1c. Evaluate $S_j \triangleq \sum_{d=1}^{N-1} L_d^j$ where $j = 1, 2, \ldots, 2(N-1)$.

Step 2: **Forming a new simplex**

2a. Determine the worst case vertex $\mathbf{L}^W$ where $W \triangleq \operatorname{argmin}_{1\le j\le 2(N-1)} S_j$ and the second worst case vertex $\mathbf{L}^w$ where $w \triangleq \operatorname{argmin}_{1\le j\le 2(N-1), j\neq W} S_j$.

2b. Compute the reflection vertex, $\mathbf{r} \triangleq \langle (1+\alpha)\mathbf{T} - \alpha\mathbf{L}^W \rangle$ of the worst case vertex $\mathbf{L}^W$ where $\mathbf{T} \triangleq (1/(2(N-1)-1)) \sum_{j=1, j\neq W}^{2(N-1)} \mathbf{L}^j$ and $\alpha = 0.5$. If $r_i \ge M_i, 1 \le i \le N-1$, then set $r_i = M_i$.

2c. Find the new vertex $\mathbf{N}$ as follows:

1) If $\mathbf{r}$ satisfies constraint (12), proceed to **2c** 2) in Step 2. If not, set $\mathbf{r} = \langle (1/2)(\mathbf{T}+\mathbf{r}) \rangle$ and redo this step.
2) If $\sum_{d=1}^{N-1} r_d > \sum_{d=1}^{N-1} L_d^w$, then set $\mathbf{N} = \mathbf{r}$. Otherwise, set $\mathbf{r} = \langle (1/2)(\mathbf{T}+\mathbf{r}) \rangle$, and then go back to **2c** 1) in Step 2.

2d. Replace $\mathbf{L}^W$ with $\mathbf{N}$.

2e. If the values of S_j for all $1 \le j \le 2(N-1)$ are identical, the optimum vertex is chosen as any one of the vertices in the simplex and the optimization procedure terminates. Otherwise, go back to Step 2 and start a new round to evolve the simplex.

REFERENCES

[1] G. J. Foschini, "Layered space time architecture for wireless communication in a fading environment when using multi-element antennas," *Bell Labs. Tech. J.*, vol. 1, no. 2, pp. 41–59, 1996.

[2] P. W. Wolniansky, G. J. Foschini, G. D. Golden, and R. A. Valenzuela, "V-BLAST: An architecture for realizing very high data rates over the rich-scattering wireless channel," in *Proc. IEEE ISSSE*, Sep. 1996, pp. 295–300.

[3] B. M. Hochwald and S. ten Brink, "Achieving near-capacity on a multiple-antenna channel," *IEEE Trans. Commun.*, vol. 51, no. 3, pp. 389–399, Mar. 2003.

[4] K. J. Kim and R. A. Iltis, "Joint detection and channel estimation algorithm for QS-CDMA signals over time-varying channels," *IEEE Trans. Commun.*, vol. 50, no. 5, pp. 845–855, May 2002.

[5] K. Higuchi, H. Kawai, N. Maeda, and M. Sawahashi, "Adaptive selection of surviving symbol replica candidates based on maximum reliability in QRM-MLD for OFCDM MIMO multiplexing," in *Proc. IEEE GLOBECOM*, Dec. 2004, pp. 2480–2486.

[6] J. B. Anderson and S. Mohan, "Sequential coding algorithms: A survey and cost analysis," *IEEE Trans. Commun.*, vol. 32, no. 2, pp. 169–176, Feb. 1984.

[7] K. J. Kim, J. Yue, R. A. Iltis, and J. D. Gibson, "A QRD-M/Kalman filter-based detection and channel estimation algorithm for MIMO-OFDM systems," *IEEE Trans. Wireless Commun.*, vol. 4, no. 2, pp. 710–721, Mar. 2005.

[8] K. J. Kim, T. Reid, and R. A. Iltis, "Iterative soft-QRD-M for turbo coded MIMO-OFDM systems," *IEEE Trans. Commun.*, vol. 56, no. 2, pp. 1043–1046, Jul. 2008.

[9] G. Strang, *Linear Algebra and Its Applications*, 3rd ed. Pacific Grove, CA: Brooks/Cole, 1988.

[10] R. J. Muirhead, *Aspects of Multivariate Statistical Theory*. New York: Wiley, 1982.

[11] K. Itoh, S. Kimura, and T. Sakata, "VLSI memory technology: Current status and future trends," in *Proc. IEEE ESSCIRC*, Sep. 1999, pp. 3–10.

[12] Y. Ma and L. Wanhammar, "A hardware efficient control of memory addressing for high-performance FFT processors," *IEEE Trans. Signal Process.*, vol. 48, no. 3, pp. 917–921, Mar. 2000.

[13] K. Inoue, V. G. Moshnyaga, and K. Murakami, "Reducing power consumption of instruction ROMs by exploiting instruction frequency," in *Proc. IEEE APCCAS*, 2002, pp. 1–6.

[14] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes in C*. Cambridge, U.K.: Cambridge Univ. Press, 1988.

[15] A. Suzuki and K. Kuzumoto, "Data slicer, data slicing method, and amplitude evaluation value setting method," U.S. Patent 2004/0095510, May 20, 2004.

