Smoothed ℓ_p-ℓ_2 Solvers for Signal Denoising

Jie Yan and Wu-Sheng Lu
Department of Electrical and Computer Engineering
University of Victoria
Victoria, BC, Canada V8W 3P6
jyan@ece.uvic.ca, wslu@ece.uvic.ca

***Abstract*—The basis pursuit denoising refers to the solution of an ℓ_1-ℓ_2 minimization formulation which is well known as an effective method for signal denoising. In this paper we investigate an ℓ_p-ℓ_2 formulation with $p \in (0,1)$ for denoising. Based on an analysis of the discontinuity of the global minimizer of the ℓ_p-ℓ_2 problem with respect to regularization parameter, we propose two smoothed ℓ_p-ℓ_2 solvers for orthogonal basis and overcomplete dictionary respectively. Experimental studies that evaluate the performance of the proposed solvers with various parameter settings are also presented.**

I. Introduction

A well-known technique for noise removal from discrete measurements is the basis pursuit denoising (BPDN) which refers to the solution of a nonsmooth convex ℓ_1-ℓ_2 unconstrained minimization problem [1]. The same ℓ_1-ℓ_2 formulation also finds applications in linear inverse problems and compressive sensing, and many efficient solvers have been proposed, see [2] and the references cited therein. In addition, several authors have investigated a nonconvex extension, namely,

$$\min \ F(\mathbf{s}) = \lambda \|\mathbf{s}\|_p^p + \|\mathbf{\Theta s} - \mathbf{y}\|_2^2 \tag{1}$$

where $\mathbf{y}$ denotes measurement, $\mathbf{\Theta}$ represents a dictionary, $\lambda > 0$ is a regularization parameter and $p \in (0,1)$, who reported results that outperform its ℓ_1-ℓ_2 counterpart [3]–[6]. In this paper, we study problem (1) for signal denoising and propose two fast solvers where $\mathbf{\Theta}$ represents either an orthogonal basis or an overcomplete dictionary. For the case of an orthogonal $\mathbf{\Theta}$, our ℓ_p-ℓ_2 solver is developed based on an analysis of the discontinuity of the global solution of (1). For the case of an overcomplete dictionary, our solver is developed based on a proximal-point technique [7] which converts the problem at hand to a set of scalar-variable problems that can be handled using the preceding solver and allows FISTA [7] type of iteration for fast convergence. The performance of the proposed solvers are evaluated by applying them to signal denoising problems where the results obtained from various parameter settings are compared with each other.

II. Basis pursuit denoising and ℓ_p-ℓ_2 formulation

A. *Basis pursuit denoising*

Let $\mathbf{y}$ be the observation of a signal $\mathbf{x}$ that is contaminated by Gaussian white noise $\mathbf{w}$, i.e., $\mathbf{y} = \mathbf{x} + \mathbf{w}$. Without loss of generality, assume that $\mathbf{x}$ admits a sparse or nearly sparse representation in a suitable dictionary $\mathbf{\Theta}$, namely $\mathbf{x} = \mathbf{\Theta s}$ where $\mathbf{s}$ is sparse. The well-known (BPDN) [1] to recover signal $\mathbf{x}$ from noisy measurement $\mathbf{y}$ refers to the solution of

$$\min \ \lambda \|\mathbf{s}\|_1 + \|\mathbf{\Theta s} - \mathbf{y}\|_2^2 \tag{2}$$

where parameter $\lambda > 0$ depends on the variance of $\mathbf{w}$ as well as the cardinality of dictionary $\mathbf{\Theta}$ [1]. Problem (2) is convex, for which effective solution methods have been developed in the past several years [2].

B. *An ℓ_p-ℓ_2 formulation for denoising*

In compressive sensing, the recovery of a sparse signal $\mathbf{s}$ using noisy linear measurement $\mathbf{y} = \mathbf{\Theta s} + \mathbf{w}$ is formulated as

$$\min \ \|\mathbf{s}\|_0 \ \text{ s.t. } \|\mathbf{\Theta s} - \mathbf{y}\|_2^2 \le \varepsilon \tag{3}$$

where $\|\mathbf{s}\|_0$ denotes the number of nonzero entries in $\mathbf{s}$ and $\varepsilon > 0$ is an upper bound for measurement noise. Problem (3) is usually relaxed to the convex problem

