

# Compressed Signal Recovery by Convex Optimization and Non-Convex Regularization

<sup>[1]</sup>A.Suban, <sup>[1]</sup>V.Karthick, <sup>[2]</sup>Mallikarjuna Nandi, <sup>[3]</sup>M. Pradeep

<sup>[1][2]</sup>Assistant Professor, <sup>[3]</sup>Lecturer

<sup>[1], [3]</sup>Department of Electronics and Communication Engineering

<sup>[2]</sup>Department of Computer Science and Engineering

<sup>[1],[2],[3]</sup>Velammal College of Engineering and Technology, Madurai

**Abstract** - We Aim to promote a standard approach for estimating sparse signals in noise is Convex optimization with sparsity-promoting convex Regularization. In order to promote sparsity more strongly than convex regularization, it is also standard practice to employ non-convex optimization. In this paper, we take a third approach. We utilize a non-convex regularization term chosen such that the total cost function (consisting of data consistency and regularization terms) is convex. Therefore, sparsity is more strongly promoted than in the standard convex formulation, but without sacrificing the attractive aspects of convex optimization. We use this idea to improve the recently developed ‘overlapping group shrinkage’ (OGS) algorithm for the de-noising of group-sparse signals. The algorithm is applied to the problem of speech enhancement with favorable results in terms of both SNR and perceptual quality.

**Index Terms**—Convex optimization, Sparse optimization, de-noising, Threshold function, non-convex optimization, speech enhancement.

## I. INTRODUCTION

This paper aims to develop an approach that promotes sparsity more strongly than convex formulation. An example of such a vector (in 2D) is the spectrogram of a speech waveform. The spectrogram of a speech waveform exhibits areas and ridges of large magnitude, but not isolated large values. The method proposed in this work will be demonstrated on the problem of speech filtering.

Convex and non-convex optimization are both common practice for the estimation of sparse vectors from noisy data. Generally, convex approaches are based on sparsity-promoting convex penalty functions (e.g., the  $l_1$  norm), while non-convex approaches are based on non-convex penalty functions

The algorithm we present is derived according to the principle of majorization-minimization (MM) . The proposed approach:

- 1) does not underestimate large amplitude components of sparse solutions to the extent that convex penalties do,
- 2) is translation invariant (due to groups in the proposed method being fully overlapping),
- 3) is computationally efficient (  $O(N)$  per iteration) with monotonically decreasing cost function.
- 4) requires no algorithmic parameters (step-size, Lagrange, etc.).

## A.Related Work

This paper develops a specific threshold function designed so as to have the three properties advocated in unbiasedness (of large coefficients), sparsity and continuity. Further, the threshold function  $\Theta$  and its corresponding penalty function  $\phi$  are parameterized by two parameters: the threshold  $T$  and the right-sided derivative of  $\Theta$  at the threshold, a measure of the threshold function’s sensitivity.

## II. PRELIMINARIES

### A. Notation

We will work with finite-length discrete signals which we denote in lower case bold. The  $n$ -point signal is written as

$$\mathbf{X} = [x(0), \dots, x(N-1)] \in \mathbb{R}^N$$

We use the notation

$$\mathbf{X}_{i,K} = [x(i), \dots, x(i+K-1)] \in \mathbb{R}^K$$

to denote the  $i$ -th group of size  $K$ . We consistently use  $K$  (a positive integer) to denote the group size. At the boundaries (i.e., for  $i < 0$  and  $i > N-K$ ).

### B. Penalty Functions

We will make the following assumptions on the penalty function,  $\phi : \mathbb{R} \rightarrow \mathbb{R}$

- 1)  $\phi$  is continuous on  $\mathbb{R}$
- 2)  $\phi$  is twice differentiable on  $\mathbb{R} \setminus \{0\}$
- 3)  $\phi(-x) = \phi(x)$  (symmetric)
- 4)  $\phi''(x) > 0$  (increasing on  $\mathbb{R}^*_+$ )
- 5)  $\phi'(0^+) = 1$  (unit slope at zero)
- 6)  $\phi''(x) \leq 0$  (concave on  $\mathbb{R}^*_+$ )
- 7)  $\phi''(0^+) \leq \phi''(x)$
- 8)  $\phi''(0^+)$  is finite.