[16] *IEEE Standard for Local and Metropolitan Area Networks*, IEEE 802.16d, 2005.

[17] J. Hagenauer, E. Offer, and L. Papke, "Iterative decoding of binary block and convolutional codes," *IEEE Trans. Inf. Theory*, vol. 42, no. 2, pp. 429–445, Mar. 1996.

[18] Y. L. C. de Jong and T. J. Willink, "Iterative tree search detection for MIMO wireless systems," *IEEE Trans. Commun.*, vol. 53, no. 6, pp. 930–935, Jun. 2005.

[19] S. L. Loyka and J. R. Mosig, "Channel capacity of $\boldsymbol{N}$-antenna BLAST architecture," *IEE Electron. Lett.*, vol. 36, no. 7, pp. 660–661, Mar. 2000.

[20] C. Ahn, "Parallel detection algorithm using multiple QR decomposition with permuted channel matrix for SDM/OFDM," *IEEE Trans. Veh. Technol.*, vol. 57, no. 4, pp. 2578–2582, Mar. 2008.

Compressive Sensing Reconstruction With Prior Information by Iteratively Reweighted Least-Squares

Cristiano Jacques Miosso, Ricardo von Borries, M. Argàez, L. Velazquez, C. Quintero, and C. M. Potes

***Abstract*—Iteratively reweighted least-squares (IRLS) algorithms have been successfully used in compressive sensing to reconstruct sparse signals from incomplete linear measurements taken in nonsparse domains. The underlying optimization problem corresponds to finding the vector that solves the ℓ_p minimization while explaining the measurements, and IRLS allows to easily control the used value of p, with effect on the number of required measurements. In this paper, we propose a weighting strategy in the reconstruction method based on IRLS in order to add prior information on the support of the sparse domain. Our simulation results show that the use of prior knowledge about positions of at least some of the nonzero coefficients in the sparse domain leads to a reduction in the number of linear measurements required for unambiguous reconstruction. This reduction occurs for all values of p, so that a further reduction can be achieved by decreasing p and using prior information. The proposed weighting scheme also reduces the computational complexity with respect to the IRLS with no prior information, both in terms of number of iterations and computation time.**

***Index Terms*—Compressive sensing, iteratively reweighted least-squares, prior information, sensor networks, sparsity.**

I. INTRODUCTION

This paper addresses compressive sensing to show how prior information on the support region of the sparse domain of a signal can be added to the reconstruction procedure using the iteratively reweighted least-squares (IRLS) method. The use of the additional prior information is shown to have advantages in terms of number of required measurements, convergence time and number of iterations. We start by briefly describing the sparsity condition and how it is exploited in

Manuscript received May 07, 2008; accepted January 28, 2009. First published March 10, 2009; current version published May 15, 2009. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Subhrakanti Dey. This work is supported in part by NSF CNS-0709438 and by the Texas Instruments Foundation Endowed Scholarship and in part by Army ARL W911NF-07-2-0027 and NSF HRD-0734825.

C. J. Miosso, R. von Borries, and C. M. Potes are with the Department of Electrical and Computer Engineering, University of Texas at El Paso, El Paso, TX 79968 USA (e-mail: miosso@ieee.org).

M. Argàez, L. Velazquez, and C. Quintero are with the Numerical Optimization Group, Department of Mathematical Sciences, University of Texas at El Paso, El Paso, TX 79968 USA.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2009.2016889

the context of compressive sensing to allow the reconstruction of signals from limited linear measurements. We then exploit two important points that can be used to improve the reconstruction procedure: prior information and a particular weighting scheme in the IRLS method.

Prior information on the positions of nonzero coefficients in the sparse domain allows the number of initially taken measurements to be reduced without leading to ambiguity during the reconstruction stage [1], [2]. In addition, an iterative procedure based on IRLS allows the reconstruction problem to be tackled in a computationally efficient way, with the possibility of easily reducing the value of p in the ℓ_p minimization, which affects the number of iterations, the computation time and the number of required measurements [3]. Using these two results, we show that prior information on the support region of the sparse domain can be added to the basic framework of IRLS, by conducting a modification in the weights used at each iteration. This information leads to a particular strategy for the choice of weights in the IRLS, which we describe and evaluate.

Compressive sensing allows discrete-time signals having a sparse representation in some domain to be unambiguously represented by a limited number of linear measurements [4], [5]. While sparsity is the only signal characteristic that is usually assumed during reconstruction from a set of linear measurements, other forms of prior information about the signal's structure have been investigated as a way to improve reconstruction. In [6] and [7], for instance, a connected tree structure is assumed in the wavelet domain, which restricts the class of signals that can be reconstructed to be piecewise smooth.

Other forms of prior information used in signal reconstruction are also found in the literature. Garcia–Frias and Esnaola show that if a signal is a realization of a stochastic process, prior information about this process (such as second order statistics) can improve the signal's reconstruction from a limited number of measurements [8]. The prior information used in this approach refers to a statistical description of the source that generates the analyzed signals.

In all these contexts, a sparse representation is assumed, and some form of prior information on this representation is used to improve signal reconstruction. Elad and Aharon, on the other hand, exploit sparse and redundant representations of images, and define global image priors that force sparsity over patches in images, to develop a method for image denoising [9], [10].

We exploit a different type of prior information, represented by positions of the support of the signal's sparse representation, in order to reduce the required number of measurements and the computational complexity during the reconstruction stage.