$$\min \ \|\mathbf{s}\|_1 \ \text{ s.t. } \|\mathbf{\Theta s} - \mathbf{y}\|_2^2 \le \varepsilon \tag{4}$$

so as to avoid the combinatorial complexity encountered in solving (3). Note that an unconstrained reformulation of (4) also leads to (2). From this perspective and the relation between (3) and (4), it is natural to investigate the ℓ_p-ℓ_2 problem with $p \in (0,1)$ for signal denoising, where $\|\mathbf{s}\|_p^p = \sum_i |s_i|^p$. We stress that for any $p < 1$, problem (1) is no longer convex. As a result, efficient solvers for problem (2) are not applicable to (1).

III. A smoothed ℓ_p-ℓ_2 solver for orthogonal basis

In this section problem (1) is investigated with an orthogonal $\mathbf{\Theta}$, i.e., $\mathbf{\Theta\Theta}^T = \mathbf{\Theta}^T\mathbf{\Theta} = \mathbf{I}$, and a fixed power $p \in (0,1)$.

A. *Discontinuity of the global solution with respect to λ*

We write the objective function in (1) as

$$F(\mathbf{s}) = \lambda \|\mathbf{s}\|_p^p + \|\mathbf{\Theta}(\mathbf{s} - \mathbf{\Theta}^T\mathbf{y})\|_2^2 = \lambda \|\mathbf{s}\|_p^p + \|\mathbf{s} - \mathbf{c}\|_2^2.$$

where $\mathbf{c} = \mathbf{\Theta}^T\mathbf{y}$. Therefore, minimizing $F(\mathbf{s})$ amounts to solving N scalar optimization problems separately, each with a single-variable objective function $\lambda|s_i|^p + (s_i - c_i)^2$. Therefore, the problem of global minimization of $F(\mathbf{s})$ amounts to global minimization of the single-variable function

$$u(s; \lambda) = \lambda |s|^p + (s - c)^2.$$

(here for simplicity we abuse the notation by dropping subindex i). Without loss of generality we assume $c > 0$ (if

978-1-4673-0046-9/12/$26.00 ©2012 IEEE

$c < 0$, we simply let $c := -c$ as this does not affect the analysis below).

By combining the graphs of λs^p and $(s-c)^2$, it is evident that the minimizers of $u(s;\lambda)$ as a function of s can only occur in $[0, c]$. Over this interval, $u(s;\lambda) = \lambda s^p + (s-c)^2$ is differentiable and a minimizer inside the interval satisfies $u'(s;\lambda) = \lambda p s^{p-1} + 2(s-c) = 0$ with $u''(s;\lambda) = 2 - \lambda p(1-p)s^{p-2} > 0$. In addition, the presence of term $\lambda |s|^p$ yields a notch at $s = 0$ which is either a local or a global minimizer, depending on the value of λ. In effect, there is a value $\hat{\lambda} > 0$ with which the two minimizers are equal and hence both become global minimizers. The $\hat{\lambda}$ and the locations of the two global minimizers, 0 and $\hat{s}$, can be determined by solving the equations $u'(\hat{s};\hat{\lambda}) = 0$ and $u(\hat{s};\hat{\lambda}) = u(0,\hat{\lambda})$ simultaneously. In doing so, we obtain

$$\hat{s} = \frac{2(1-p)c}{2-p} \quad \text{and} \quad \hat{\lambda} = \frac{\hat{s}^{(2-p)}}{1-p}. \tag{5}$$

Note that the $\hat{s}$ in (5) satisfies $0 < \hat{s} < c$ and $u''(\hat{s};\hat{\lambda}) = 2 - p > 0$ hence $\hat{s}$ is indeed a minimizer inside $[0, c]$. From (5) it follows that

$$\hat{\lambda} = \gamma c^{2-p}, \ \gamma = \frac{1}{1-p} \cdot \left[\frac{2(1-p)}{2-p}\right]^{2-p}. \tag{6}$$