### C. Threshold Functions

The fact that the soft threshold function reduces large values by a constant amount is considered its deficiency. In the estimation of sparse signals in AWGN, this behavior results in a systematic underestimation (bias) of large magnitude signal values. Hence, threshold functions that are asymptotically unbiased are often preferred to the soft threshold function, and the penalty functions from which they are derived promote sparsity more strongly than the  $l_1$  norm.

Threshold functions corresponding to several penalty functions. The threshold function corresponding to the absolute value penalty function is called the *soft* threshold function. Notice that, except for the soft threshold function, the threshold functions approach the identity function

asymptotically. The atan threshold function approaches identity the fastest.

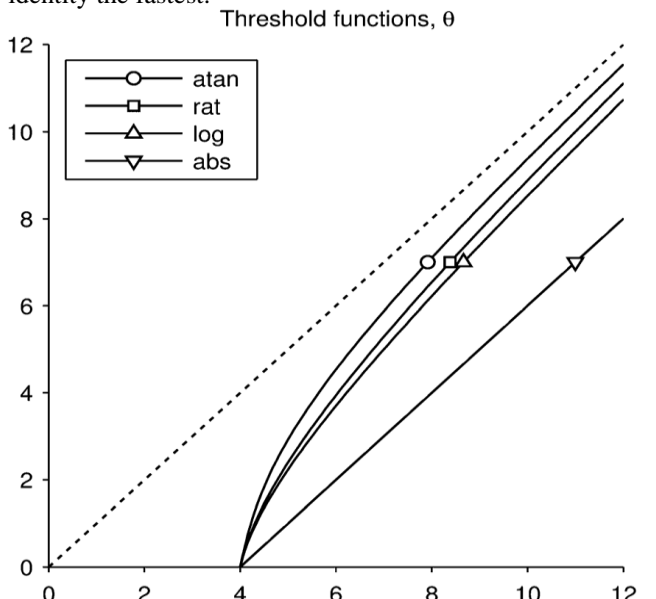


Fig.. Threshold functions derived from the four penalty functions given in Section II-B; three of which are non-convex.

### III. OGS WITH NON-CONVEX REGULARIZATION

For de-noising group-sparse signals in AWGN, we propose to minimize the cost function,

$$F(\mathbf{x}) = \frac{1}{2} \|\mathbf{y} - \mathbf{x}\|_2^2 + \lambda \sum_{i \in \mathcal{Z}} \phi(\|\mathbf{x}_{i,K}\|_2; a)$$

where  $\phi$  is a (non-convex) sparsity promoting penalty function satisfying the assumptions in Section II-B, and  $\lambda > 0$ . The notation  $\mathbf{x}_{i,K}$ , denotes a  $K$ -point group starting at index  $i$ . The group size, (a positive integer), should be selected based roughly on the size of the groups (clusters) arising in the data. This constitutes one's 'prior knowledge' regarding the data to be denoised and may need to be set through some trial-and-error. We note that does not impose any strict constraint on the size of groups, not does it define the boundaries of groups. The current work addresses the case where  $\phi$  is a non-convex regularizer, so as to promote group sparsity more strongly in comparison to convex regularization.

#### A. Overlapping Group Thresholding[OGS]

Using the results above, we can find a condition on  $a$  to ensure  $F$  is strictly convex. The result permits the use of non-convex regularization to strongly promote group sparsity while preserving strict convexity of the total cost function. A small  $a$ , in turn, limits the non-convexity of the regularizer. Hence, it appears the benefit of the proposed non-convex regularization method is diminished for large  $K$ . However, two considerations offset this reasoning. First, for larger  $K$ , a smaller value of  $a$  is needed so as to achieve a fixed level of noise suppression. Secondly, for larger  $K$ , there is greater overlap between adjacent groups because the groups are fully-overlapping; so, regularization may be more sensitive to  $a$ .