In different applications where sparse signals must be reconstructed from limited linear measurements, incomplete or complete prior information on the support of the sparse domain may be available, and this information can be used to reduce the computation time and the required number of measurements. As an example, in X-ray medical tomographic imaging, if a sparse representation is given by the gradient of the image to be reconstructed, as in [11], some of the positions of the nonzero values of the sparse domain can be made available from the regions of highest derivatives in images in a corresponding medical record. Also, if several successive slices are being analyzed using a tomographic technique, to reconstruct a tridimensional structure, prior information on the support of the domain of a slice can be obtained from the highest gradient values of the previously reconstructed slice.

Another condition in which prior information on the support of the sparse representation can be made available is through the analysis of the intermediary results of the iterative reconstruction algorithm. In [5], a procedure is described by which the highest iterates obtained in a single reconstruction stage are used to determine some of the positions of the support. The next stages then use the information on these positions to improve the reconstruction.

The purpose of this paper is then to show how this prior information on the support of the sparse domain can be efficiently used in the IRLS reconstruction procedure. We first pose the reconstruction problem using prior information to restrict the class of signals that can be reconstructed, thus allowing less linear measurements to represent a signal without generating ambiguity. Next, we show how the reconstruction problem can be tackled using the IRLS and how the prior information can be added to this method through an appropriate weighting scheme. Finally, we show that the resulting algorithm leads to a reduction in the number of required measurements to attain reconstruction, as well as a reduction in the total number of iterations and computation time.

The remaining of the paper is organized as follows. Section II briefly reviews the basic concepts of compressive sensing and presents the idea of adding prior information on the support of the sparse representation to the reconstruction procedure. Following, Section III presents the problem formulation using IRLS, and specifically proposes a weighting strategy for adding prior information. The corresponding reconstruction procedure is described, leading to the algorithm proposed in Section IV. Section V presents the numerical results we obtained by applying this algorithm to discrete-time signals having a sparse representation in an arbitrary, random domain. Finally, Section VI presents our conclusions.

II. Compressive Sensing and Prior Information

The compressive sensing framework allows the efficient representation of discrete-time signals that are sparse in some known domain. The main characteristic common to these signals is that their projections on the basis functions of the sparse domain are mostly zero. This distinguishing property ultimately means that a signal of length N can be unambiguously represented by $\eta < N$ values by taking an appropriate transformation and coding both the nonzero coefficients and their positions.

An important discovery in compressive sensing was that random linear measurements are generally valid for the reconstruction of signals that are sparse in some domain [4], [12]. In this context, let $\mathbf{x}$ represent an N-dimensional signal with an η-sparse transform $\hat{\mathbf{x}}$ ($\eta < N$ coefficients of $\hat{\mathbf{x}}$ are nonzero) and $\mathbf{T}$ represent the transformation matrix, meaning

$$\hat{\mathbf{x}} = \mathbf{T}\mathbf{x}.$$

Then, if $\mathbf{M}_{\ell\times N}$, with $\ell < N$, is a random matrix with normal independent identically distributed entries, then the linear measurements defined by

$$\mathbf{b} = \mathbf{M}\mathbf{x} \tag{1}$$

allow the determination of all the N components of $\mathbf{x}$ provided that the amount ℓ of measurements taken is high enough compared to the sparsity of $\hat{\mathbf{x}}$ [12].

Although the ℓ linear measurements defined by (1) unambiguously represent the original N samples of $\mathbf{x}$, a problem still remains on how to compute these samples from the available measurements. This problem is commonly referred to as signal reconstruction; it is approached by finding the sparsest vector $\hat{\mathbf{x}}$ such that its inverse transform $\mathbf{T}^{-1}$ generates the same given measurements. The key idea is to indirectly solve the minimization problem

$$\min_{\hat{\mathbf{x}}} \|\hat{\mathbf{x}}\|_0, \quad \text{subject to } \mathbf{M}\mathbf{T}^{-1}\hat{\mathbf{x}} = \mathbf{b} \tag{2}$$

where $\|\hat{\mathbf{x}}_0\|$ corresponds to the number of nonzero components of the vector $\hat{\mathbf{x}}$.

Note that a direct approach to (2) leads to combinatorial complexity, which is not viable even for moderately sized signals. An approximate solution that is largely used is the minimization of the 1-norm instead of the objective function $\|\hat{\mathbf{x}}_0\|$ previously defined, so the reconstruction problem in this case is

$$\min_{\hat{\mathbf{x}}} \|\hat{\mathbf{x}}\|_1, \quad \text{subject to } \mathbf{A}\hat{\mathbf{x}} = \mathbf{b} \tag{3}$$

where $\mathbf{A} = \mathbf{M}\mathbf{T}^{-1}$.

The solution to (3) leads to polynomial complexity and it is possibly the most common approach to signal reconstruction in compressive sensing. A second approach is the ℓ_p minimization of $\hat{\mathbf{x}}$, with $0 < p < 1$ or, equivalently,

$$\min_{\hat{\mathbf{x}}} \frac{1}{2}\|\hat{\mathbf{x}}\|_p^p, \quad \text{subject to } \mathbf{A}\hat{\mathbf{x}} = \mathbf{b} \tag{4}$$

and it has been shown that, by reducing the value of p, it is possible to reduce the number of required linear measurements ℓ with respect to that attained for $p = 1$ [3]. In [13], the minimization of the p-norm-like diversity measures, for $p \leq 1$, as well as of the Gaussian and Shannon entropies, is used also to compute sparse solutions to underdetermined systems, in the problem of optimal basis selection; the approach includes the ℓ_p minimization case when $p > 0$.

