Convex Fused Lasso Denoising with Non-Convex Regularization and its use for Pulse Detection

Ankit Parekh and Ivan W. Selesnick

Abstract—We propose a convex formulation of the fused lasso signal approximation problem consisting of non-convex penalty functions. The fused lasso signal model aims to estimate a sparse piecewise constant signal from a noisy observation. Originally, the ℓ_1 norm was used as a sparsity-inducing convex penalty function for the fused lasso signal approximation problem. However, the ℓ_1 norm underestimates signal values. Non-convex sparsity-inducing penalty functions better estimate signal values. In this paper, we show how to ensure the convexity of the fused lasso signal approximation problem with non-convex penalty functions. We further derive a computationally efficient algorithm using the majorization-minimization technique. We apply the proposed fused lasso method for the detection of pulses.

Index Terms—Sparse signal, total variation denoising, fused lasso, non-convex regularization, pulse detection.

I. INTRODUCTION

We consider the problem of estimating a sparse piecewise constant signal x from its noisy observation y, i.e.,

$$y = x + w, \qquad y, x, w \in \mathbb{R}^N, \qquad (1)$$

where w represents zero-mean additive white Gaussian noise. Estimation of sparse piecewise continuous signals arise in transient removal [28], genomic hybridization [24], [32], signal and image denoising [1], [9], prostate cancer analysis [31], sparse trend filtering [13], [23], [33], [34] and biophysics [15]. In order to estimate sparse piecewise constant signals, it has been proposed [31] to solve the following sparse-regularized optimization problem

$$\arg\min_x \left\{ \frac{1}{2} \|y - x\|_2^2 + \lambda_0 \|x\|_1 + \lambda_1 \|Dx\|_1 \right\}, \qquad (2)$$

where $\lambda_0 > 0$ and $\lambda_1 > 0$ are the regularization parameters and the matrix D is defined as

$$D = \begin{bmatrix} -1 & 1 & & \\ & \ddots & \ddots & \\ & & -1 & 1 \end{bmatrix}, \quad D \in \mathbb{R}^{(N-1)\times N}. \qquad (3)$$

The optimization problem (2) is well-known as the ℓ_1 fused lasso signal approximation (FLSA) problem [31]. The FLSA problem has been explored to aid the diagnosis of Alzheimer's disease [36] and background subtraction problems [37], however with a different data-fidelity term. Note that when $\lambda_0 = 0$,

problem (2) reduces to the total variation denoising (TVD) problem [26].

It is known that the ℓ_1 norm, when used as a sparsity-inducing regularizer, underestimates the signal values. The ℓ_1 norm is generally not the tightest convex envelope for sparsity [12]. In order to better estimate signal values, non-convex penalty functions are often favored over the ℓ_1 norm [2], [4], [19], [22], [25], [35]. However, the use of non-convex penalty functions generally leads to non-convex optimization, which suffer from several issues (spurious local minima, initialization, convergence, etc.).

In this paper, we propose to estimate sparse piecewise constant signals via the following convex non-convex (CNC) FLSA problem

$$\arg\min_x \Bigg\{ F(x) = \frac{1}{2}\|y - x\|_2^2 + \lambda_0 \sum_{n=1}^{N} \phi(x_n; a_0)$$
$$+ \lambda_1 \sum_{n=1}^{N-1} \phi([Dx]_n; a_1) \Bigg\}, \qquad (4)$$

where $\phi\colon \mathbb{R} \to \mathbb{R}$ with $a \geqslant 0$ is a non-convex sparsity-inducing regularizer. Specifically, we propose that the regularization terms be chosen so that the objective function F in (4) is convex. As a result, the CNC FLSA approach avoids the drawbacks of non-convex optimization. The non-convex penalty function ϕ aims to induce sparsity more strongly than the ℓ_1 norm and thus better estimate the signal values. The parameter a controls the degree of non-convexity of ϕ; a higher value of a indicates a higher degree of non-convexity for ϕ. As a main result, we state and prove a condition that a_0 and a_1 must satisfy to ensure the objective function F in (4) is strictly convex. As a consequence of the convexity condition, well-known convex optimization techniques can be used to reliably obtain the global minimum of the objective function F. As a second main result, we provide an efficient fast converging algorithm for the proposed CNC FLSA problem (4). The algorithm is derived using the majorization-minimization (MM) procedure [8].

The idea of formulating a convex problem with non-convex regularization was described by Blake and Zisserman [3], and Nikolova [17], [18]. The idea is to balance the positive second-derivatives of the data-fidelity term with the negative second-derivatives of the non-convex penalty function. This approach has been successfully applied to various signal processing applications (e.g., [6], [14], [21], [29] and the references therein). Using this technique, a modified formulation of the ℓ_1 FLSA problem (2) was proposed with an aim to induce sparsity more strongly than the ℓ_1 norm [1]. Further, a two-step

procedure was used to obtain the solution to the modified fused lasso problem. However, the modified fused lasso (MDFL) problem [1] considers only the first of the two regularization terms as non-convex.