#### B. Minimization Algorithm

To derive an algorithm minimizing the strictly convex function, we use the majorization-minimization (MM) procedure. The MM procedure replaces a single minimization problem by a sequence of (simpler) ones. Specifically, MM is based on the iteration

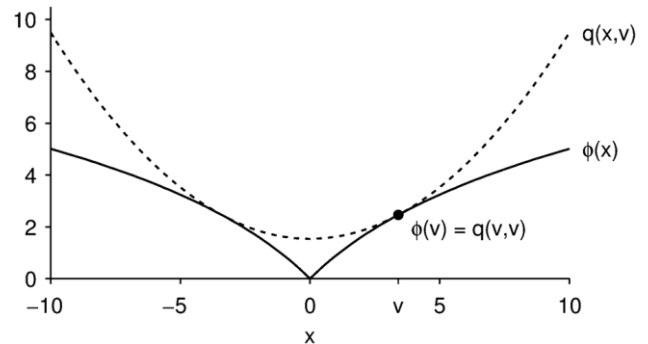


Fig.. Majorization of non-convex  $\phi(x)$  by  $q(x,v)$ .

To specify a majorizer of the cost function  $F$ , we first specify a majorizer of the penalty function  $\phi$ . To simplify notation, we suppress the dependence of  $\phi$  on  $a$ .

penalty	$\phi(u)$	$\phi'(u)/u$
abs	$ u $	$\frac{1}{ u }$
log	$\frac{1}{a} \log(1 + a u )$	$\frac{1}{ u (1 + a u )}$
atan	$\frac{2}{a\sqrt{3}} \left( \tan^{-1} \left( \frac{1 + 2a u }{\sqrt{3}} \right) - \frac{\pi}{6} \right)$	$\frac{1}{ u (1 + a u  + a^2 u ^2)}$
rational	$\frac{ u }{1 + a u /2}$	$\frac{1}{ u (1 + a u )^2}$

The OGS algorithm proceeds by iteratively reducing  $x(i)$ ,  $i \in \mathcal{S}$ , toward their optimal values (including zero). The attenuation is multiplicative, so the value never equals zero, even though it may converge to zero. But if a value reaches 'machine epsilon' then a divide-by-zero error may subsequently occur in the implementation. Hence, to avoid possible divide-by-zero errors due to finite precision arithmetic, the OGS algorithm updates  $\mathcal{S}$  at the end of the loop. The small number  $\epsilon$ , may be set to 'machine epsilon', which for single precision floating point is about  $10^{-6}$ .

### IV. EXPERIMENTAL RESULTS

#### A. Example 1: One-Dimensional Signal Denoising

This example compares the proposed non-convex regularized

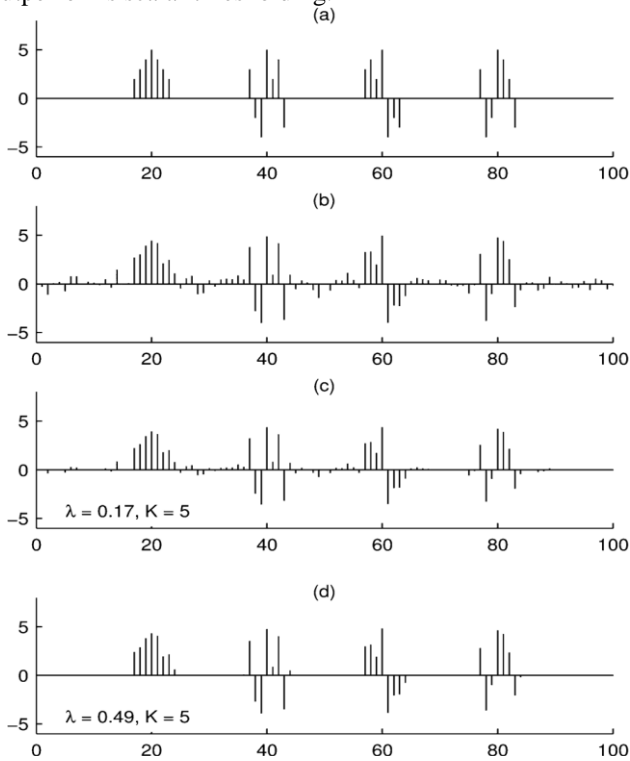
OGS algorithm with the earlier (convex regularized) version of OGS and with scalar thresholding. The SNRs are summarized in below Table.

The noisy signal was obtained by adding white Gaussian noise (AWGN) with SNR of 10 dB.