A possible approach to (4) is based on IRLS, as we will discuss in Section III. An advantage of this method over interior point methods is that it allows the reduction of p in the ℓ_p minimization in a straightforward manner, so that the same algorithm implemented for a certain value of p can be used for a different one by changing a single parameter. A different approach is presented in [13], based on a factorization of the gradient of the Lagrangian function and on the successive relaxation of this function.

In this paper, we propose a particular weighting strategy in the IRLS approach in order to add prior information on the support region of the sparse representation. As it has been shown in [1] and [2], if such information is available it is possible to reduce the amount of taken measurements and still unambiguously reconstruct the underlying signal. In fact, let Φ be the subset of positions in $\{1, 2, \ldots, N\}$ which are known to belong to the support region of $\hat{\mathbf{x}}$, meaning

$$\hat{x}_k \neq 0 \; \forall \, k \in \Phi. \tag{5}$$

It was shown in [1] that the information represented by (5) can be added to the reconstruction procedure based on the minimization of the 1-norm, by solving, instead of (3),

$$\min_{\hat{\mathbf{x}}} \sum_{\substack{k=1 \\ k \notin \Phi}}^{N} |\hat{x}_k|, \quad \text{subject to } \mathbf{A}\hat{\mathbf{x}} = \mathbf{b}. \tag{6}$$

In [1] and [2], problem (6) is solved by using an interior point method approach. The corresponding simulation results show that, for a prespecified frequency of correct reconstructions, (6) leads to a reduction by φ in the number of required measurements, where $\varphi = |\Phi|$ is the number of known positions of nonzero coefficients of $\hat{\mathbf{x}}$.

Note that the reason why (3) can be replaced by (6) when the prior information (5) is available comes from the fact that the minimization of the 1-norm in (3) is actually aimed at finding the sparsest solution that explains the measurements: the 1-norm is minimized in order to find the vector $\hat{\mathbf{x}}$ with most null components that satisfies $\mathbf{A}\hat{\mathbf{x}} = \mathbf{b}$. If the positions in Φ are known to contain nonzero components of $\hat{\mathbf{x}}$, then, during the search for a sparse solution to $\mathbf{A}\hat{\mathbf{x}} = \mathbf{b}$, the alternative is to minimize the number of nonzeros in the other positions only (those which do not belong to Φ). The improvement of (6) over (3) is then related to trying to minimize the number of nonzeros only outside the region where $\hat{\mathbf{x}}$ is already supposed to be nonzero; hence, (6) gives preference to a solution with more zeros outside the specified set Φ.

This paper shows that the prior information represented by (5) can also be added to the IRLS approach to the reconstruction. The corresponding minimization problem, with the prior information, is reformulated as

$$\min_{\hat{\mathbf{x}}} \frac{1}{2} \sum_{\substack{k=1 \\ k \notin \Phi}}^{N} |\hat{x}_k|^p, \quad \text{subject to } \mathbf{A}\hat{\mathbf{x}} = \mathbf{b}. \tag{7}$$

Our results show that the prior-information weighting scheme, which allows the reconstruction based on (7), leads to a reduction in the number of required measurements with respect to (4). This occurs for all used values of p, so a further reduction with respect to (3) and (4) can be attained by simultaneously using prior information and reducing p.

An important characteristic of (7) is that a solution $\hat{\mathbf{x}}$ is not explicitly constrained to be nonzero in the locations specified by Φ; rather, the corresponding values are determined from the equality constraint and the minimization of the objective function associated to the remaining positions. This is specially important in the cases in which the prior information is not perfectly reliable, so that some positions in Φ can actually not belong to the support (in this case, the reconstruction procedure should allow the computation of null elements inside Φ). By solving (7), we can still reconstruct the underlying signals, but more measurements may be required compared to the case when no wrong locations are present. In fact the $\hat{\mathbf{x}}$ components in the positions Φ are removed in (7) from the minimization function, so if some zero components' locations are mistakenly attributed to Φ the local sparsity of those components is not exploited during the reconstruction (we emphasize that the possibility of reconstructing signals from limited measurements in compressive sensing is based on exploiting the sparsity). Even in this case, however, we observed that if most of the components of Φ belong to the support, our proposed method provides an improvement in the reconstruction, in terms of computational cost and of number of required measurements. More details on this are presented in Section V.