This paper is organized as follows. In Section II we describe the class of non-convex penalty functions. In Section III we provide the convexity condition for the objective function F in (4). We derive an algorithm to solve the CNC FLSA problem (4) based on the MM procedure in Section IV. In Section V we apply the proposed CNC FLSA approach to the problem of detecting ECG pulses in strong additive white Gaussian noise (AWGN).

II. PRELIMINARIES

A. Notation

We denote vectors and matrices by lower and upper case letters respectively. The N-point signal y is represented by the vector

$$y = [y_0, \ldots, y_{N-1}]^T, \qquad y \in \mathbb{R}^N, \qquad (5)$$

where $[\cdot]^T$ represents the transpose. The ℓ_1 and ℓ_2 norms of the vector y are defined as

$$\|y\|_1 = \sum_n |y(n)|, \quad \|y\|_2 = \left(\sum_n |y(n)|^2 \right)^{1/2}. \qquad (6)$$

The soft-threshold function [7] for $\lambda > 0, \lambda \in \mathbb{R}$ is defined as

$$\mathrm{soft}(x; \lambda) = \begin{cases} x + \lambda, & x < -\lambda \\ 0, & -\lambda \leqslant x \leqslant \lambda \\ x - \lambda, & x > \lambda. \end{cases} \qquad (7)$$

For $x \in \mathbb{R}^N$, the notation $\mathrm{soft}(x; \lambda)$ implies that the soft-threshold function is applied element-wise to x with a threshold of λ.

Definition 1: The total variation denoising (TVD) problem [26] is defined as

$$\mathrm{tvd}(y; \lambda) = \arg \min_x \left\{ \frac{1}{2} \|y - x\|_2^2 + \lambda \|Dx\|_1 \right\}, \qquad (8)$$

where $\lambda > 0$ is the regularization parameter.

We note the following lemma, which provides an efficient two-step solution to the ℓ_1 FLSA problem (2).

Lemma 1: [9, Lemma A.1] The solution x^* to the ℓ_1 FLSA problem (2) is given by

$$x^* = \mathrm{soft}\big(\mathrm{tvd}(y, \lambda_1), \lambda_0\big). \qquad (9)$$

B. Non-convex penalty functions

We propose to use parameterized non-convex penalty functions, with a view to induce sparsity more strongly than the ℓ_1 norm. We assume such non-convex penalty functions have the following properties.

Assumption 1: The penalty function $\phi \colon \mathbb{R} \to \mathbb{R}$ satisfies the following

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

3) $\phi''(x) \leqslant 0, \forall x > 0$
4) $\phi'(0^+) = 1$
5) $\inf\limits_{x \neq 0} \phi''(x; a) = \phi''(0^+; a) = -a$

An example of a penalty function, which satisfies Assumption 1, is the logarithmic penalty function [4] defined as

$$\phi(x; a) = \begin{cases} \dfrac{1}{a} \log(1 + a|x|), & a > 0 \\ |x|, & a = 0. \end{cases} \qquad (10)$$

Note that when $a = 0$, this penalty function reduces to the ℓ_1 norm. Other examples of non-convex penalty functions satisfying Assumption 1 include the arctangent and the rational penalty functions [11], [27]. Note that the ℓ_p norm does not satisfy Assumption 1.

We note the following lemma, which we will use to obtain a convexity condition for optimization problem (4).

Lemma 2: [21] Let $\phi \colon \mathbb{R} \to \mathbb{R}$ satisfy Assumption 1. The function $s \colon \mathbb{R} \to \mathbb{R}$ defined as

$$s(x; a) = \phi(x; a) - |x|, \qquad (11)$$

is twice continuously differentiable and concave with

$$-a \leqslant s''(x; a) \leqslant 0. \qquad (12)$$

III. CONVEXITY CONDITION

In this section, we seek to find a condition on the parameters a_0 and a_1 to ensure that the objective function F in (4) is strictly convex. The following theorem provides the required condition on a_0 and a_1.