Param.	Estimator				
	Hard thr.	Soft thr.	OGS[abs]	OGS[log]	OGS[atan]
max SNR	13.84	12.17	12.30	14.52	15.37
$10^{-2}\sigma$	6.74	3.86	8.01	12.07	13.92
$10^{-3}\sigma$	5.05	2.17	6.23	9.69	11.54

For each of soft and hard thresholding, we used the threshold,  $T$ , that maximizes the SNR. The SNR values are summarized in the top row of above Table.

We selected  $T$  and  $\lambda$  for each method, so as to reduce the noise standard deviation,  $\sigma$ , down to  $0.01\sigma$ . The resulting SNRs, given in the second row of Table, are much lower. (This method does not maximize SNR, but it does ensure residual noise is reduced to the specified level.) The low SNR in these cases is due to the attenuation (bias) of large magnitude values. However, it can be observed that OGS, especially with non-convex regularization, significantly outperforms scalar thresholding.



Example 1: Group-sparse signal denoising. (a) Signal; (b) Signal+noise(SNR=10dB) (c) OGS[abs](SNR=12.30db); (d) OGS[atan]

**B. Example 2: Speech Denoising**

This example evaluates the use of the proposed OGS algorithm for the problem of speech enhancement (denoising). We compare the OGS algorithm with several other algorithms. For the evaluation, we use female and male speakers, multiple sentences, two noise levels, and two sampling rates. Throughout this example, we use the non-convex arctangent penalty function with  $a$  set to its maximum value of  $a = (1/K1K2\lambda)$ . In all cases, we use a fixed number of 25 iterations within the OGS algorithm.

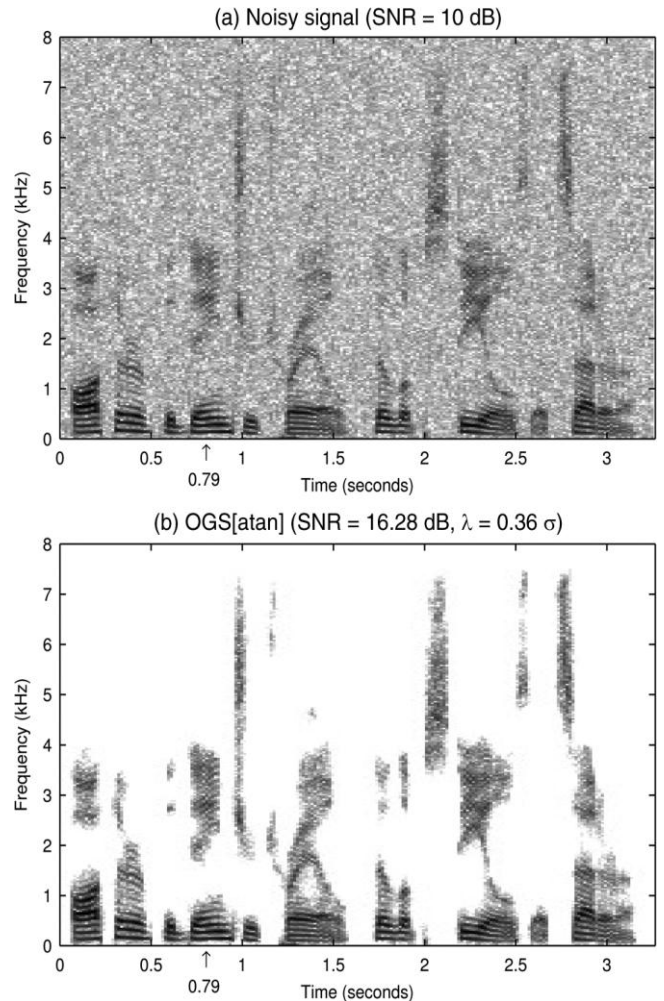


Fig.. Spectrograms before and after denoising (male speaker). (a) Noisy signal. (b) OGS[atan] with group size(K=(8,2)). Gray scale represents decibels.

