Convex Denoising using Non-Convex Tight Frame Regularization

Ankit Parekh* and Ivan W. Selesnick†

April 7, 2015

Abstract

This paper considers the problem of signal denoising using a sparse tight-frame analysis prior. The $\ell_1$ norm has been extensively used as a regularizer to promote sparsity; however, it tends to under-estimate non-zero values of the underlying signal. To more accurately estimate non-zero values, we propose the use of a non-convex regularizer, chosen so as to ensure convexity of the objective function. The convexity of the objective function is ensured by constraining the parameter of the non-convex penalty. We use ADMM to obtain a solution and show how to guarantee that ADMM converges to the global optimum of the objective function. We illustrate the proposed method for 1D and 2D signal denoising.

1 Introduction

A standard technique for estimating sparse signals is through the formulation of an inverse problem with the $\ell_1$ norm as convex proxy for sparsity [31]. In particular, consider the problem of estimating a signal $x \in \mathbb{R}^n$ from a noisy observation $y \in \mathbb{R}^n$,

$$ y = x + w, $$

where $w$ represents zero mean additive white Gaussian noise. We assume the underlying signal to be sparse with respect to an overcomplete tight-frame $A \in \mathbb{R}^{m \times n}$, $m \geq n$, which satisfies the Parseval frame condition, i.e.,

$$ A^T A = rI, \quad r > 0. $$

---

* A. Parekh (ankit.parekh@nyu.edu) is with the Department of Mathematics, School of Engineering, New York University, Brooklyn, NY.

† I. Selesnick (selesi@nyu.edu), is with the Department of Electrical and Computer Engineering, School of Engineering, New York University, Brooklyn, NY.
Using an analysis prior, we formulate the signal denoising problem as
\[
\arg\min_x \left\{ F(x) := \frac{1}{2}\|y - x\|^2_2 + \sum_{i=1}^m \lambda_i \phi([Ax]_i; a_i) \right\},
\] (3)
where \(\lambda_i > 0\) are the regularization parameters, and \(\phi: \mathbb{R} \rightarrow \mathbb{R}\) is a non-smooth sparsity inducing penalty function. The parameter \(a_i\) controls the non-convexity of \(\phi\) in case it is non-convex. The analysis prior is used in image processing and computer vision applications [6, 7, 15, 26, 28, 32, 34]. Commonly, the \(\ell_1\) norm is used to induce sparsity, i.e., \(\phi(x) = |x|\) [10]. In that case, problem (3) is strictly convex and the global optimum can be reliably obtained.

The \(\ell_1\) norm is not the tightest envelope of sparsity [19]. It under-estimates the non-zero values of the underlying signal [8, 22]. Non-convex regularizers estimate the significant values more accurately. Non-convex regularization in an analysis model has been used for MRI reconstruction [9], EEG signal reconstruction [21], and for computer vision problems [25]. However, the use of non-convex regularizers comes at a price: the objective function is generally non-convex. Consequently, several issues arise (spurious local minima, a perturbation of the input data can change the solution unpredictably, convergence is guaranteed to the local minima only, etc.).

In order to maintain convexity of the objective function while using non-convex regularizers, we propose to restrict the parameter \(a_i\) of the non-convex regularizer \(\phi\). By controlling the degree of non-convexity of the regularizer we guarantee that the total objective function \(F\) is convex. This idea which dates to Blake and Zisserman [3], and Nikolova [22], has been applied to image restoration and reconstruction [23, 24], total variation denoising [29], and wavelet denoising [13].

In this paper we provide a critical value of parameter \(a\) to ensure \(F\) in (3) is strictly convex (even though \(\phi\) is non-convex). In contrast to the above works, we consider transform domain regularization and prove that ADMM [5] applied to the problem (3) converges to the global optimum. The convergence of ADMM is guaranteed, provided the augmented Lagrangian parameter \(\mu > 1/r\).