For a $\lambda < \hat{\lambda}$, the interior minimizer $s(\lambda)$ determined by $u'(s;\lambda) = 0$ with $u''(s;\lambda) > 0$ is the unique global minimizer of $u(s;\lambda)$; for a $\lambda > \hat{\lambda}$, the origin $s = 0$ becomes the unique global minimizer; and the global minimizer jumps between the origin and the interior point $\hat{s}$ (computed from (5)) as λ varies across the critical value $\hat{\lambda}$ given by (6). Fig. 1 illustrates our analysis for the case of $p = 0.5$ and $c = 1$ in that (6) gives $\hat{\lambda} = 1.0887$. Fig. 1(a)-(c) show the minimizers of $u(s;\lambda)$ for (a) $\lambda = 1.08 < \hat{\lambda}$, (b) $\lambda = \hat{\lambda} = 1.0887$, and (c) $\lambda = 1.09 > \hat{\lambda}$. The global minimizer of $u(s;\lambda)$, denoted by $s^*(\lambda)$, as a function of λ is depicted in Fig. 1(d) where its discontinuity at $\hat{\lambda} = 1.0887$ is evident.

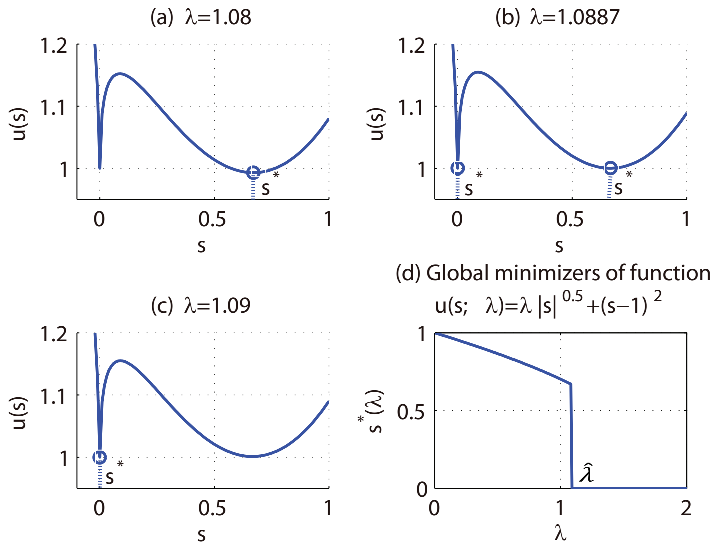


Fig. 1: Global minimizer $s^*(\lambda)$ of $u(s;\lambda) = \lambda |s|^{0.5} + (s-1)^2$ (a) $\lambda = 1.08$, (b) $\lambda = \hat{\lambda} = 1.0887$, (c) $\lambda = 1.09$, (d) discontinuity of $s^*(\lambda)$ at $\hat{\lambda} = 1.0887$.

B. A smoothed ℓ_p-ℓ_2 solver

The discontinuity of $s^*(\lambda)$ is undesirable as it degrades the stability and predictability of the denoising process based on formulation (1). Below we propose a solution strategy that prefers a stable solution rather than a global solution in case parameter λ is in a vicinity of the discontinuity point $\hat{\lambda}$.

Like problem (2), when formulation (1) is employed for denoising, parameter λ is a prescribed positive real. In what follows we assume the value of λ falls within an interval $[\lambda_L, \lambda_H]$. On the other hand, vector $\mathbf{c}$ can be computed based on the given orthonormal basis $\mathbf{\Theta}$ and measurement $\mathbf{y}$ (see Sec. II-A). Using (6) we evaluate the critical $\hat{\lambda}_i$ for each component c_i as $\hat{\lambda}_i = \gamma c_i^{2-p}$ for $1 \le i \le N$. Each component s_i^* of the solution vector $\mathbf{s}^*$ is found as follows:

(a) If $\hat{\lambda}_i \notin [\lambda_L, \lambda_H]$, solution jump will not occur, hence the global solution s_i^* can be found with stability: if $\hat{\lambda}_i < \lambda_L$, set $s_i^* = 0$; if $\hat{\lambda}_i > \lambda_H$, take the minimizer inside $[0, c_i]$ as the solution s_i^*. This solution can be efficiently identified using a one-dimensional search technique such as a golden section or bisection search. We refer the reader to [6] for details.

(b) If $\hat{\lambda} \in [\lambda_L, \lambda_H]$, to prevent solution jump, we take the unique global solution of $u(s;\lambda)$ with $p = 1$ as s_i^*, which is simply computed by a soft-shrinkage operation as $s_i^* = \text{sgn}(c_i) \cdot \max\{|c_i| - \lambda/2, 0\}$.