III. Problem Formulation

The IRLS method for reconstructing sparse signals from linear measurements is based on iteratively solving (4) with a modified objective function, that at each iteration function approaches $\sum_{k=1}^{N} |\hat{x}_k|^p$. More specifically, consider the optimization problem

$$\min_{\hat{\mathbf{x}}} \frac{1}{2} \sum_{k=1}^{N} w_k^{p-2} \hat{x}_k^2, \quad \text{subject to } \mathbf{A}\hat{\mathbf{x}} = \mathbf{b} \tag{8}$$

where w_k is a weighting parameter [3]. Note that (8) can be solved in just one iteration, as we will describe, but if the problem is repeatedly solved changing the values of w_k at each time, so that w_k approaches $\hat{x}_k$, the objective function in (8) will approach that of (7). In fact, let

$$\mathbf{w}^{(m)} = \left|\hat{\mathbf{x}}^{(m-1)}\right|$$

where $\mathbf{w}^{(m)} = [\, w_1^{(m)} \quad w_2^{(m)} \quad \cdots \quad w_N^{(m)} \,]^T$ is the value of the weighting vector to be used in the mth iteration and $\hat{\mathbf{x}}^{(m-1)}$ is the $(m-1)$th iterate. After convergence, $\hat{\mathbf{x}}^{(m-1)}$ will be sufficiently close to $\hat{\mathbf{x}}^{(m)}$, according to a specified tolerance, so $\sum_{k=1}^{N} (w_k^{(m)})^{p-2} (\hat{x}_k^{(m)})^2 = \sum_{k=1}^{N} \left|\hat{x}_k^{(m-1)}\right|^{p-2} (\hat{x}_k^{(m)})^2$ will be close to $\sum_{k=1}^{N} \left|\hat{x}_k^{(m)}\right|^p$, which is the original objective function.

Now, in order to add the prior information on the positions of nonzero coefficients in the sparse domain, we must do the ℓ_p minimization over the vector components in the remaining positions only [1], [2]. Also, since the sparse signal must still match all the linear measurements, the equality constraint is the same, and the new minimization problem is

$$\min_{\hat{\mathbf{x}}} \frac{1}{2} \sum_{\substack{k=1 \\ k \notin \Phi}}^{N} |\hat{x}_k|^p, \quad \text{subject to } \mathbf{A}\hat{\mathbf{x}} = \mathbf{b}. \tag{9}$$

Now, a local solution to (9) can be obtained by iteratively solving

$$\min_{\hat{\mathbf{x}}} \frac{1}{2} \sum_{\substack{k=1 \\ k \notin \Phi}}^{N} w_k^{p-2} \hat{x}_k^2, \quad \text{subject to } \mathbf{A}\hat{\mathbf{x}} = \mathbf{b} \tag{10}$$

and changing w_k at each iteration so that $w_k^{p-2}\hat{x}_k^2$ is sufficiently close to $|\hat{x}_k|^p \ \forall k \notin \Phi$ after convergence.

By defining

$$w_k = 0 \, \forall k \in \Phi \tag{11}$$

the minimization in (10) is posed in the form of (8), which has a closed-form solution as indicated below. Since w_k must approach $\hat{x}_k$ for $k \notin \Phi$, we then define

$$w_k^{(m)} = \begin{cases} \left|\hat{x}_k^{(m-1)}\right|, & \text{if } k \notin \Phi \\ \tau \left|\hat{x}_k^{(m-1)}\right|, & \text{otherwise} \end{cases} \tag{12}$$

where τ is a specified small constant. Note that for $\tau = 0$, the second expression in (12) reduces to 0 as required by (11), but a small $\tau > 0$ is necessary for obtaining a closed solution based on the procedure described next.

A signal $\mathbf{x}$ with sparse representation $\hat{\mathbf{x}}$ can then be reconstructed from a sufficient amount of linear measurements $\mathbf{b}$ by solving for $m = 1, 2, \ldots$ and until convergence

$$\min_{\hat{\mathbf{x}}} \frac{1}{2} \sum_{k=1}^{N} \left|w_k{}^{(m)}\right|^{p-2} \hat{x}_k^2, \quad \text{subject to } \mathbf{A}\hat{\mathbf{x}} = \mathbf{b} \tag{13}$$

where $w_k^{(m)}$ is given by (12).

Our approach to (13) using (12) leads to the IRLS with prior information. Note that (13) corresponds to the minimization of a quadratic form with a linear equality constraint, which leads to the solution

$$\hat{\mathbf{x}}^{(m)} = \mathbf{Q}^{(m)}\mathbf{A}^T \left(\mathbf{A}\mathbf{Q}^{(m)}\mathbf{A}^T\right)^{-1} \mathbf{b} \tag{14}$$

where

$$\mathbf{Q} = \mathrm{diag}(q_1, q_2, \ldots, q_N) \tag{15}$$

with

$$q_k = \begin{cases} \left|\hat{x}_k^{(m-1)}\right|^{2-p}, & \text{if } k \notin \Phi \\ \tau^{2-p} \left|\hat{x}_k^{(m-1)}\right|^{2-p}, & \text{otherwise.} \end{cases} \tag{16}$$

In summary, the sparse representation $\hat{\mathbf{x}}$ can be reconstructed from sufficient linear measurements $\mathbf{b}$ by solving (14) for $m = 1, 2, \ldots$, until convergence.

IV. IRLS With Prior Information

We now summarize the basic steps needed to reconstruct sparse signals with the weighting strategy with prior information in the IRLS. Note that a regularization procedure is needed when defining the components of the main diagonal of $\mathbf{Q}^{(m)}$ according to (16). Because of the matrix inversion in (14), we must guarantee that these components do not approach zero, so a constant μ is added to $|\hat{x}_k|$ when defining the weights. With this regularization, (16) becomes

$$q_k = \begin{cases} \left|\hat{x}_k^{(m-1)}\right|^{2-p} + \mu, & \text{if } k \notin \Phi \\ \tau^{2-p} \left|\hat{x}_k^{(m-1)}\right|^{2-p} + \mu, & \text{otherwise} \end{cases} \tag{17}$$

where we used $\tau^{2-p} = 10^{-3}$. Our numerical experimentations show the proposed method to be stable regarding the choice of values for τ and the resulting factor τ^{2-p}. In fact, different orders of magnitude of τ were tested, with a large range of values allowing equivalent results. More details on the behavior of the algorithm with respect to this parameter are presented in Section V.