2 Sparse Signal Estimation

2.1 Non-convex Penalty Functions

In order to induce sparsity strongly than the \(\ell_1\) norm, we use non-convex penalty functions \(\phi: \mathbb{R} \rightarrow \mathbb{R}\) parameterized by the parameter \(a \geq 0\). We make the following assumption of such penalty functions.

Assumption 1. The non-convex penalty function \(\phi: \mathbb{R} \rightarrow \mathbb{R}\) satisfies the following

1. \(\phi\) is continuous on \(\mathbb{R}\), twice differentiable on \(\mathbb{R} \setminus \{0\}\) and symmetric, i.e., \(\phi(-x; a) = \phi(x; a)\)
2. $\phi'(x) > 0, \forall x > 0$

3. $\phi''(x) \leq 0, \forall x > 0$

4. $\phi'(0^+) = 1$

5. $\inf_{x \neq 0} \phi''(x; a) = \phi''(0^+; a) = -a$

6. $\phi(x; 0) = |x|.$

Since $\phi(x; 0) = |x|$, the $\ell_1$ norm is recovered as a special case of the penalty function $\phi$. The parameter $a$ controls the degree of non-convexity of the penalty function $\phi$. Note that the $\ell_p$ norm does not satisfy assumption 1. The rational penalty function [17],

$$\phi(x; a) = \frac{|x|}{1 + a|x|/2},$$

(4)

the logarithmic, and the arctangent penalty functions [8, 27] are examples that satisfy Assumption 1. The rational penalty $\phi$ for several values of $a$ is shown in Fig. 1(a).

The proximity operator of $\phi$ [12], $\text{prox}_\phi : \mathbb{R} \to \mathbb{R}$, is defined as,

$$\text{prox}_\phi(y; \lambda, a) := \arg \min_{x \in \mathbb{R}} \left\{ \frac{1}{2} (y - x)^2 + \lambda \phi(x; a) \right\}.$$  

(5)

For $\phi(x; a)$ satisfying Assumption 1, with $a < 1/\lambda$, the proximity operator is a continuous non-linear threshold function with $\lambda$ as the threshold value, i.e., $\text{prox}_\phi(y; \lambda, a) = 0, \forall |y| < \lambda$. The proximity operator of the absolute value function is the soft-thresholding function. There is a constant gap between the identity function and the soft-threshold function due to which the non-zero values are underestimated [16]. On the other hand, non-convex penalty functions satisfying Assumption 1 are specifically designed so that the threshold function approaches identity asymptotically. The non-convex penalty functions do not underestimate large values when $a > 0$.

2.2 Convexity Condition

In order to benefit from convex optimization principles in solving (3), we seek to ensure $F$ in (3) is convex by controlling the parameter $a_i$. For later, we note the following lemma.

**Lemma 1.** Let $\phi : \mathbb{R} \to \mathbb{R}$ satisfy Assumption 1. The function $s : \mathbb{R} \to \mathbb{R}$ defined as

$$s(x; a) := \phi(x; a) - |x|,$$

(6)
is twice continuously differentiable and concave with

\[-a \leq s''(x; a) \leq 0. \tag{7}\]

**Proof.** Since \(\phi\) and the absolute value function are twice continuously differentiable on \(\mathbb{R} \setminus \{0\}\), we need only show \(s'(0^+) = s'(0^-)\) and \(s''(0^+) = s''(0^-)\). From assumption 1, we have \(\phi'(0^+) = 1\), hence \(s'(0^+) = \phi'(0^+) - 1 = 0\). Again by assumption 1 we have \(\phi'(0^-) = -\phi'(0^+) = -1\), hence \(s'(0^-) = \phi'(0^-) + 1 = 0\). Further, \(s''(0^+) = \phi''(0^+)\) and \(s''(0^-) = \phi''(0^-) = \phi''(0^+) = s''(0^+)\). Thus the function \(s\) is twice continuously differentiable. The function \(s\) is concave since \(s''(x) = \phi''(x) \leq 0\), \(\forall x \neq 0\). Using Assumption 1 it follows that \(-a \leq s''(x; a) \leq 0\). \(\square\)

Figure 1(b) displays the function \(s(x; a)\), which is twice continuously differentiable even though \(\phi\) in Fig. 1(a) is not differentiable. The following theorem states the critical value of parameter \(a_i\) to ensure the convexity of \(F\) in (3).