**Regularization Parameter:**

We have found empirically, that setting to maximize SNR yields speech with noticeable undesirable perceptual artifacts (‘musical noise’). This known phenomenon is due to residual noise in the STFT domain. Therefore, we instead set the regularization parameter, using the noise suppression approach described. In particular, we set so as to reduce the noise standard deviation  $\sigma$ . We have selected this value so as to optimize the perceptual quality of the denoised speech according to informal listening tests. In particular, this value is effective at suppressing the ‘musical noise’ artifact. We also note that this approach leads to greater regularization (higher) than SNR-optimization.

**Group Size:** The perceptual quality of speech denoised using OGS depends on the specified group size. As we apply OGS to a time-frequency spectrogram, the size of the group with respect to both the temporal and spectral dimensions must be specified. We let  $K1$  and  $K2$  denote the number of spectral and temporal samples, respectively.

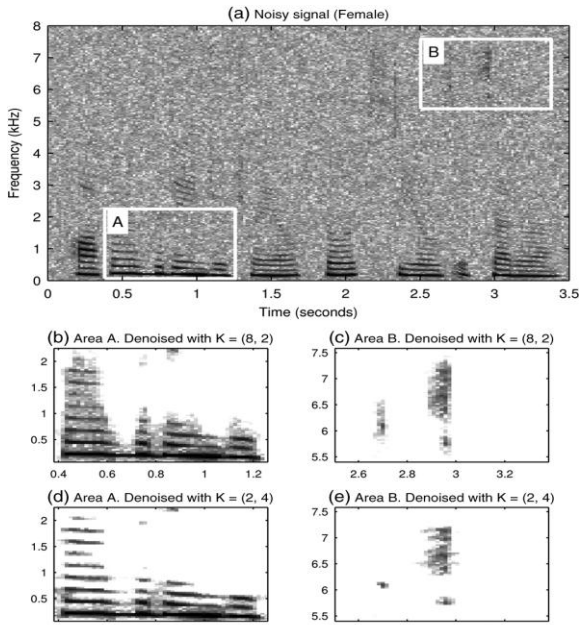


Fig.. Denoised spectrograms; female speaker. (a) Noisy spectrogram with SNR=10dB. (b, c) Areas A and B, denoised with group size (8, 2). (d, e) Areas A and B, denoised with group size (2, 4).

Table V  
(a)

Method	Male / Input SNR (dB)		Female / Input SNR (dB)	
	5	10	5	10
SS	9.44 (10.96)	13.63 (14.99)	13.36 (14.59)	16.86 (17.93)
LMA	10.24 (11.64)	13.30 (15.25)	13.30 (15.16)	15.71 (18.13)
SUB	11.28 (12.31)	13.94 (16.11)	13.39 (15.31)	15.05 (18.48)
BT	12.00 (12.49)	15.61 (16.10)	15.09 (15.69)	18.18 (18.78)
PS	10.75 (12.00)	14.17 (15.73)	12.67 (14.71)	16.39 (18.13)
OGS[abs]	10.48 (12.36)	13.92 (16.00)	12.91 (15.53)	16.24 (18.60)
OGS[atan]	12.93 (12.98)	16.58 (16.58)	15.37 (15.83)	18.68 (19.02)

(b)

Method	Male / Input SNR (dB)		Female / Input SNR (dB)	
	5	10	5	10
SS	10.73 (11.75)	14.57 (15.54)	10.45 (11.59)	14.38 (15.47)
LMA	10.66 (12.00)	13.75 (15.61)	9.34 (11.05)	12.51 (14.85)
SUB	10.83 (12.29)	14.03 (16.06)	9.57 (11.53)	13.25 (15.55)
BT	11.80 (12.48)	15.45 (16.10)	11.54 (12.40)	15.12 (16.00)
PS	10.45 (12.20)	13.64 (15.75)	9.11 (11.20)	13.52 (15.47)
OGS[abs]	9.96 (12.25)	13.42 (15.87)	9.34 (11.91)	12.81 (15.70)
OGS[atan]	12.80 (12.97)	16.41 (16.53)	12.10 (12.62)	15.84 (16.31)

AVERAGE SNR FOR SIX SPEECH ENHANCEMENT ALGORITHMS. (a)  $f_z=16$  kHz (average of 30 samples); (b)  $f_z=8$  kHz (average of 15 samples)

*Algorithm Comparisons:* In Table V we compare the OGS[atan] algorithm with several other speech enhancement algorithms. The table summarizes the output SNR for two sampling rates, male and female speakers, and two input SNR (noise) levels. Each SNR value is averaged over 30 or 15 sentences, depending on the sampling rate. It can be observed that the proposed algorithm, OGS[atan], achieves the highest SNR in each case. (We also note that in all cases, OGS is used not with SNR-optimized, but with the larger, set according to the noise suppression method. The SNR of OGS could be further increased, but at the cost of perceptual quality.