Theorem 1: Let $\phi \colon \mathbb{R} \to \mathbb{R}$ be a non-convex penalty function satisfying Assumption 1. The function $F \colon \mathbb{R}^N \to \mathbb{R}$ defined as

$$F(x) = \frac{1}{2} \|y - x\|_2^2 + \lambda_0 \sum_{n=1}^{N} \phi(x_n; a_0) + \lambda_1 \sum_{n=1}^{N-1} \phi([Dx]_n; a_1), \qquad (13)$$

is strictly convex if

$$0 \leqslant a_0 \lambda_0 + 4 a_1 \lambda_1 \leqslant 1. \qquad (14)$$

Proof: Consider the function $G \colon \mathbb{R}^N \to \mathbb{R}$ defined as

$$G(x) = \frac{1}{2} \|y - x\|_2^2 + \lambda_0 \sum_{n=1}^{N} s(x_n; a_0) + \lambda_1 \sum_{n=1}^{N-1} s([Dx]_n; a_1), \qquad (15)$$

where $s(x; a) = \phi(x; a) - |x|$. From Lemma 2, the function G is twice continuously differentiable and its Hessian can be written as

$$\nabla^2 G = I + \lambda_0 \Gamma(x; a_0) + \lambda_1 D^T \Gamma(Dx; a_1) D, \qquad (16)$$

where

$$\Gamma(x; a) = \begin{bmatrix} s_1''(x_1; a) & & \\ & \ddots & \\ & & s_n''(x_n; a) \end{bmatrix}. \qquad (17)$$

For the strict convexity of G, we need to ensure that $\nabla^2 G$ is positive definite. To this end, from the assumptions on ϕ, it
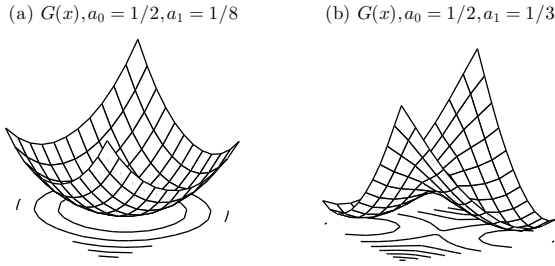
Fig. 1. Surface plots illustrating the convexity condition. (a) The function $G(x)$ (15) is convex for $a_0 = 1/2$ and $a_1 = 1/8$. (b) The function $G(x)$ is not convex for $a_0 = 1/2$ and $a_1 = 1/3$ (these values violate Theorem 1).
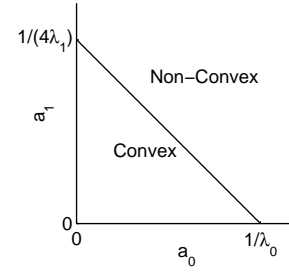

Fig. 2. Region of convexity for the function F in (13). The function F is strictly convex for any values of a_0 and a_1 inside the triangular region.

follows that

$$\Gamma(x;a_0) \succeq -a_0 I, \quad x \in \mathbb{R}^N. \tag{18}$$

Moreover, we can write

$$D^T \Gamma(Dx;a_1)D \succeq -a_1 D^T D \tag{19}$$
$$\succ -4a_1 I. \tag{20}$$

The inequality (20) is obtained using the eigenvalues[1] of the matrix $D^T D$. Using (16), (18) and (20), $\nabla^2 G \succ 0$ if

$$(1 - a_0\lambda_0 - 4a_1\lambda_1)I \succeq 0, \tag{21}$$

or equivalently if,

$$1 - a_0\lambda_0 - 4a_1\lambda_1 \geqslant 0. \tag{22}$$

From (11), (13) and (15) it is straighforward that

$$F(x) = G(x) + \lambda_0\|x\|_1 + \lambda_1\|Dx\|_1. \tag{23}$$

Hence, F in (13) is strictly convex as long as the inequality (22) holds true (the function F is a sum of a strictly convex function G, the convex ℓ_1 norm, and the convex TV penalty). $\square$

The following example illustrates the convexity condition (22) for $N = 2$. Let $\lambda_0 = \lambda_1 = 1$ and $y = 0$. As per Theorem 1, the function G (by extension the function F) is strictly convex if $a_0 + 4a_1 \leqslant 1$. Figure 1(a) shows the function G with the values $a_0 = 1/2$ and $a_1 = 1/8$. These values satisfy Theorem 1 and as a result the function G is strictly convex. On the other hand, Fig. 1(b) shows the function G when $a_0 = 1/2$ and $a_1 = 1/3$. These values of a_0 and a_1 violate the Theorem 1; consequently the function G is non-convex as seen in Fig. 1(b).

The convexity condition given by Theorem 1 in (14) implies that the values of a_0 and a_1 must lie on or below the line given by $a_0\lambda_1 + 4a_1\lambda_1 = 1$. Figure 2 displays the values of a_0 and a_1 for which the function F is strictly convex. In order to maximally induce sparsity, we choose the values of a_0 and a_1 on the line. Specifically, we propose to select a value of $a_0 \in (0, 1/\lambda)$ and set the value of a_1 as

$$a_1 = \frac{1 - a_0\lambda_0}{4\lambda_1}. \tag{24}$$

[1]The eigenvalues of $D^T D$ are given by $\{2 - 2\cos(k\pi/N)\}$ for $k = 0, \ldots, N-1$ [30].

