{h}_{\mu,B} \in \Gamma_0(\mathbb{R}^n)$.*

Besides the infimal convolution formulation, an equivalent representation of GICS is the Fenchel conjugate or dual formulation using the property (5).

Proposition 2 *Given the properties of GICS in Proposition 1, GICS can be rewritten in a Fenchel conjugate or dual formulation*

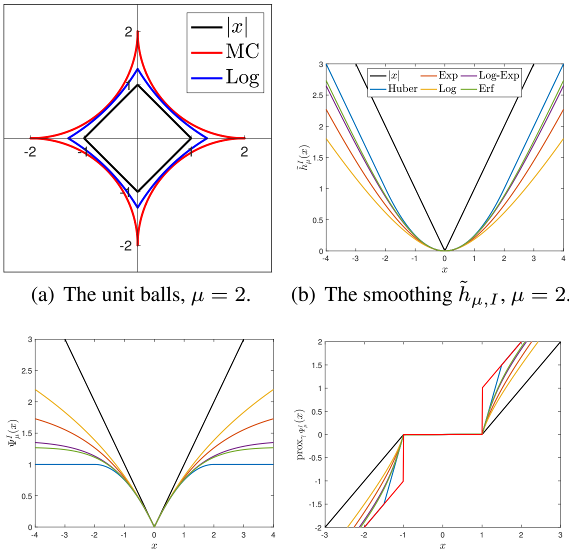
$$\tilde{h}_{\mu,B}(x) = \sup_z \left\{ z^T Bx - h^*(B^T z) - \mu\, \Phi^*(z) \right\} \tag{13}$$

As expected, when $B = I$, then the GICS dual formulation in (13) reduces to the Nesterov's smoothing [11]. We next address the gradient of GICS.

Proposition 3 *Let $v_{\mu,B}^\star(x)$ and $z_{\mu,B}^\star(x)$ be the solutions of the infimal convolution* (11) *and the dual* (13) *formulations, respectively. Then, the GICS technique, $\tilde{h}_{\mu,B}$, is differentiable and the gradient is obtained as*

$$\nabla \tilde{h}_{\mu,B}(x) = B^T z_{\mu,B}^\star(x) = B^T \nabla \Phi_\mu\left(B(x - v_{\mu,B}^\star(x))\right) \tag{14}$$

with Lipschitz constant $\tilde{L}_{\mu,B} = \|B\|_2^2/(\mu\sigma)$.



(a) The unit balls, $\mu = 2$. (b) The smoothing $\tilde{h}_{\mu,I}$, $\mu = 2$.

(c) The SSR penalty, $\mu = 2$. (d) The proximal operators with $\gamma = 1, \mu = 1.5$.

Fig. 1: Examples of non-convex sparse penalties generated by the proposed SSR framework.

Interestingly, when $B = I$ and $\phi(x) = (1/2)x^2$, then (14) reduces to the celebrated proximal operator identity

$$\mu \operatorname{prox}_{\mu^{-1}h^*}(x/\mu) = x - \operatorname{prox}_{\mu h}(x) \tag{15}$$

where $v^\star_{\mu,I} = \operatorname{prox}_{\mu h}$ and $z^\star_{\mu,I} = \operatorname{prox}_{\mu^{-1}h^*}$.

Furthermore, we can show that GISC has the following upper and lower bounds properties

$$0 \leqslant \tilde{h}_{\mu,B}(x) \leqslant \|Wx\|_1, \tag{16}$$

consequently, the SSR framework penalties have the following bounds

$$0 \leqslant \Psi^B_\mu(Wx) \leqslant \|Wx\|_1. \tag{17}$$

3.2. Examples and Interpretations

Figures 1(a) shows the contour of ℓ_1-norm and two non-convex penalties, MC and logarithm, that are recovered by the SSR as special cases. Note that the edges of non-convex penalties are sharper than the ℓ_1-norm unit ball, thus, the name sharpening spare regularizers (SSR).

In order to demonstrate the generalization power of the SSR framework, consider the simple case of $B = W = I$. Figures 1(b) and 1(c) illustrates several examples of the the smoothed functions $\tilde{h}^I_{\mu,I}$ and non-convex sparse penalties Ψ^I_μ. These penalties are obtained by varying the smoothing kernel ϕ. The framework is capable of recovering several well known non-convex sparse penalties in the literature such as minimax concave (MC) [3], logarithm [4] and exponential [8] penalties. In addition, it generates new non-convex sparse penalties that are based on log-sum-exp, error function, and inverse tangent function. That is, the SSR framework is a unified approach for creating non-convex sparse-inducing penalties.