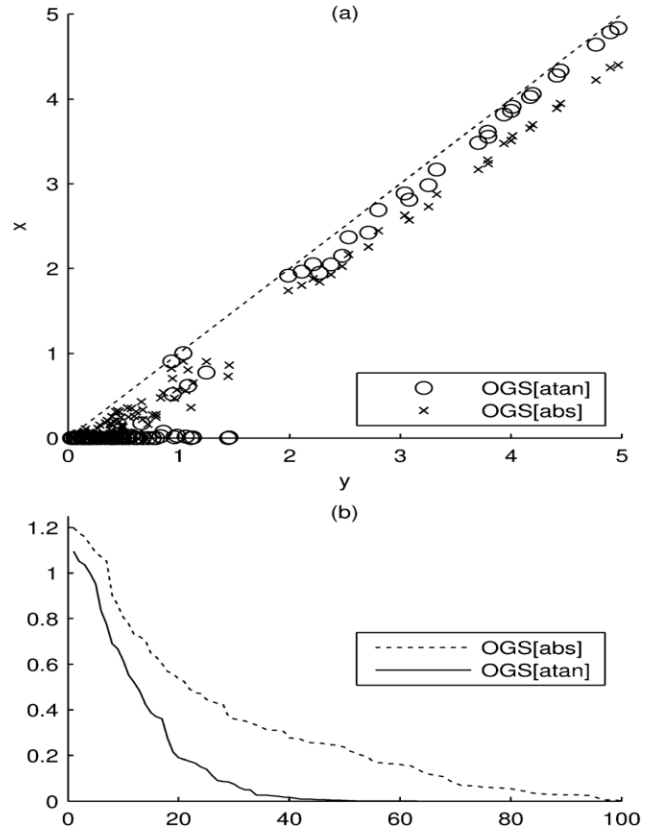


Fig.. Example 1. Comparison of OGS[abs] and OGS[atan] (a) Output versus input; (b) sorted error.

The proposed algorithm, OGS[atan], achieves the highest SNR for both noise levels and genders. For example, for the male speaker with an input SNR of 10 dB, OGS[atan] attains the highest output SNR of 16.58 dB. BT achieves the second highest, 15.61 dB. In terms of perceptual quality, SS and LMA have clearly audible artifacts; BT and PS have slight audible artifacts; OGS[atan], OGS[abs] and SUB have the least audible artifacts. However, SUB has a high computational complexity due to eigenvalue factorization. Compared to OGS[abs] and SUB, OGS[atan] better preserves the perceptual quality of high frequencies. Similar results can be observed for different noise levels and the female speaker.

Empirical Wiener post-processing (EWP) improves the SNR for all methods at all noise levels, but least for OGS[atan]. EWP is effective for increasing SNR because it effectively rescales large-amplitude STFT coefficients that are unnecessarily attenuated by these algorithms (the results of which are biased toward zero). The fact that EWP yields the least improvement for OGS[atan] demonstrates that this algorithm inherently induces less bias than the other algorithms.