IV. Optimization Algorithm

Due to Theorem 1, we can reliably obtain via convex optimization the global minimum of (4) as long as the parameters a_0 and a_1 are chosen to satisfy (14). We derive an algorithm for the proposed CNC fused lasso method using the majorization-minimization (MM) procedure [8], such that

$$x^{k+1} = \arg\min_x F^{\mathsf{M}}(x, x^k), \tag{25}$$

where F^{M} denotes a majorizer of the function F in (4), and where k is the iteration index. The MM procedure guarantees that each iteration monotonically decreases the value of the objective function F in (4). We use the absolute value function and a linear function to majorize the non-convex penalty function. With this particular choice of majorizer, each MM update iteration involves solving the ℓ_1 FLSA problem (2).

To derive a majorizer of the function ϕ, note that $\phi(x;a) = s(x;a) + |x|$. As a result, it suffices to majorize the function s with a linear term in order to obtain a majorizer of the function ϕ. Observe that since s is a concave function, the tangent line to s at a point v always lies above the function s. Using the tangent line to the function s, a majorizer of the function ϕ is given by $\phi^{\mathsf{M}} \colon \mathbb{R} \times \mathbb{R} \to \mathbb{R}$, defined as

$$\phi^{\mathsf{M}}(x, v; a) = |x| + s'(v;a)(x - v) + s(v;a), \tag{26}$$

for $x, v \in \mathbb{R}$. It follows straightforwardly that

$$\phi^{\mathsf{M}}(x, v; a) \geqslant \phi(x;a), \quad \forall x, v \in \mathbb{R}, \tag{27}$$
$$\phi^{\mathsf{M}}(v, v; a) = \phi(v;a), \quad \forall v \in \mathbb{R}. \tag{28}$$

Figure 3(a) shows the absolute value function $|x|$. The twice continuously differentiable function $s(x;a)$ is shown in Fig. 3(b), along with the tangent line to $s(x;a)$ at $x = 1$. Figure 3(c) shows the non-convex penalty function ϕ and its majorizer $\phi^{\mathsf{M}}(x, v; s)$ given by (26). The majorizer is the sum of the absolute value function in Fig. 3(a) and the tangent line to $s(x;a)$ in Fig. 3(b).

Using (27) and (28), we note that

$$\sum_n \phi^{\mathsf{M}}(x_n, v_n; a) \geqslant \sum_n \phi(x_n; a), \tag{29}$$
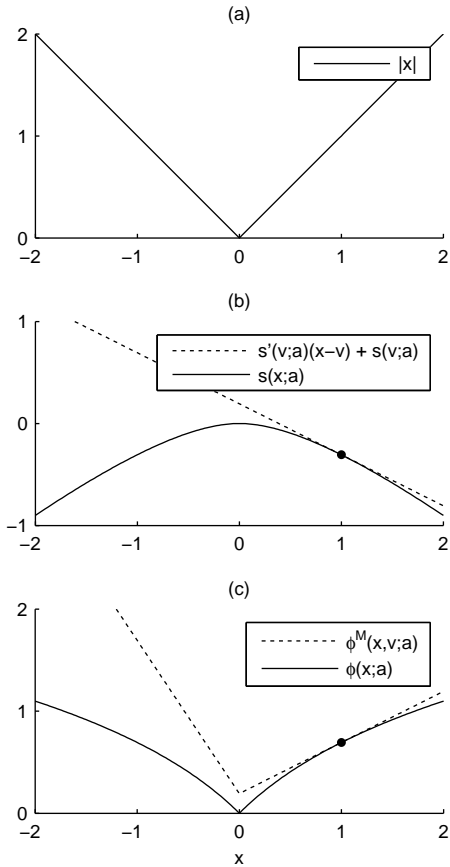$$\sum_n \phi^{\mathsf{M}}([Dx]_n, [Dv]_n; a) \geqslant \sum_n \phi([Dx]_n; a), \tag{30}$$

Fig. 3. (a) The absolute value function $|x|$. (b) The twice continuously differentiable function $s(x; a)$ and the tangent line at $x = 1$. (c) The nonconvex penalty function ϕ and its majorizer $\phi^{\mathsf{M}}(x, v; a)$ given in (26).