**Theorem 1.** Let \(\phi(x; a)\) be a non-convex penalty function satisfying Assumption 1 and \(A\) be a transform satisfying \(A^T A = rI\), \(r > 0\). The function \(F : \mathbb{R}^n \rightarrow \mathbb{R}\) defined in (3) is strictly convex if

\[0 \leq a_i < \frac{1}{r\lambda_i}. \tag{8}\]

**Proof.** Consider the function \(G : \mathbb{R}^n \rightarrow \mathbb{R}\) defined as

\[G(x) := \frac{1}{2}\|y - x\|^2 + \sum_{i=1}^{m} \lambda_i s([Ax]_i; a_i). \tag{9}\]

Since \(G\) is twice continuously differentiable (using Lemma 1), the Hessian of \(G\) is given by

\[\nabla^2 G(x) = I + A^T \text{diag} (\lambda_1 d_1, \ldots, \lambda_m d_m) A, \tag{10}\]
where $d_i = s''([Ax]_i; a_i)$. Using (2), we write the Hessian as

$$
\nabla^2 G(x) = A^T \left( \frac{1}{r} I + \text{diag}(\lambda_1 d_1, \ldots, \lambda_m d_m) \right) A.
$$

(11)

$$
= A^T \text{diag} \left( \frac{1}{r} + \lambda_1 d_1, \ldots, \frac{1}{r} + \lambda_m d_m \right) A.
$$

(12)

The transform $A$ has full column rank, from (2), hence $\nabla^2 G(x)$ is positive definite if

$$
\frac{1}{r} + \lambda_i d_i > 0, \quad i = 1, \ldots, m.
$$

(13)

Thus, $\nabla^2 G(x)$ is positive definite if

$$
s''([Ax]_i; a_i) > -\frac{1}{r \lambda_i}.
$$

(14)

Using Lemma 1, we obtain the critical value of $a_i$ to ensure the convexity of $G$, i.e.,

$$
0 \leq a_i < \frac{1}{r \lambda_i}.
$$

(15)

It is straightforward that

$$
F(x) = G(x) + \sum_{i=1}^{m} \lambda_i |[Ax]_i|.
$$

(16)

Thus, being a sum of a strictly convex function and a convex function, $F$ is strictly convex.

Note that if $a_i > 1/(r \lambda_i)$, then the function $G(x)$ is not convex, as the Hessian of $G(x)$ is not positive definite. As a result, $1/(r \lambda_i)$ is the critical value of $a_i$ to ensure the convexity of the function $F$. The following corollary provides a convexity condition for the situation where the same regularization parameter is applied to all coefficients.

**Corollary 1.** For $\lambda_i = \lambda, i = 1, \ldots, m$, the function $F$ in (3) is strictly convex if $0 \leq a_i < 1/(r \lambda)$.

□

We illustrate the convexity condition using a simple example with $n = 2$. We set

$$
A^T = \begin{bmatrix}
1 & 1 & 1 & 1 \\
1 & 1 & -1 & -1 \\
\end{bmatrix}, \quad A^T A = 4I,
$$

(17)

and $\lambda_1 = \lambda_2 = 1$. Theorem 1 states that the function $G$ defined in (9) is strictly convex for $a_i < 1/4$ and non-convex for $a_i > 1/4$. It can be seen in Fig. 2 that the function $G$ is convex for $a_i = 0.25$,
Figure 2: Surface plots of the rational penalty function and the function $G$, for two different values of $a$.