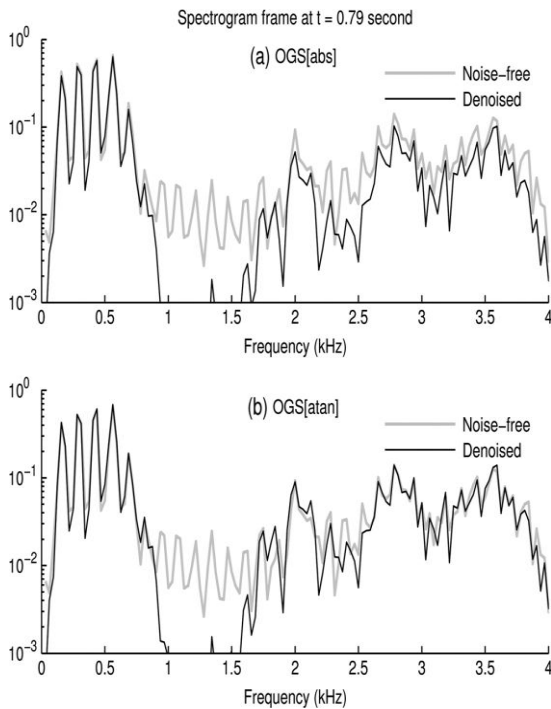


Fig.. Frequency spectrum of denoised spectrograms at t=0.79 seconds.(a) OGS[abs]. (b) OGS[atan]. The group size is K=8,2 in both cases. The noise-free spectrum is in gray.

V. REMARKS

Several aspects of the non-convex regularized OGS algorithm are sufficiently similar to those of the convex regularized OGS algorithm. In particular, remarks regarding the convergence behavior, implementation issues, computational complexity, and relationship of OGS, apply also to the version of OGS presented here. The proximal framework has proven effective for convex optimization problems arising in sparse signal estimation and reconstruction. The proposed non-convex regularized OGS algorithm resembles a proximity operator; however, a proximity operator is defined in terms of a convex penalty function. Hence, the proposed approach appears to fall outside the proximal framework.

The proximal framework has proven effective for convex optimization problems arising in sparse signal estimation and reconstruction. The proposed non-convex regularized OGS algorithm resembles a proximity operator; however, a proximity operator is defined in terms of a convex penalty function. Hence, the proposed approach appears to fall outside the proximal framework. Due to the effectiveness of the proximal framework for solving inverse problems much more general than denoising (e.g. deconvolution), it will be of interest in future work to explore the extent to which the proposed method can be used for more general inverse problems by using proximal-like techniques.

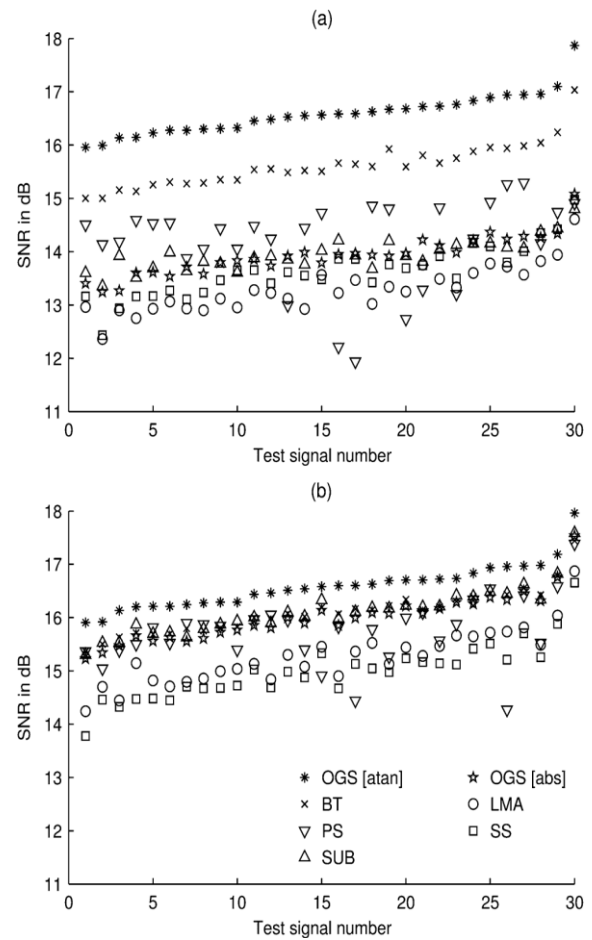


Fig.. SNR comparison of speech enhancement algorithms (30 male sentences, input SNR of 10 dB). Each algorithm is used without EWP (a) and with EWP (b). The sentences are ordered according the SNR of OGS[atan].