Although not a truly global solver, the solution procedure proposed above eliminates the jump phenomenon and offers a stable yet nearly global solution $\mathbf{s}^*$. In the rest of the paper, we call it the smoothed ℓ_p-ℓ_2 solver whose performance for denoising will be examined in Sec. V.

C. Parallel implementation of the smoothed ℓ_p-ℓ_2 solver

The proposed solver admits a fast implementation which solves the N single-variable ℓ_p-ℓ_2 problem in parallel. For notation simplicity, denote by $\mathbf{a}.*\mathbf{b}$ the component-wise product of vectors $\mathbf{a}$ and $\mathbf{b}$, and by $\mathbf{a}.^p$ the vector whose ith component is a_i^p. Let Λ be a subset of $\{1, 2, ..., N\}$, $\mathbf{c}$ be a vector of length-N and $\mathbf{b}$ be a vector of length-K. We use $\mathbf{c}(\Lambda)$ to denote a vector of length K that retains those components of $\mathbf{c}$ whose indices are in Λ; $\mathbf{c}(\Lambda) = \mathbf{b}$ to denote a vector of length-N obtained by updating the components of $\mathbf{c}$, whose indices are in Λ, with those of $\mathbf{b}$. A step-by-step description of a parallel implementation (we refer it to as Algorithm 1) of the proposed ℓ_p-ℓ_2 solver is given in Table I where the data are processed in a vector-wise rather than component-wise manner.

IV. A Smoothed ℓ_p-ℓ_2 Solver for Overcomplete Dictionary

In this section, we investigate the circumstance that $\mathbf{\Theta} \in R^{M \times N}$ with $M < N$ is an overcomplete dictionary in space

Input	$\mathbf{c}$, λ, p, λ_L and λ_H.
Output	$\mathbf{s}^*$.
Step 1	Compute $\mathbf{c}_+ = \text{sign}(\mathbf{c}).*\mathbf{c}$.
Step 2	Compute $\hat{\mathbf{s}} = \dfrac{2(1-p)\mathbf{c}_+}{2-p}$ and $\hat{\boldsymbol{\lambda}} = \dfrac{\hat{\mathbf{s}}.^{2-p}}{1-p}$.
Step 3	Define $\mathcal{J} = \{i : \hat{\lambda}_i \in [\lambda_L, \lambda_H]\}$ and $\mathcal{C} = \{i : \hat{\lambda}_i \notin [\lambda_L, \lambda_H]\}$. Define $\mathbf{c}_{\mathcal{J}} = \mathbf{c}(\mathcal{J})$ and $\mathbf{c}_{\mathcal{C}} = \mathbf{c}(\mathcal{C})$.
Step 4	Compute $\mathbf{s}_{\mathcal{J}} = \text{sign}(\mathbf{c}_{\mathcal{J}}).*\max(\|\mathbf{c}_{\mathcal{J}}\| - \lambda/2, 0)$ and $\mathbf{s}_{\mathcal{C}} = \text{argmin}\ \{\lambda\|\mathbf{s}\|_p^p + \|\mathbf{s}-\mathbf{c}_{\mathcal{J}}\|_2^2\}$.
Step 5	Set $\mathbf{s}^*(\mathcal{J}) = \mathbf{s}_{\mathcal{J}}$ and $\mathbf{s}^*(\mathcal{C}) = \mathbf{s}_{\mathcal{C}}$. Return $\mathbf{s}^*$.

TABLE I: Algorithm 1 for (1) with orthogonal basis $\boldsymbol{\Theta}$

R^M. We deal with the non-orthogonality of $\boldsymbol{\Theta}$ by an iterative technique that is in spirit similar to a proximal-point method employed in [7]: iterate $\mathbf{s}_k$ in the kth iteration is updated to

$$\mathbf{s}_{k+1} = \underset{\mathbf{s}}{\text{argmin}}\ \{\lambda\|\mathbf{s}\|_p^p + \frac{L}{2}\|\mathbf{s}-\mathbf{c}_k\|_2^2\} \tag{7}$$