Then, the values of $\hat{\mathbf{x}}$ are updated according to (14), starting from an initial value $\hat{\mathbf{x}}^{(0)}$ and until the relative change between the norms of two consecutive iterates is below a specified tolerance. After this, the regularization parameter is then reduced and the iterations in (14) are again conducted until convergence. The process is repeated until μ becomes sufficiently small. In our simulations, the initial value $\hat{\mathbf{x}}^{(0)}$ is computed by finding the minimum 2-norm solution to the equality constraint (least squares solution), so

$$\hat{\mathbf{x}}^{(0)} = \mathbf{Q}^{(0)}\mathbf{A}^T(\mathbf{A}\mathbf{Q}^{(0)}\mathbf{A}^T)^{-1}\mathbf{b} \tag{18}$$

with the initial inverse weight matrix given by $\mathbf{Q}^{(0)} = \mathrm{diag}(q_1^{(0)}, q_2^{(0)}, \ldots, q_N^{(0)})$, where

$$q_k^{(0)} = \begin{cases} 1, & \text{if } k \notin \Phi \\ \tau^{2-p}, & \text{otherwise.} \end{cases} \tag{19}$$

In defining the convergence criterion for each iteration stage, our results have shown that the iterative procedure with prior information given by (14) can follow the same strategy proposed in [3]. In this scheme, (14) is repeated, first with $\mu = 1$, until

$$\frac{\left\|\left|\hat{\mathbf{x}}^{(m)} - \hat{\mathbf{x}}^{(m-1)}\right\|\right|}{1 + ||\hat{\mathbf{x}}^{(m-1)}||} < \frac{\sqrt{\mu}}{100}. \tag{20}$$

After (20) is attained, μ is reduced by a factor of 10, and the iterative procedure is repeated until $\mu \le 10^{-8}$ [3].

Algorithm 1 summarizes the procedures for reconstructing a sparse transform $\hat{\mathbf{x}}$ from the linear measurements $\mathbf{b}$ with prior information on the support domain, by solving (7). Note that $\mathbf{A} = \mathbf{M}\mathbf{T}^{-1}$; also, the signal $\mathbf{x}$, from which the measurements are taken, can be reconstructed by taking the inverse transform $\mathbf{x} = \mathbf{T}^{-1}\hat{\mathbf{x}}$.

Algorithm 1. IRLS Method for Signal Reconstruction in Compressive Sensing With Prior Information

Inputs: $p > 0$, $\mathbf{A}$, $\mathbf{b}$, Φ, μ, τ.
Step 1. Initialize $\hat{\mathbf{x}}^{(0)}$ and $\mathbf{Q}^{(0)}$ using (18) and (19).
Step 2. Do the inner loop:
- **2.1** Initialize $m := 1$.
- **2.2** Update $\mathbf{Q}^{(m)}$ using (15) and (17).
- **2.3** Compute $\hat{\mathbf{x}}^{(m)}$ using (14).
- **2.4** If (20) is satisfied, go to Step 3; otherwise, let $m := m+1$ and go to Step 2.2.

Step 3. Update the regularization parameter, $\mu := \mu/10$.
Step 4. If $\mu < 10^{-8}$, finish; else, go to Step 2.

V. Simulation Results

Algorithm 1 was first evaluated in different conditions using 500 test signals with length 256 (larger signals are also considered later). We used a procedure to guarantee that each of these signals, although randomly generated, is η-sparse in some specified domain. First, for generality, we defined a random orthogonal transformation matrix $\mathbf{T}$