Table 1: Iterative algorithm for the solution to (3).

| Input: $y, \lambda_i, r, a_i, \mu$ |
|------------------------------------|
| Initialization: $u = 0, d = 0$     |
| Repeat:                           |
| $x \leftarrow \frac{1}{1 + \mu r}(y + \mu A^H(u - d))$ |
| $u_i \leftarrow \text{prox}_\phi([Ax + d]_i; \lambda_i/\mu_i, a_i)$ |
| $d \leftarrow d - (u - Ax)$        |
| Until convergence                 |

even though the penalty function is not convex. However, when $a_i > 0.25$, the function $G$ (hence $F$) is non-convex.

3 Algorithm

A benefit of ensuring convexity of the objective function is that we can utilize convex optimization approaches to obtain the solution. In particular, for $\phi(x) = |x|$, the widely used methods for solving (3) are proximal methods [12] and ADMM [5, 18].

The convergence of ADMM to the optimum solution is guaranteed when the functions appearing in the objective function are convex [14]. The following theorem states that ADMM can be used to solve (3) with guaranteed convergence, provided the augmented Lagrangian parameter $\mu$ is appropriately set. Note that $\mu$ does not affect the solution to which ADMM converges, rather the speed at which it converges.

**Theorem 2.** Let $\phi$ satisfy Assumption 1 and the transform $A$ satisfy the Parseval frame condition (2). Let $a_i < 1/(r_i \lambda_i)$. The iterative algorithm in Table 1 converges to the global minimum of the
function $F$ in (3) if

$$\mu > \frac{1}{r}. \quad (18)$$

**Proof.** We re-write the problem (3) using variable splitting [1] as

$$\arg \min_{u,x} \left\{ \frac{1}{2} \| y - x \|_2^2 + \sum_{i=1}^{m} \lambda_i \phi(u_i; a_i) \right\} \quad (19a)$$

s.t. $u = Ax \quad (19b)$

The minimization is separable in $x$ and $u$. Applying ADMM to (19) yields the following iterative procedure with the augmented Lagrangian parameter $\mu$.

$$x \leftarrow \arg \min_{x} \left\{ \frac{1}{2} \| y - x \|_2^2 + \frac{\mu}{2} \| u - Ax - d \|_2^2 \right\} \quad (20a)$$

$$u \leftarrow \arg \min_{u} \left\{ \sum_{i=1}^{m} \lambda_i \phi(u_i; a_i) + \frac{\mu}{2} \| u - Ax - d \|_2^2 \right\}_{R(u)} \quad (20b)$$

$$d \leftarrow d - (u - Ax) \quad (20c)$$

The sub-problem (20a) in $x$ can be solved explicitly as

$$x = (I + \mu A^H A)^{-1} (y + \mu A^H (u - d)) \quad (21)$$

$$= \frac{1}{1 + \mu r} \left( y + \mu A^H (u - d) \right), \quad (22)$$

using (2). The sub-problem (20b) in $u$ can be solved using $\text{prox}_\phi$, provided the function $R$ is convex. Consider the function $Q: \mathbb{R}^m \rightarrow \mathbb{R}$ defined as

$$Q(u) := \sum_{i=1}^{m} \lambda_i s(u_i; a_i) + \frac{\mu}{2} \| u - Ax - d \|_2^2. \quad (23)$$

From Lemma 1 and the proof of Theorem 1, $\nabla^2 Q(u)$ is positive definite if

$$s''(u_i; a_i) > \frac{-\mu}{\lambda_i} \quad \Rightarrow \quad \mu > a_i \lambda_i. \quad (24)$$

Since $a_i < 1/(r \lambda_i)$, it follows that $\nabla^2 Q(u)$ is positive definite if $\mu > 1/r$. Hence $Q$ is strictly convex for $\mu > 1/r$. Note that $R(u) = Q(u) + \| u \|_1$. Hence, the function $R$, being the sum of a convex and
a strictly convex function, is convex. As such, the minimization problem in (20b) is well-defined and its solution can be efficiently computed using the proximity operator of $\phi$ (5), i.e.,

$$u_i \leftarrow \text{prox}_\phi \left( [Ax + d]_i; \lambda_i/\mu_i, a_i \right).$$