where $\mathbf{c}_k = \mathbf{s}_k - \frac{2}{L}\boldsymbol{\Theta}^T(\boldsymbol{\Theta}\mathbf{s}_k - \mathbf{y})$ and L is the Lipschitz constant of the gradient of $\|\boldsymbol{\Theta}\mathbf{s}-\mathbf{y}\|_2^2$ given by $L = 2\lambda_{\max}(\boldsymbol{\Theta}^T\boldsymbol{\Theta})$. Note that for an orthogonal basis $\boldsymbol{\Theta}$, we have $L = 2$, $\mathbf{c}_k = \boldsymbol{\Theta}^T\mathbf{y} = \mathbf{c}$ (see Sec. III-A) and (7) becomes $\mathbf{s}^* = \text{argmin}_{\mathbf{s}}\ \{\lambda\|\mathbf{s}\|_p^p + \|\mathbf{s}-\mathbf{c}\|_2^2\}$ which is exactly the case addressed Sec. III. Also note that the formulation differs from that of [7] as here we deal with a nonconvex objective function because $p \in (0,1)$. The primary reason to emply (7) is that it is again a separable objective function whose solution was analyzed in detail in Sec. III. Furthermore, formulation (7) allows us to incorporate FISTA [7] type of iteration into this formulation so as to accelerate the algorithm without substantial increase in computational complexity. Essentially a FISTA iteration modifies vector $\mathbf{c}_k$ to $\mathbf{c}_k = \mathbf{b}_k - \frac{2}{L}\boldsymbol{\Theta}^T(\boldsymbol{\Theta}\mathbf{b}_k - \mathbf{y})$ where $\mathbf{b}_k$ is updated using two previous iterates $\mathbf{s}_{k-1}$ and $\mathbf{s}_{k-2}$, see Table II for more algorithmic details. Below we refer the algorithm proposed above as Algorithm 2.

Input	$\mathbf{y}$, $\boldsymbol{\Theta}$, λ, p, λ_L, λ_H and $\mathbf{s}_0$.
Output	$\mathbf{s}^*$.
Step 1	Compute the Lipschitz constant $L = 2\lambda_{\max}(\boldsymbol{\Theta}^T\boldsymbol{\Theta})$. Set the number of iterations K.
Step 2	Set $\mathbf{b}_1 = \mathbf{s}_0$, $t_1 = 1$ and $k = 1$.
Step 3	Compute $\mathbf{c}_k = \frac{2}{L}\boldsymbol{\Theta}^T(\mathbf{y} - \boldsymbol{\Theta}\mathbf{b}_k) + \mathbf{b}_k$, apply Algorithm 1 to solve $\mathbf{s}_k = \text{argmin}_{\mathbf{s}}\ \{\frac{2\lambda}{L}\|\mathbf{s}\|_p^p + \|\mathbf{s}-\mathbf{c}_k\|_2^2\}$ and compute $t_{k+1} = \dfrac{1+\sqrt{1+4t_k^2}}{2}$, $\mathbf{b}_{k+1} = \mathbf{s}_k + \left(\dfrac{t_k - 1}{t_{k+1}}\right)(\mathbf{s}_k - \mathbf{s}_{k-1})$, $k = k+1$
Step 4	If $k = K$, output $\mathbf{s}_k$ as solution $\mathbf{s}^*$ and terminate; otherwise repeat from Step 3.

TABLE II: Algorithm 2 for (1) with overcomplete dictionary $\boldsymbol{\Theta}$

V. Performance Evaluation and Comparisons

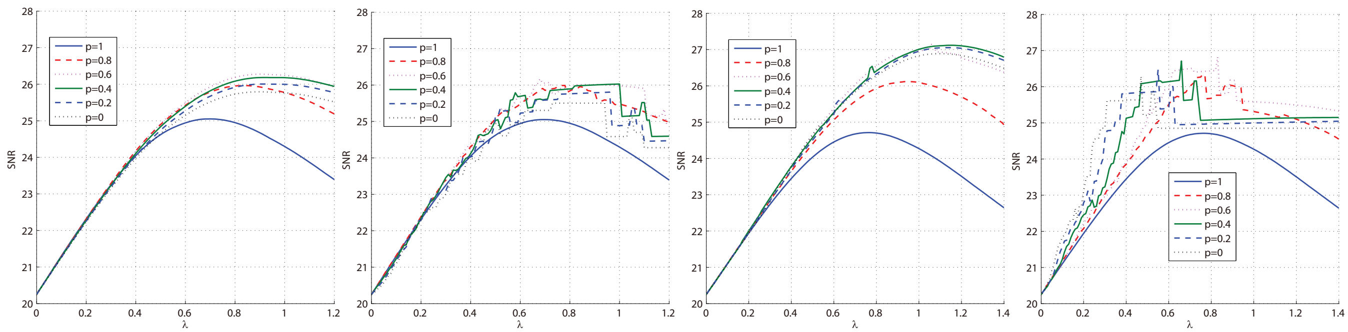
The ℓ_p-ℓ_2 solvers proposed in Sec. III and IV were applied to denoising 1-D measurements and the results obtained from various settings are compared with each other.