Fig. 4. Denoising a synthetic signal using CNC fused lasso (4) and ℓ_1 fused lasso (2).

with equality if $x = v$. Further, note that

$$\sum_n \phi^{\mathsf{M}}(x_n, v_n; a) = \|x\|_1 + s'(v; a)^T(x - v) + C_1, \quad (31)$$

where $s'(v; a)$ is the vector defined as $[s'(v; a)]_n = s'(v_n; a)$, i.e., the derivative of the function s is applied element-wise to the vector v. Further, note that C_1 is a constant that does not depend on x. Similarly, we write

$$\sum_n \phi^{\mathsf{M}}([Dx]_n, [Dv]_n; a)$$
$$= \|Dx\|_1 + s'(Dv; a)^T D(x - v) + C_2, \quad (32)$$

where C_2 is a constant that does not depend on x. Therefore, using (31) and (32), a majorizer of the objective function F in (4) is given by $F^{\mathsf{M}} : \mathbb{R}^N \times \mathbb{R}^N \to \mathbb{R}$, defined as

$$F^{\mathsf{M}}(x, v) = \frac{1}{2}\|y - x\|_2^2 + \lambda_0\|x\|_1 + \lambda_0 s'(v; a)^T(x - v)$$
$$+ \lambda_1\|Dx\|_1 + \lambda_1 s'(Dv; a)^T D(x - v) + C, \quad (33)$$

where C is a constant that does not depend on x. Completing the square, we write (33) as

$$F^{\mathsf{M}}(x, v) = \frac{1}{2}\|\tilde{y}(v) - x\|_2^2 + \lambda_0\|x\|_1 + \lambda_1\|Dx\|_1 + C, \quad (34)$$
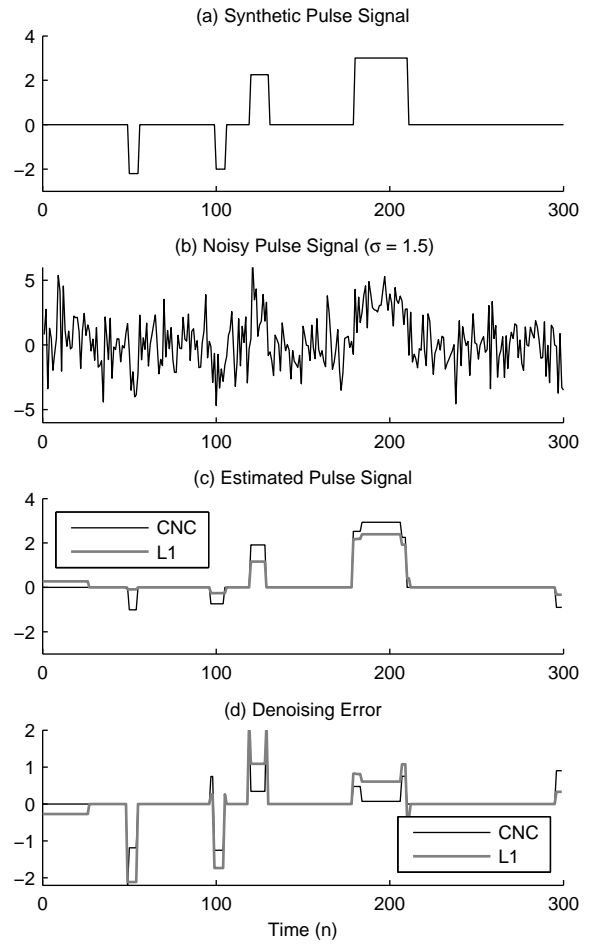
where

$$\tilde{y}(v) = y - \lambda_0 s'(v; a) - \lambda_1 D^T s'(Dv; a). \quad (35)$$

Therefore, each MM iteration consists of minimizing the function F^{M} (34), which is the ℓ_1 FLSA problem (2) with $\tilde{y}(v)$ as the input. Consequently, using (9) the MM update (25) can be written as

$$\tilde{y}^k = y - \lambda_0 s'(x^k; a) - \lambda_1 D^T s'(Dx^k; a) \quad (36a)$$
$$x^{k+1} = \text{soft}(\text{tvd}(\tilde{y}^k, \lambda_1), \lambda_0). \quad (36b)$$

Equation (36) constitutes a fast converging, computationally efficient algorithm to solve the proposed CNC FLSA problem (4). To implement (36b), we use the fast (finite-time) exact TV denoising algorithm based on the 'taut-string' method [5], which has a worst case complexity of $\mathcal{O}(n)$. We initialize the iteration with the solution (9) to the ℓ_1 FLSA problem (2). Note that the MM update (36) does not consist of any matrix inverses.

V. Examples

We consider the problem of estimating pulses of varying width in the presence of high additive white Gaussian noise (AWGN). We model the pulse signal as sparse piecewise
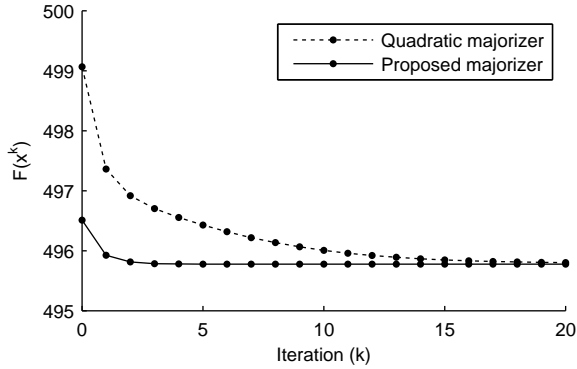
Fig. 5.   The value of the objective function F in (4) at each iteration of the MM algorithm using quadratic and proposed majorizers. The proposed MM algorithm (36) converges within 5 iterations.
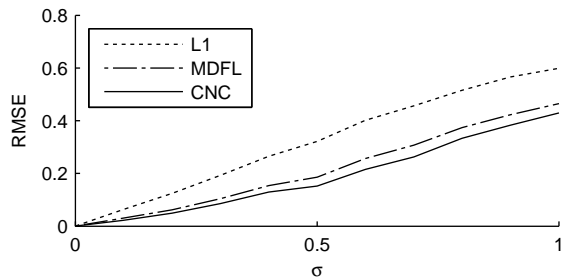