(25)

Since $A$ has full column rank with $\mu$ being sufficiently large, ADMM converges to a stationary point of the objective function (despite having a non-convex function in the objective) [2,4,20,33]. Moreover, the function $F$ is convex (by Theorem 1) and the sub-problems of the ADMM are strictly convex for $\mu > 1/r$. As a result, the iterative procedure (20) converges to the global minimum of $F$.

4 Examples

4.1 1D Signal Denoising

We consider the problem of denoising a 1D signal that is sparse with respect to the undecimated wavelet transform [11], which satisfies the condition (2) with $r = 1$. In particular, we use a 4-scale undecimated wavelet transform with three vanishing moments. The noisy signal is generated using Wavelab (http://www-stat.stanford.edu/~wavelab/) with additive white Gaussian noise (AWGN) of $\sigma = 4.0$. We set the regularization parameters $\lambda_j = \beta \sigma 2^{-j/2}, 1 \leq j \leq 4$. We use the same $\lambda_j$ for all the coefficients in scale $j$. The value of $\beta$ is chosen to obtain the lowest RMSE for the convex and the non-convex regularization respectively. To maximally induce sparsity we set $a_i = 1/\lambda_i$. For the 1D signal denoising example, we use the non-convex arctangent penalty and its corresponding threshold function [27].

Figure 3 shows that the denoised signal obtained using non-convex regularization has a lower RMSE while preserving the discontinuities. Further, the peaks are less attenuated compared to the denoised signal obtained using $\ell_1$ norm regularization.

For further comparison, we generate the noisy signal in Fig. 3 for $1 \leq \sigma \leq 4$, and denoise it with non-convex and convex regularization. We also denoise the noisy signal by direct non-linear thresholding of the noisy wavelet coefficients. We use the same $\beta$ values as in Fig. 3 for non-convex and convex regularization. The value of $\beta$ for direct non-linear thresholding is also chosen to obtain the lowest RMSE. As seen in Fig. 4, non-convex regularization consistently yields the lowest RMSE. The RMSE values are obtained by averaging over 15 realizations for each $\sigma$. 

8
Figure 3: 1D denoising example. Non-convex regularization yields lower RMSE than convex regularization.

Figure 4: RMSE values as a function of the noise level $\sigma$ for the 1D signal denoising example.

Figure 5: Image denoising. Wavelet artifacts are more prominent when using $\ell_1$ norm regularization.

4.2 2D Image Denoising

We consider the problem of denoising a 2D image corrupted with AWGN. We use the 2D dual-tree complex wavelet transform (DT-CWT) [30], which is 4-times expansive and satisfies (2) with $r = 1$. The noisy ‘peppers’ image has peak signal-to-noise ratio (PSNR) value of 14.6 dB. We use the same $\lambda$ for all the sub-bands. As in the previous example, we set the value of $\lambda$ for each case (convex and non-convex) as a constant multiple of $\sigma$, which gives the highest PSNR.

Figure 5 shows that the denoised image (non-convex case) contains fewer wavelet artifacts and
Figure 6: Relative performance of convex and non-convex regularization for image denoising. (a) PSNR as a function of $\lambda$. (b) PSNR as a function of $\sigma$.

has a higher PSNR. Figure 6(a) shows the PSNR values (convex and non-convex case) for different values of $\lambda$. To further assess the performance of tight-frame non-convex regularization, we realize several noisy ‘peppers’ images with $0.1 \leq \sigma \leq 0.3$. As in the case of the 1D signal denoising, Fig. 6 shows that non-convex regularization offers higher PSNR consistently across different noise-levels.