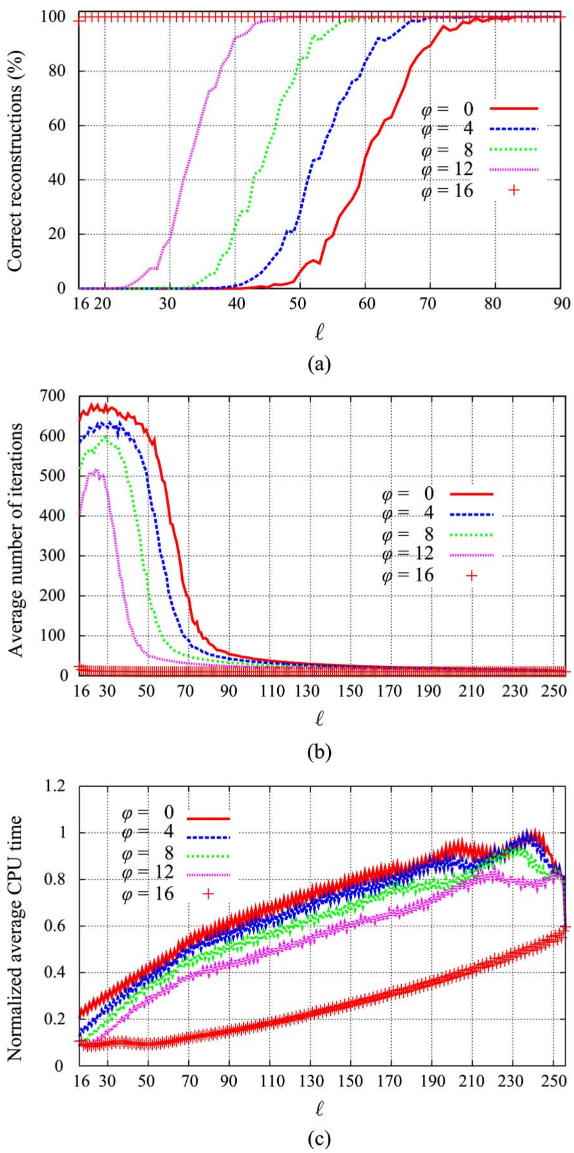

Fig. 1. Numerical results obtained using Algorithm 1 with 1-norm, as a function of the number of samples taken in the nonsparse domain and for different values of known positions (φ) in the support region of the sparse domain: (a) Percentage of correct reconstructions; (b) average number of iterations; (c) average time to convergence. All 500 test signals are of length $N = 256$ and sparsity $\eta = 16$ in an arbitrary, randomly determined transformed domain.

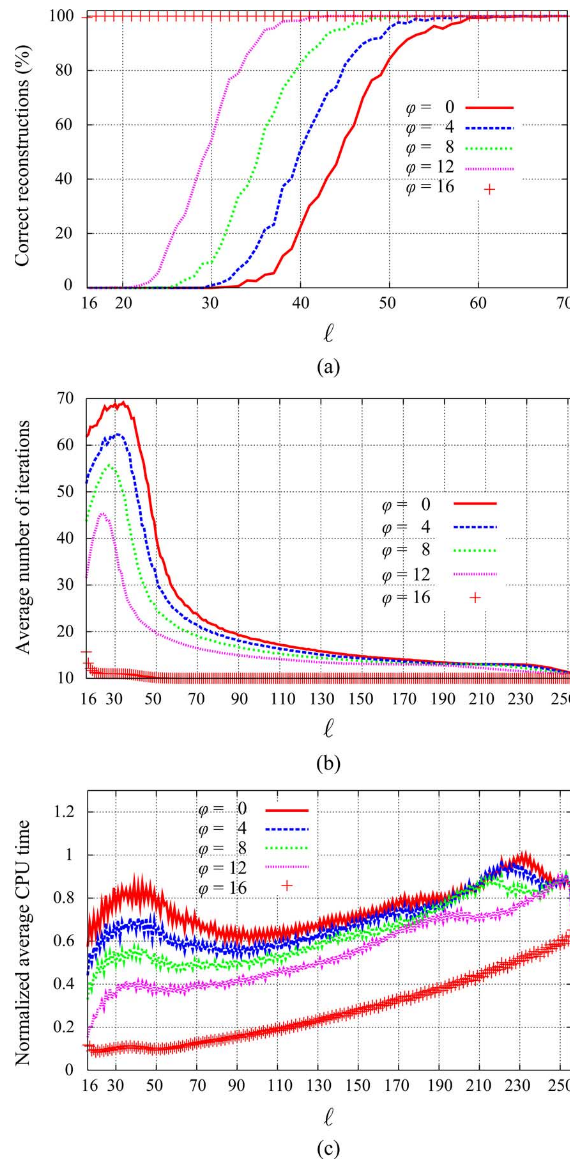

Fig. 2. Numerical results obtained using Algorithm 1 with $p = 0.1$, as a function of the number of samples taken in the nonsparse domain and for different values of known positions (φ) in the support region of the sparse domain: (a) Percentage of correct reconstructions; (b) average number of iterations; (c) average time to convergence. All 500 test signals are of length $N = 256$ and sparsity $\eta = 16$ in an arbitrary, randomly determined transformed domain.

using Givens decomposition [14]. Each signal was then generated first in the corresponding desired sparse domain; with this purpose, $\eta = 16$ nonzero values were determined using a Gaussian pseudo-random generator, while their positions were assigned by a generator with uniform distribution. From the sparse vector $\hat{\mathbf{x}}$ thus obtained, the time-domain signal was finally computed by taking the inverse transform $\mathbf{x} = \mathbf{T}^{-1}\hat{\mathbf{x}}$.

Each test consisted on taking an specific number of linear measurements, ℓ, from each of the 500 signals, and applying IRLS method with prior information to reconstruct it using different amounts of known positions φ belonging to the support of the sparse domain. For each possible combination of ℓ and φ, we then evaluated if the signal was correctly reconstructed. In this classification, we considered as correct reconstructions only the cases for which the normalized energy of the error between the original signal and the solution to (9) was below a prespecified tolerance of 10^{-3}.

In Figs. 1 and 2, we show the results using Algorithm 1 with $p = 1$ and $p = 0.1$, respectively. The percentages of correct reconstructions are shown in Figs. 1(a) and 2(a) as functions of ℓ and φ. While these percentages generally increase with the amount of available measurements, as expected, they also increase with φ. Indeed, as the number of known support positions increases by 4, the resulting curve for the percentage of correct reconstructions is shifted to the left, showing that for the same fixed percentage, the number of required measurements is reduced. This shows that the information represented by the φ positions is appropriately used by the IRLS method through the weighting scheme given by (12).