Fig. 6.   Average RMSE as a function of σ. The proposed CNC FLSA yields the lowest RMSE across different values of the noise variance (σ^2).
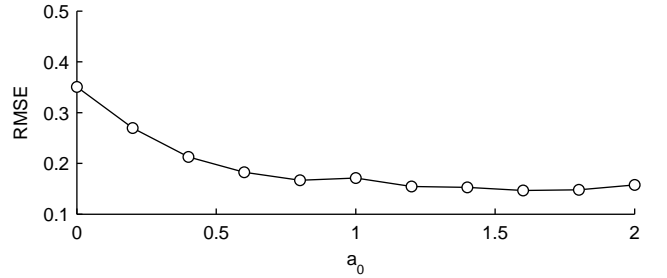

Fig. 7.   Average RMSE as a function of a_0 for the synthetic test signal example in Fig. 4.
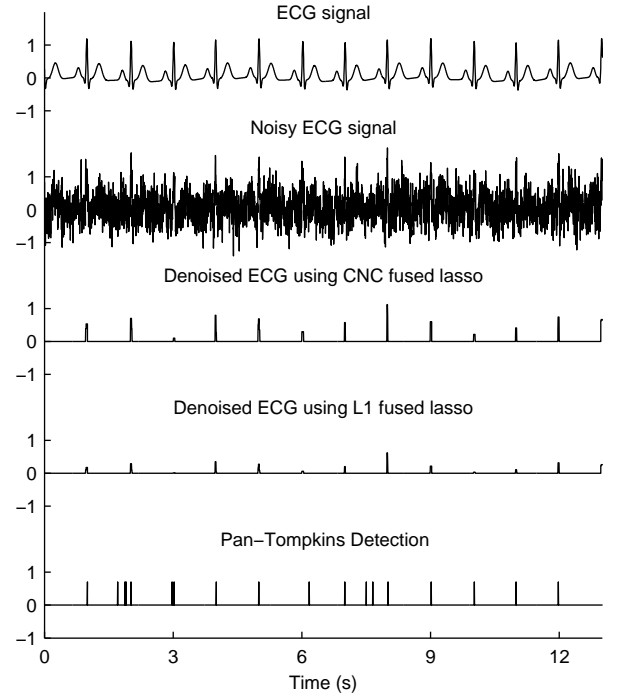

Fig. 8.   Denoising of ECG signal in strong AWGN. The ℓ_1 FLSA underestimates the signal values. The Pan-Tompkins detects several false-positive R-waves.

constant and apply the proposed CNC FLSA problem (4) to estimate the individual pulses. We set the value of λ_1 as in [29], i.e., $\lambda_1 = \beta\sqrt{N}\sigma$ where β is a constant (usually $1/4$) and σ represents the standard deviation of the additive white Gaussian noise. We manually set the value of λ_0 to obtain the lowest RMSE.

Figure 4(a) illustrates the synthetic clean pulse signal and Fig. 4(b) shows the noisy pulse signal. Shown in Fig. 4(c) are the estimates obtained using the standard ℓ_1 norm and the non-convex atan penalty function [19, equation (23)]. It can be seen that the proposed CNC FLSA method estimates the pulses more accurately than the ℓ_1 FLSA method. The relative performance of the CNC fused lasso in estimating pulses is also highlighted by the denoising error shown in Fig. 4(d).

The value of the objective function F in (4), after each iteration of the MM algorithm (36), is shown in Fig. 5. The MM algorithm derived in [28, Table II] for a more general problem can also be used to solve the CNC FLSA problem (4). However, the MM algorithm in [28] utilizes a quadratic majorizer for the non-convex penalty function ϕ. Figure 5 shows that the proposed MM algorithm (36) converges much faster than the MM algorithm in [28]. The proposed MM algorithm (36) converges in about 5 iterations.