Moreover, looking at the corresponding proximal operator of each penalty provides further insight. The proximal operator for a proper lsc function $g : \mathbb{R}^n \to \mathbb{R} \cup \{\pm\infty\}$ and parameter $\gamma > 0$ is defined as

$$\operatorname{prox}_{\gamma h}(x) = \arg\min_v \left\{h(v) + 1/(2\gamma)\|x - v\|_2^2\right\} \tag{18}$$

The next proposition shows the condition under which $\operatorname{prox}_{\gamma \Psi^B_\mu}$ is single-valued.

Proposition 4 *The operator $\operatorname{prox}_{\gamma\,\Psi^B_\mu}$ is well-defined, and is a single-valued mapping if and only if $\mu\sigma/\|B\|_2^2 > \gamma$.*

Figure 1(d) shows the proximal operators of the previous examples. Observe that $\operatorname{prox}_{\gamma\,\Psi^I_\mu}$ approximates the hard-thresholding operator ($\operatorname{prox}_{\gamma\,\ell_0}$) better than the soft-thresholding operator ($\operatorname{prox}_{\gamma\,\ell_1}$). In fact, as $\mu \to \infty$, we have that $\operatorname{prox}_{\gamma\,\Psi^I_\mu} \to \operatorname{prox}_{\gamma\,\ell_1}$.

3.3. Convexity conditions

One of the main advantages of the SSR framework is that despite using non-convex penalties, it can preserve the convexity of the cost function (1), thus global convergence can be obtained. The next theorem addresses the condition under which the linear inverse problem (1) with SSR penalty remains convex.

Theorem 1 *Consider a regularized least square linear inverse problem* (1) *with SSR penalty* (9). *Then, the cost function*

$$F(x) = (1/2)\|y - Ax\|_2^2 + \lambda\,\Psi^B_\mu(Wx) \tag{19}$$

is convex if

$$(1/2)\|y - Ax\|_2^2 - \frac{\lambda}{2\sigma\mu}\|Bx\|_2^2 \quad \textit{is convex,} \tag{20}$$

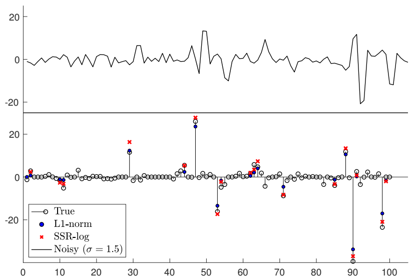
i.e., the data fidelity term is $\lambda/(\sigma\mu)$-strongly convex relative to $(1/2)\,\|Bx\|_2^2$.

Note in particular that the data fidelity term does not need to be strongly convex (singular A is allowed). Instead, the requirement is relaxed using the steering matrix, B, which introduces non-convex penalties along directions where there is additional curvature to exploit. In other words, relative strong convexity of the data fidelity term is required [26,27], but not classical strong convexity. The convexity condition (20) can be rewritten as

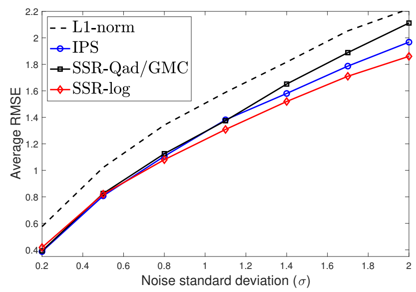
$$A^T A \succeq \lambda/(\sigma\mu)\, B^T B. \tag{21}$$

This suggests it is reasonable to select B as a scaled version of A such as $B = \sqrt{\gamma\sigma\mu/\lambda}\,A$, where $0 \leqslant \gamma \leqslant 1$ is a non-convexity parameter. This selection is not necessarily the best, better selection schemes are on our search agenda.

Remark 1 *The SSR framework penalty Ψ^B_μ will generally be non-separable when the steering matrix B is non-diagonal. This non-separability is a requirement to be able to introduce non-convexity and maintain the cost function convexity. Without this property, whenever A is singular, the only sparse penalty that preserves the cost function convexity is the convex ℓ_1-norm.*



(a) Sparse deconvolution example at $\sigma = 1.5$.



(b) Average RMSE as a function of σ

Fig. 2: Sparse deconvolution example.