A. Signal denoising with orthogonal basis

A test signal of length $N = 256$ known as "HeaviSine" [8] was corrupted with additive white Gaussian noise $\mathbf{n}$ with zero mean and standard deviation $\sigma = 0.3$. The signal-to-noise ratio (SNR) of the noisy signal was found to be 20.25dB. Matrix $\boldsymbol{\Theta}$ represents an orthogonal 8-level Daubechies wavelet D8 basis. The lower and upper bounds for λ were set to $\lambda_L = 0$ and $\lambda_H = 1.2$. With p fixed as one of the six values $\{1, 0.8, 0.6, 0.4, 0.2, 0\}$, Algorithm 1 was applied to solve problem (1) with 121 uniformly placed λ from 0 to 1.2. The SNR obtained versus λ for each p are depicted as six curves in Fig. 2(a). It is observed that (a) for each fixed λ, using a $p < 1$ offers improved SNR relative to that obtained with $p = 1$ (BPDN); (b) for a fixed p, the SNR is a smooth function of λ, and the value of λ achieving peak SNR gradually increases as p decreases; and (c) the best performance was achieved with $p = 0.6$ at $\lambda = 0.92$ offering an SNR of 26.27dB compared to an SNR of 25.05dB obtained by BPDN (with $p = 1$ at $\lambda = 0.70$). For comparison, Fig. 2(b) depicts the SNR profiles obtained by global solutions of (1) (if a λ used in (1) happened to be equal to the critical value $\hat{\lambda}$ in (6), the global solution $\hat{\mathbf{s}}$ given by (5) was used). Most of the SNRs associated with $p < 1$ exhibit considerable oscillations – a sharp departure from the smooth concave SNR profiles obtained from Algorithm 1 where the peak SNR (for each p) is unique and predicable.

B. Signal denoising with overcomplete dictionary

The "HeaviSine" signal $\mathbf{x}$ and its noisy version constructed in part A were also used here. A $\boldsymbol{\Theta} = [\boldsymbol{\Theta}_1\ \boldsymbol{\Theta}_2]$ of size 256×512 with $\boldsymbol{\Theta}_1$ the 8-level Daubechies D8 wavelet basis and $\boldsymbol{\Theta}_2$ the 1-level Haar wavelet basis was used as an overcomplete dictionary. The lower and upper bounds of λ were set to $\lambda_L = 0$ and $\lambda_H = 1.4$. Because both $\boldsymbol{\Theta}_1$ and $\boldsymbol{\Theta}_2$ are orthogonal, the Lipschitz constant $L = 2\lambda_{\max}(\boldsymbol{\Theta}\boldsymbol{\Theta}^T) = 4$. Algorithm 2 was applied to each of the six cases of $p \in \{1, 0.8, 0.6, 0.4, 0.2, 0\}$, where problem (1) was solved for each of 141 λ's that were equally placed over $[0, 1.4]$. In our implementation of Algorithm 2, the solution $\mathbf{s}(\lambda)$ obtained from a given λ was used as the initial point for the algorithm to proceed with the subsequent value of λ. The use of this better initial point was found helpful in reducing the number of iterations required. The SNRs obtained are shown in Fig. 2(c). We see that the observations made in part A



(a) (b) (c) (d)

Fig. 2: SNRs produced by denoising signal "HeaviSine" by (a) Algorithm 1 with orthogonal $\mathbf{\Theta}$; (b) global solution with orthogonal $\mathbf{\Theta}$; (c) Algorithm 2 with overcomplete $\mathbf{\Theta}$; and (d) replacing the 2nd sub-step in Step 3 of Algorithm 2 by $\mathbf{s}_k =$ global minimizer of $\{\frac{2\lambda}{L}||\mathbf{s}||_p^p + ||\mathbf{s} - \mathbf{c}_k||_2^2\}$ with overcomplete $\mathbf{\Theta}$.