5 Conclusion

This paper considers the problem of signal denoising using a sparse tight-frame analysis prior. The underlying signal is assumed to be sparse with respect to an overcomplete tight frame. We propose the use of parameterized non-convex regularizers to maximally induce sparsity while maintaining the convexity of the total problem. The convexity of the objective function is ensured by restricting the parameter $a$ of the non-convex regularizer. We use ADMM to obtain the solution to the convex objective function (consisting of a non-convex regularizer), and guarantee its convergence to the global optimum, provided the augmented Lagrangian parameter $\mu > 1/r$. The proposed method outperforms the $\ell_1$ norm regularization for 1D and 2D signal denoising.

References

[1] M. Afonso, J. Bioucas-Dias, and M. Figueiredo, “Fast image recovery using variable splitting and constrained optimization.” *IEEE Trans. Image Process.*, vol. 19, no. 9, pp. 2345–2356, Sep. 2010.

[2] H. Attouch, J. Bolte, P. Redont, and A. Soubeyran, “Proximal alternating minimization and projection methods for nonconvex problems: An approach based on the Kurdyka-Lojasiewicz inequality,” *Math. Oper. Res.*, vol. 35, no. 2, pp. 438–457, May 2010.

[3] A. Blake and A. Zisserman, “Visual reconstruction,” *MIT Press*, 1987.
[4] J. Bolte, S. Sabach, and M. Teboulle, “Proximal alternating linearized minimization for non-convex and nonsmooth problems,” *Math. Program.*, vol. 146, no. 1-2, pp. 459–494, Aug. 2014.

[5] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, “Distributed optimization and statistical learning via the alternating direction method of multipliers,” *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, 2010.

[6] J.-F. Cai, B. Dong, S. Osher, and Z. Shen, “Image restoration: total variation, wavelet frames, and beyond,” *J. Am. Math. Soc.*, vol. 25, no. 4, pp. 1033–1089, May 2012.

[7] J.-F. Cai, S. Osher, and Z. Shen, “Split Bregman methods and frame based image restoration,” *Multiscale Model. Simul.*, vol. 8, no. 2, pp. 337–369, Dec. 2010.

[8] E. J. Candès, M. B. Wakin, and S. P. Boyd, “Enhancing sparsity by reweighted L1 minimization,” *J. Fourier Anal. Appl.*, vol. 14, no. 5-6, pp. 877–905, Dec. 2008.

[9] R. Chartrand, “Fast algorithms for nonconvex compressive sensing: MRI reconstruction from very few data,” *IEEE Int. Symp. Biomed. Imaging ISBI*, pp. 262–265, 2009.

[10] S. Chen, D. L. Donoho, and M. A. Saunders, “Atomic decomposition by basis pursuit,” *SIAM J. Sci. Comput.*, vol. 20, no. 1, pp. 33–61, 1998.

[11] R. R. Coifman and D. L. Donoho, “Translation-invariant de-noising,” *Lect. Notes Stat.*, vol. 103, pp. 125–150, 1995.

[12] P. L. Combettes and J.-C. Pesquet, “Proximal thresholding algorithm for minimization over orthonormal bases,” *SIAM J. Optim.*, vol. 18, no. 4, pp. 1351–1376, Nov. 2007.

[13] Y. Ding and I. W. Selesnick, “Artifact-free wavelet denoising: non-convex sparse regularization, convex optimization,” *Signal Process. Lett. IEEE*, vol. 22, no. 9, pp. 1364–1368, Sep. 2015.

[14] J. Eckstein and D. P. Bertsekas, “On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators,” *Math. Program.*, vol. 55, no. 1-3, pp. 293–318, Apr. 1992.

[15] M. Elad, P. Milanfar, and R. Rubinstein, “Analysis versus synthesis in signal priors,” *Inverse Probl.*, vol. 23, no. 3, pp. 947–968, Jun. 2007.