In order to assess the relative performance of the proposed CNC fused-lasso method, we use 15 realizations of the noisy synthetic pulse signal in Fig. 4 and denoise them using both the original ℓ_1 fused lasso and the proposed CNC fused lasso methods. We also compare with the modified fused lasso

(MDFL) [1], which is a special case of (4) with $a_1 = 0$; i.e., only the first regularization term is non-convex. It can be seen in Fig. 6 that the proposed CNC FLSA (4) approach offers the lowest RMSE values across different noise levels. Further, for the test signal in Fig. 4(a), the average RMSE as a function of a_0 is shown in Fig. 7. Note that a_1 is set according to (24).

As an another example, we consider the problem of detecting the QRS peaks in an ECG signal in AWGN with high variance (σ^2). Wearable heart-rate monitors suffer from strong noise due to abrupt motion artifacts. Several methods for detecting the QRS peaks in ECG signals were studied in [10], [20]. We evaluate the detection of ECG R-waves in strong AWGN using the proposed CNC fused lasso method. Using a sampling frequency of 256 Hz, we simulate the ECG signal using the synthetic ECG waveform generator, ECGSYN [16]. The clean and noisy ECG ($\sigma = 0.4$) are shown in Fig. 8. We set the parameters $\lambda_0 = 0.6$, $\lambda_1 = 0.9$, $a_0 = 0.9/\lambda_0$ and $a_1 = 0.1/(4\lambda_1)$. We use 20 iterations for the proposed CNC

FLSA algorithm (36).

Figure 8 illustrates the denoised ECG signal using the ℓ_1 FLSA and the proposed CNC FLSA methods. It can be seen that the ℓ_1 FLSA does not detect all the R-waves. Moreover, the amplitudes of the pulses detected using the proposed CNC FLSA (4) are relatively high compared to those detected using the ℓ_1 FLSA. The ℓ_1 norm tends to underestimate signal values. Also shown in Fig. 8 are the R-waves detected using the Pan-Tompkins real-time QRS detector [20]. Note that the Pan-Tompkins detector was not designed for ECG signals with high-noise variance. As a result, the Pan-Tompkins algorithm detects several false-positive R-waves.

VI. CONCLUSION

The fused lasso signal approximation (FLSA) problem aims to estimate sparse piecewise constant signals. In order to improve the accuracy of the ℓ_1 FLSA approach, we use non-convex penalty functions as sparsity-inducing regularizers. In this paper we generalize the results of [1], which addresses the case wherein only one of the two regularization terms is non-convex. We prove that the proposed FLSA objective function is convex when the non-convex penalty parameters are suitably set. We also derive a computationally efficient algorithm using the majorization-minimization technique. The proposed CNC FLSA algorithm does not consist of any calculations involving a matrix inverse. We apply the proposed method to the problem of pulse detection under high additive white Gaussian noise. An illustration is provided for the detection of R-waves in an ECG signal.

VII. ACKNOWEDGEMENT

The authors would like to thank Ilker Bayram for valuable comments on an earlier version of this paper.

REFERENCES

[1] I. Bayram, P.-Y. Chen, and I. W. Selesnick, "Fused lasso with a non-convex sparsity inducing penalty," *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, no. 1, pp. 4184–4188, 2014.

[2] A. Bruckstein, D. Donoho, and M. Elad. "From sparse solutions of systems of equations to sparse modeling of signals and images." *SIAM Rev.*, vol. 51, no. 1, pp. 34–81, 2009.

[3] A. Blake and A. Zisserman, "Visual reconstruciton," *MIT Press*, 1987.

[4] E. J. Candès, M. B. Wakin, and S. P. Boyd, "Enhancing sparsity by reweighted l1 minimization," *J. Fourier Anal. Appl.*, vol. 14, no. 5, pp. 877–905, Dec. 2008.

[5] L. Condat, "A direct algorithm for 1D total variation denoising," *IEEE Signal Process. Lett.*, vol. 20, no. 11, pp. 1054–1057, Nov. 2013.

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

[7] D. L. Donoho, "De-noising by soft-thresholding," *IEEE Trans. Inf. Theory*, vol. 41, no. 3, pp. 613–627, 1995.

[8] M. A. T. Figueiredo, J. M. Bioucas-Dias, and R. D. Nowak, "Majorization-minimization algorithms for wavelet-based image restoration," *IEEE Trans. Image Process.*, vol. 16, no. 12, pp. 2980–2991, 2007.

[9] J. Friedman, T. Hastie, H. Höfling, and R. Tibshirani, "Pathwise coordinate optimization," *Ann. Appl. Stat.*, vol. 1, no. 2, pp. 302–332, Dec. 2007.

[10] G. M. Friesen, T. C. Jannett, M. A. Jadallah, S. L. Yates, S. R. Quint *et al.*, "A comparison of the noise sensitivity of nine QRS detection algorithms." *IEEE Trans. Biomed. Eng.*, vol. 37, no. 1, pp. 85–98, 1990.

[11] 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.