for the case of orthogonal basis also hold here, except that the best performance in the present case was achieved with $p = 0.4$ at $\lambda = 1.17$, offering an SNR of 27.12dB which is 0.9dB higher than the maximum SNR obtained by Algorithm 1. For comparison, Fig. 2(d) depicts the SNRs obtained by replacing the 2nd sub-step in Step 3 of Algorithm 2 with $\mathbf{s}_k =$ global minimizer of $\{\frac{2\lambda}{L}||\mathbf{s}||_p^p + ||\mathbf{s} - \mathbf{c}_k||_2^2\}$. Like the case in part A, the SNRs with $p < 1$ show a great deal of instability with respect to λ.

Fig. 3 illustrates the clean "HeaviSine" signal, the noise-corrupted signal, the denoised signal obtained by BPDN at $p = 1$ and $\lambda = 0.75$ and the denoised signal obtained by Algorithm 2 with $p = 0.4$ and $\lambda = 1.17$ using the overcomplete dictionary.

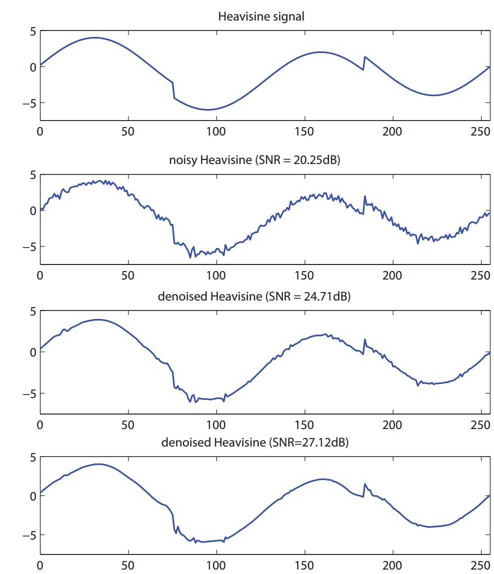


Fig. 3: From top to bottom: original HeaviSine; its noisy version; denoised by BP with $p = 1$ and $\lambda = 0.7$; and denoised by Algorithm 2 with $p = 0.4$ and $\lambda = 1.17$.

VI. Conclusion

Two smoothed ℓ_p-ℓ_2 solvers with $p \in (0, 1)$ for signal spaces with orthogonal basis or overcomplete dictionary have been proposed. Both solvers are computationally efficient because the solver with orthogonal basis is non-iterative while the solver with overcomplete dictionary admits FISTA type iterations for fast convergence. By applying them to signal denoising problems, the proposed solvers are demonstrated to outperform their ℓ_1-ℓ_2 counterparts.

Acknowledgement

The work was supported by the Natural Sciences and Engineering Research Council of Canada.

References

[1] S. Chen, D. Donoho, and M. Saunders, "Atomic decomposition by basis pursuit," *SIAM Review*, vol. 43, no. 1, pp. 129–159, 2001.

[2] M. Zibulevsky and M. Elad, "L1-L2 optimization in signal and image processing," *Signal Processing Magazine, IEEE*, vol. 27, no. 3, pp. 76–88, 2010.

[3] R. Chartrand, "Exact reconstruction of sparse signals via nonconvex minimization," *Signal Processing Letters, IEEE*, vol. 14, no. 10, pp. 707–710, 2007.

[4] R. Chartrand and W. Yin, "Iteratively reweighted algorithms for compressive sensing," in *ICASSP 2008*, pp. 3869–3872.

[5] S. Foucart and M. Lai, "Sparsest solutions of underdetermined linear systems via lq-minimization for $0 < q < 1$," *Applied and Computational Harmonic Analysis*, vol. 26, no. 3, pp. 395–407, 2009.

[6] J. Yan and W.-S. Lu, "New algorithms for sparse representation of discrete signals based on ℓ_p-ℓ_2 optimization," in *PacRim 2011*, pp. 73-78, Victoria, BC, Aug. 2011.

[7] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM Journal on Imaging Sciences*, vol. 2, no. 1, pp. 183–202, 2009.

[8] D. Donoho and I. Johnstone, "Adapting to unknown smoothness via wavelet shrinkage," *Journal of the American Statistical Association*, vol. 90, no. 432, pp. 1200–1224, 1995.