VI. CONCLUSION

This paper formulates group-sparse signal denoising as a convex optimization problem with a non-convex regularization term. The regularizer is based on overlapping groups so as to promote group-sparsity. The regularizer, being concave on the positive real line, promotes sparsity more strongly than any convex regularizer can. For several non-convex penalty functions, parameterized by a variable,  $a$ , it has been shown how to constrain  $a$  to ensure the optimization problem is strictly convex. We also develop a fast iterative algorithm for the proposed approach. Numerical experiments demonstrate the effectiveness of the proposed method for speech enhancement.

## REFERENCES

- [1] F. Bach, R. Jenatton, J. Mairal, and G. Obozinski, "Optimization with sparsity-inducing penalties," *Found. Trends Machine Learn.*, vol. 4, no. 1, pp. 1–106, 2012.
- [2] I. Bayram, "Mixed norms with overlapping groups as signal priors," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, May 2011, pp. 4036–4039.
- [3] I. Bayram and O. D. Akyildiz, "Primal-dual algorithms for audio decomposition using mixed norms," *Signal, Image Video Process.*, vol. 8, no. 1, pp. 95–110, Jan. 2014.
- [4] M. Berouti, R. Schwartz, and J. Makhoul, "Enhancement of speech corrupted by acoustic noise," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, April 1979, vol. 4, pp. 208–211.
- [5] A. Blake and A. Zisserman, *Visual Reconstruction*. Cambridge, MA, USA: MIT Press, 1987.
- [6] A. Blumensath, "Accelerated iterative hard thresholding," *Signal Process.*, vol. 92, no. 3, pp. 752–756, 2012.
- [7] 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 in Machine Learn.*, vol. 3, no. 1, pp. 1–122, 2011.
- [8] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge University Press, 2004.
- [9] E. J. Candes, C. A. Sing-Long, and J. D. Trzasko, "Unbiased risk estimates for singular value thresholding and spectral estimators," *IEEE Trans. Signal Process.*, vol. 61, no. 19, pp. 4643–4657, Oct. 2013.
- [10] E. J. Candès, M. B. Wakin, and S. Boyd, "Enhancing sparsity by reweighted  $\ell_1$  minimization," *J. Fourier Anal. Appl.*, vol. 14, no. 5, pp. 877–905, Dec. 2008.
- [11] R. Chartrand, "Exact reconstruction of sparse signals via non-convex minimization," *IEEE Signal Process. Lett.*, vol. 14, no. 10, pp. 707–710, 2007.
- [12] R. Chartrand and B. Wohlberg, "A nonconvex ADMM algorithm for group sparsity with sparse groups," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, May 2013.
- [13] P.-Y. Chen and I. W. Selesnick, "Translation-invariant shrinkage/thresholding of group sparse signals," *Signal Process.*, vol. 94, pp. 476–489, Jan. 2014.
- [14] X. Chen, Q. Lin, S. Kim, J. G. Carbonell, and E. P. Xing, "Smoothing proximal gradient method for general structured sparse regression" *Ann. Appl. Stat.*, vol. 6, no. 2, pp. 719–752, 2012.
- [15] I. Cohen, "Optimal speech enhancement under signal presence uncertainty using log-spectral amplitude estimator," *IEEE Signal Process. Lett.*, vol. 9, no. 4, pp. 113–116, Apr. 2002.
- [16] 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, 2008.
- [17] P. L. Combettes and J.-C. Pesquet, "Proximal splitting methods in signal processing," in *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, H.H. Bauschke, Ed. et al. New York, NY, USA: Springer-Verlag, 2011, pp. 185–212.
- [18] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3d transform-domain collaborative filtering," *IEEE Trans. Image Process.*, vol. 16, no. 8, pp. 2080–2095, Aug. 2007.
- [19] W. Deng, W. Yin, and Y. Zhang, "Group Sparse Optimization by Alternating Direction Method," Rice Univ. CAAM Tech. Rep. TR11-06, 2011, 2011.
- [20] D. L. Donoho and I. M. Johnstone, "Ideal spatial adaptation by wavelet shrinkage," *Biometrika*, vol. 81, no. 3, pp. 425–455, 1994.
- [21] Y. C. Eldar and M. Mishali, "Robust recovery of signals from a structured union of subspaces," *IEEE Trans. Inf. Theory*, vol. 55, no. 11, pp. 5302–5316, Nov. 2009.