4. OPTIMIZATION ALGORITHMS

The SSR regularizer (9) does not always have a closed-form for arbitrary W, B, and ϕ. Similarly, evaluating the SSR gradient (14) or proximal operator (18) requires solving an optimization problem at each iteration. Nonetheless, given the convexity condition are satisfied, the RLS linear inverse problem (1) with SSR regularizer can be solved globally using an efficient first-order algorithm, which is based on a generalization of the primal-dual algorithms [28,29]. The optimization algorithm is based on reformulating the SSR-RLS as a saddle-point problem such as

$$\inf_{x} \sup_{v} \left\{ \mathcal{L}(x, v) + \lambda \left(\|Wx\|_1 - \|Wv\|_1 \right) \right\} \tag{22}$$

where $\mathcal{L}(x, v) = (1/2)\|y - Ax\|_2^2 - \lambda \, \Phi_\mu(B(x - v))$ is a differentiable function, convex in x and concave in v. The only required assumption is that $\|Wx\|_1$ has an efficient proximal operator, e.g., ℓ_1-norm, 1D-TV, or tight frame W. This problem structure lends itself to the generalization of the classical primal-dual algorithm [29]. In case the proximal operator can not be efficiently computed, other efficient forward-backward algorithms can used, but this is beyond the scope of this paper.

5. NUMERICAL EXAMPLES

The first example considers sparse deconvolution. The original sparse signal x is generated from a mixed Bernoulli-double Pareto distribution as shown in Figure 2(a). Then, the signal is filtered with a 5^{th}-order bandpass FIR with linear phase and corrupted with zero mean Gaussian noise. We let the noise standard deviation span the interval $0.2 \leqslant \sigma \leqslant 2$. The parameters are set as following: $\lambda = \beta \, \sigma \|h\|_2$, where h is the filter impulse response, $\beta \approx 2$, μ is set in each case to minimize the root-mean-square-error (RMSE), $B = \sqrt{\gamma\sigma\mu/\lambda}\; A$, and $\gamma = 0.95$. We compare among the ℓ_1-norm, the iterative p-shrinkage (IPS) penalties [30], and the SSR penalty with logarithm and quadratic smoothing kernels. The average RMSE is calculated for 100 independent realizations. Figure 2(b) illustrates that non-convex penalties outperform the ℓ_1-norm solution. Moreover, as expected from Bayesian perspective, the SSR-log outperforms the GMC penalty (SSR-Quad) [7] and IPS penalties, especially with high noise level.

The second example is wavelet-based image restoration. The images are first scaled into the range $[0, 1]$, then blurred with Gaussian filter of size 5×5 and standard deviation of 1 followed by an additive zero-mean white Gaussian noise with standard deviation $\sigma = 0.05$. We solve a synthesis inverse problem in (1) with $W = I$ and $A = HW^H$, where H is the Gaussian blur linear operator and W^H is the inverse 4-scales 2D dual-tree complex wavelet transform (DT-CWT) operator. We set the regularizer parameter λ as scaled value of σ to give the highest peak signal-to-noise ration (PSNR) for each algorithm. The image restoration results are shown in Figure 3. As expected the SSR non-convex penalties outperform the ℓ_1-norm solution. In general, we were able to gain 1 dB in PSNR while maintaining the cost function convexity.

Fig. 3: Wavelet-based image restoration example.

6. CONCLUSION

This paper proposes a unified framework to design non-convex penalties that induce sparsity more effectively than the classical ℓ_1-norm, but without sacrificing convexity of the cost function. We call the framework *sharpening sparse regularizers* (SSR). The construction process resembles high-pass filtering in signal processing. The SSR framework is applicable to any sparsity regularized least square ill-posed linear inverse problem. Furthermore, it generates new penalties as well as recovering and generalizing several well known non-convex sparse penalties in the literature. The SSR framework theory generalizes our recent generalized minimax-concave (GMC) work [7], and numerical results demonstrate that the new framework can yields better results than GMC. In summary, the SSR framework asserts that under suitable conditions, the ℓ_1-norm can be outperformed without sacrificing the convexity of the cost function.

7. REFERENCES

[1] Jean-Luc Starck, Fionn Murtagh, and Jalal Fadili, *Sparse image and signal processing: Wavelets and related geometric multiscale analysis*, Cambridge university press, 2015.

[2] Robert Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 267–288, 1996.

[3] Cun-Hui Zhang et al., "Nearly unbiased variable selection under minimax concave penalty," *The Annals of statistics*, vol. 38, no. 2, pp. 894–942, 2010.

[4] Emmanuel J Candes, Michael B Wakin, and Stephen P Boyd, "Enhancing sparsity by reweighted ℓ_1 minimization," *Journal of Fourier analysis and applications*, vol. 14, no. 5-6, pp. 877–905, 2008.

[5] Rick Chartrand, "Shrinkage mappings and their induced penalty functions," in *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*. IEEE, 2014, pp. 1026–1029.

[6] Ali Gholami and S Mohammad Hosseini, "A general framework for sparsity-based denoising and inversion," *IEEE transactions on signal processing*, vol. 59, no. 11, pp. 5202–5211, 2011.