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

[13] S. J. Kim, K. Koh, S. P. Boyd, and D. Gorinevsky, "L1 trend filtering," *Siam Rev.*, vol. 51, no. 2, pp. 339–360, 2009.

[14] A. Lanza, S. Morigi, and F. Sgallari, "Convex image denoising via non-convex regularization," in *Scale Sp. Var. Methods Comput. Vis.*, ser. Lecture Notes in Computer Science, J.-F. Aujol, M. Nikolova, and N. Papadakis, Eds. Springer, vol. 9087, pp. 666–677, 2015.

[15] M. A. Little and N. S. Jones, "Generalized methods and solvers for noise removal from piecewise constant signals. I. Background theory," *Proc. R. Soc. London A Math. Phys. Eng. Sci.*, Jun. 2011.

[16] P. E. McSharry, G. D. Clifford, L. Tarassenko, and L. A. Smith, "A dynamical model for generating synthetic electrocardiogram signals," *IEEE Trans. Biomed. Eng.*, vol. 50, no. 3, pp. 289–294, 2003.

[17] M. Nikolova, "Estimation of binary images by minimizing convex criteria," *Proc. IEEE Int. Conf. Image Process. (ICIP)*, vol. 2, 1998.

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

[19] M. Nikolova, "Energy minimization methods," in *Handbook of mathematical methods in imaging*, O. Scherzer, Ed. Springer, pp. 138–186, 2011

[20] J. Pan, W. Tompkins, "A real-time QRS detection algorithm," in *IEEE Trans. Biomed.*, vol. 32, no. 3, pp. 230–236, Mar. 1985.

[21] A. Parekh and I. W. Selesnick, "Convex denoising using non-convex tight frame regularization," *IEEE Signal Process. Lett.*, vol. 22, no. 10, pp. 1786–1790, 2015.

[22] J. Portilla and L. Mancera, "L0-based sparse approximation: Two alternative methods and some applications." *Proc. SPIE*, Aug. 2007.

[23] A. Ramdas and R. Tibshirani, "Fast and flexible ADMM algorithms for trend filtering," *Preprint arXiv1406.2082*, vol. 12, no. 1, pp. 1–19, 2014.

[24] F. Rapaport, E. Barillot, and J. P. Vert, "Classification of array CGH data using fused SVM," *Bioinformatics*, vol. 24, no. 13, 2008.

[25] A. Repetti, E. Chouzenoux and J. C. Pesquet, "A nonconvex regularized approach for phase retrieval," *Proc. IEEE Int. Conf. Image Process. (ICIP)*, pp. 1753–1757, Oct. 2014.

[26] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Phys. D Nonlinear Phenom.*, vol. 60, no. 1-4, pp. 259–268, Nov. 1992.

[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, H. L. Graber, Y. Ding, T. Zhang, and R. L. Barbour, "Transient artifact reduction algorithm (TARA) based on sparse optimization," *IEEE Trans. Signal Process.*, vol. 62, no. 24, pp. 6596–6611, 2014.

[29] I. W. Selesnick, A. Parekh, and I. Bayram, "Convex 1-D total variation denoising with non-convex regularization," *IEEE Signal Process. Lett.*, vol. 22, no. 2, pp. 141–144, Feb. 2015.

[30] G. Strang, "The discrete cosine transform," *SIAM Rev.*, vol. 41, no. 1, pp. 135–147, 1999.

[31] R. Tibshirani, M. Saunders, S. Rosset, J. Zhu, and K. Knight, "Sparsity and smoothness via the fused lasso," *J. R. Stat. Soc. Ser. B Stat. Methodol.*, vol. 67, no. 1, pp. 91–108, 2005.

[32] R. Tibshirani and P. Wang, "Spatial smoothing and hot spot detection for CGH data using the fused lasso," *Biostatistics*, vol. 9, no. 1, pp. 18–29, 2008.

[33] R. J. Tibshirani, "Adaptive piecewise polynomial estimation via trend filtering," *Ann. Stat.*, vol. 42, no. 1, pp. 285–323, 2014.

[34] Y.-X. Wang, J. Sharpnack, A. Smola, and R. J. Tibshirani, "Trend filtering on graphs," *Preprint arXiv1410.7690*, pp. 1–32, 2014.

[35] J. Woodworth and R. Chartrand, "Compressed sensing recovery via nonconvex shrinkage penalties," *Preprint arXiv:1504.02923*, pp. 1-29, Apr. 2015.

[36] B. Xin, Y. Kawahara, Y. Wang, and W. Gao, "Efficient generalized fused lasso and its application to the diagnosis of Alzheimer's Disease," *Proc. Conf. Artif. Intell. (AAAI)*, pp. 2163–2169, 2014.

[37] B. Xin, Y. Tian, Y. Wang, and W. Gao, "Background subtraction via generalized fused lasso foreground modeling," *Preprint arXiv1504.03707*, pp. 1–9, 2014.