[16] J. Fan and R. Li, “Variable selection via nonconcave penalized likelihood and its oracle properties,” *J. Am. Stat. Assoc.*, vol. 96, no. 456, pp. 1348–1360, Dec. 2001.
[17] D. Geman and G. Reynolds, “Constrained restoration and the recovery of discontinuities,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 14, no. 3, pp. 367–383, Mar. 1992.

[18] T. Goldstein, B. O’Donoghue, S. Setzer, and R. Baraniuk, “Fast alternating direction optimization methods,” *SIAM J. Imaging Sci.*, vol. 7, no. 3, pp. 1588–1623, Jan. 2014.

[19] V. Jojic, S. Saria, and D. Koller, “Convex envelopes of complexity controlling penalties: the case against premature envelopment,” *Proc. Conf. Artif. Intell. Stat.*, vol. 15, pp. 399–406, 2011.

[20] G. Li and T. K. Pong, “Global convergence of splitting methods for nonconvex composite optimization,” *arXiv1407.0753 Preprint*, Dec. 2014.

[21] A. Majumdar and R. K. Ward, “Non-convex row-sparse multiple measurement vector analysis prior formulation for EEG signal reconstruction,” *Biomed. Signal Process. Control*, vol. 13, pp. 142–147, Sep. 2014.

[22] M. Nikolova, “Estimation of binary images by minimizing convex criteria,” *Proc. IEEE Int. Conf. Image Process.*, vol. 2, 1998.

[23] M. Nikolova, “Markovian reconstruction using a GNC approach,” *IEEE Trans. Image Process.*, vol. 8, no. 9, pp. 1204–1220, Sep. 1999.

[24] M. Nikolova, M. K. Ng, and C. P. Tam, “Fast nonconvex nonsmooth minimization methods for image restoration and reconstruction,” *IEEE Trans. Image Process.*, vol. 19, no. 12, pp. 3073–3088, Dec. 2010.

[25] P. Ochs, A. Dosovitskiy, T. Brox, and T. Pock, “An iterated L1 algorithm for non-smooth non-convex optimization in computer vision,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 1759–1766, Jun. 2013.

[26] J. Portilla, “Image restoration through L0 analysis-based sparse optimization in tight frames,” *Proc. IEEE Int. Conf. Image Process.*, pp. 3909–3912, Nov. 2009.

[27] I. W. Selesnick and I. Bayram, “Sparse signal estimation by maximally sparse convex optimization,” *IEEE Trans. Signal Process.*, vol. 62, no. 5, pp. 1078–1092, Mar. 2014.

[28] I. W. Selesnick and M. Figueiredo, “Signal restoration with overcomplete wavelet transforms: comparison of analysis and synthesis priors,” in *Proc. SPIE, Wavelets XIII*, vol. 7446, Aug. 2009.
[29] I. W. Selesnick, A. Parekh, and I. Bayram, “Convex 1-D total variation denoising with non-convex regularization,” *Signal Process. Lett. IEEE*, vol. 22, no. 2, pp. 141–144, Feb. 2015.

[30] I. W. Selesnick, R. Baraniuk, and N. Kingsbury, “The dual-tree complex wavelet transform,” *IEEE Signal Process. Mag.*, vol. 22, Nov. 2005.

[31] J. A. Tropp, “Just relax: convex programming methods for identifying sparse signals in noise,” *IEEE Trans. Inf. Theory*, vol. 52, no. 3, pp. 1030–1051, Mar. 2006.

[32] J. S. Turek, I. Yavneh, and M. Elad, “On MAP and MMSE estimators for the co-sparse analysis model,” *Digit. Signal Process.*, vol. 28, no. 1, pp. 57–74, May 2014.

[33] F. Wang, Z. Xu, and H.-K. Xu, “Convergence of Bregman alternating direction method with multipliers for nonconvex composite problems,” *arXiv1410.8625 Preprint*, Dec. 2014.

[34] S. Xie and S. Rahardja, “Alternating direction method for balanced image restoration,” *IEEE Trans. Image Process.*, vol. 21, no. 11, pp. 4557–4567, Nov. 2012.