[7] I. Selesnick, "Sparse regularization via convex analysis," *IEEE Transactions on Signal Processing*, vol. 65, no. 17, pp. 4481–4494, Sept 2017.

[8] Mariano Rivera and Jose L Marroquin, "Efficient half-quadratic regularization with granularity control," *Image and Vision Computing*, vol. 21, no. 4, pp. 345–357, 2003.

[9] Amir Beck and Marc Teboulle, "Smoothing and first order methods: A unified framework," *SIAM Journal on Optimization*, vol. 22, no. 2, pp. 557–580, 2012.

[10] Jean-Jacques Moreau, "Proximité et dualité dans un espace hilbertien," *Bull. Soc. Math. France*, vol. 93, no. 2, pp. 273–299, 1965.

[11] Yu Nesterov, "Smooth minimization of non-smooth functions," *Mathematical programming*, vol. 103, no. 1, pp. 127–152, 2005.

[12] Andrew Blake and Andrew Zisserman, *Visual reconstruction*, MIT press, 1987.

[13] Mila Nikolova, "Energy minimization methods," in *Handbook of mathematical methods in imaging*, pp. 139–185. Springer, 2011.

[14] Mila Nikolova, Michael K Ng, and Chi-Pan Tam, "Fast nonconvex nonsmooth minimization methods for image restoration and reconstruction," *IEEE Transactions on Image Processing*, vol. 19, no. 12, pp. 3073–3088, 2010.

[15] P. Chen and I. W. Selesnick, "Group-sparse signal denoising: Non-convex regularization, convex optimization," *IEEE Transactions on Signal Processing*, vol. 62, no. 13, pp. 3464–3478, July 2014.

[16] Alessandro Lanza, Serena Morigi, and Fiorella Sgallari, "Convex image denoising via non-convex regularization with parameter selection," *Journal of Mathematical Imaging and Vision*, vol. 56, no. 2, pp. 195–220, 2016.

[17] M. Malek-Mohammadi, C. R. Rojas, and B. Wahlberg, "A class of nonconvex penalties preserving overall convexity in optimization-based mean filtering," *IEEE Transactions on Signal Processing*, vol. 64, no. 24, pp. 6650–6664, Dec 2016.

[18] A. Parekh and I. W. Selesnick, "Enhanced low-rank matrix approximation," *IEEE Signal Processing Letters*, vol. 23, no. 4, pp. 493–497, April 2016.

[19] I. W. Selesnick and Ilker Bayram, "Sparse signal estimation by maximally sparse convex optimization," *IEEE Transactions on Signal Processing*, vol. 62, no. 5, pp. 1078–1092, March 2014.

[20] Ivan Selesnick, "Total variation denoising via the moreau envelope," *IEEE Signal Processing Letters*, vol. 24, no. 2, pp. 216–220, 2017.

[21] R Tyrrell Rockafellar and Roger J-B Wets, *Variational analysis*, vol. 317, Springer Science & Business Media, 2009.

[22] Ralph Tyrell Rockafellar, *Convex analysis*, Princeton university press, 2015.

[23] Heinz H Bauschke, Patrick L Combettes, et al., *Convex analysis and monotone operator theory in Hilbert spaces*, vol. 2011, Springer, 2017.

[24] Patrick L Combettes and Jean-Christophe Pesquet, "Proximal splitting methods in signal processing," in *Fixed-point algorithms for inverse problems in science and engineering*, pp. 185–212. Springer, 2011.

[25] Peter J Huber et al., "Robust estimation of a location parameter," *The annals of mathematical statistics*, vol. 35, no. 1, pp. 73–101, 1964.

[26] Haihao Lu, Robert M Freund, and Yurii Nesterov, "Relatively smooth convex optimization by first-order methods, and applications," *SIAM Journal on Optimization*, vol. 28, no. 1, pp. 333–354, 2018.

[27] Heinz H Bauschke, Jérôme Bolte, and Marc Teboulle, "A descent lemma beyond lipschitz gradient continuity: first-order methods revisited and applications," *Mathematics of Operations Research*, vol. 42, no. 2, pp. 330–348, 2016.

[28] Nikos Komodakis and Jean-Christophe Pesquet, "Playing with duality: An overview of recent primal-dual approaches for solving large-scale optimization problems," *IEEE Signal Processing Magazine*, vol. 32, no. 6, pp. 31–54, 2015.

[29] Erfan Yazdandoost Hamedani and Necdet Serhat Aybat, "A primal-dual algorithm for general convex-concave saddle point problems," *arXiv preprint arXiv:1803.01401*, 2018.

[30] Joseph Woodworth and Rick Chartrand, "Compressed sensing recovery via nonconvex shrinkage penalties," *Inverse Problems*, vol. 32, no. 7, pp. 075004, 2016.