An important observation regarding the reductions in required measurements when using Algorithm 1 with prior information is that they occur for both values of p (0 and 0.1). This result is also verified for different values of p in the range $0 < p \leq 1$, as will be shown. In fact, in Figs. 1(a) and 2(a), the percentage curves are shifted by the same amount to the left for increasing values of φ. This indicates that, with respect to IRLS with no prior information, a more significant reduction in the amounts of measurements can be attained by reducing p and at the same time using prior information.

We emphasize that the evaluated amounts of correct reconstructions are based on the defined criterion of error (normalized energy of the difference below 10^{-3}), and that the amounts of measurements taken correspond to this criterion. A less strict criterion shifts the curves in Figs. 1(a) and 2(a) to the left, meaning less measurements being taken, but the distances between the curves corresponding to different values of φ are preserved. Furthermore, without prior information ($\varphi = 0$) the required number of measurements to attain reconstruction in our simulations matches the results we achieve using the algorithm in [3] for the same reconstruction error.

We have also observed that, by using prior information, Algorithm 1 allows a reduction in both the number of iterations and the total time to convergence. These aspects are illustrated, respectively, in parts (b) and (c) of Figs. 1 and 2. The number of iterations is counted as the number of times (14) is executed. Note that, as φ is increased, the curves corresponding to both the number of iterations and the convergence time are shifted to the bottom, meaning a consistent reduction in those quantities for all values of ℓ.

Note that the results in Figs. 1 and 2 follow from Algorithm 1 when the given prior information is correct for all given locations. This means that all the components of the given set Φ really belong to the support of the sparse domain, or at least to the set of points that are counted as possibly nonzeros when establishing the signals' sparsity. This could be the case, for instance, of signals that were bandpass filtered, so that the locations corresponding to the passband contain the potentially nonzero coefficients; in this condition, the set Φ corresponds to the passband.

As stated in Section II, however, it may be the case that the prior information is not perfectly reliable, meaning that some of the components of the considered set Φ belong to the support but others are misplaced, and thus actually associated to null components. In this situation, Algorithm 1 can still reconstruct the underlying signals (the computed components are not constrained to be zero anywhere—even inside the given set Φ), but more measurements may be required compared to the case when no wrong locations are present. In fact, according to (9), the elements of Φ are removed from the minimization function, so if some zero components are mistakenly attributed to Φ the local sparsity of those components is not exploited during the reconstruction. Indeed, in our experiments we observed an improvement in performance that depends on the difference between the number of correct positions and wrong positions in Φ.

In Fig. 3, we exemplify cases in which Φ contains both correct prior information (c locations belonging to the support) and incorrect prior information (w locations that do not belong to the support). Algorithm 1 was applied to 500 signals with length $N = 256$ and sparsity $\eta = 16$, and for different combinations of c and $w < c$. Note that the percentages of correct reconstructions in the cases $(c = 14, w = 2)$, $(c = 11, w = 3)$, $(c = 8, w = 4)$ are, respectively, greater than in the cases $(c = 12, w = 0)$, $(c = 8, w = 0)$, $(c = 4, w = 0)$, suggesting that it is better to have c correct positions and w wrong positions than to have $c - w$ correct positions only. Similarly, in all these conditions the results indicated an improvement with respect to the case with no prior information at all $(c = 0, w = 0)$, provided that most of the elements in Φ really belong to the support $(c > w)$.

Regarding the reductions in the number of required measurements, the amounts of iterations, and the computation time when using Algorithm 1 with prior information, we observed that they occur for all tested values of p in the objective function (30 values of p in the range $0 < p \leq 1$ were analyzed). In Fig. 4, we describe the results for all the tested values of p in the range $0 < p \leq 1$ and for a fixed number of linear measurements, $\ell = 2.5\eta = 160$; these results correspond to applying Algorithm 1 to 500 different signals with length $N = 1024$

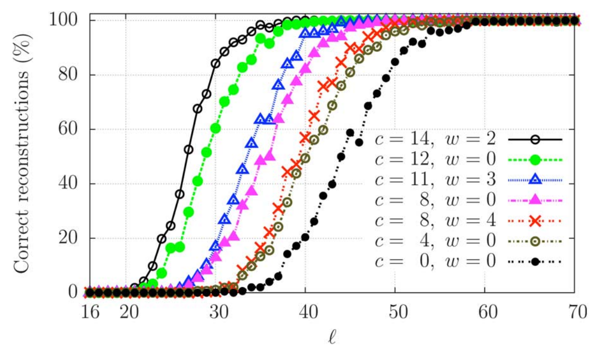

Fig. 3. Percentages of sparse signals that are correctly reconstructed using Algorithm 1 with partially correct and partially incorrect prior information on the support of the sparse domain. All 500 test signals are of length $\boldsymbol{N = 256}$ and sparsity $\boldsymbol{\eta = 16}$, and $\boldsymbol{\ell}$ represents the number of linear measurements available for reconstruction. In each line, $\boldsymbol{c}$ is the number of components of the given set $\boldsymbol{\Phi}$ that really belong to the support (correct prior information), whereas $\boldsymbol{w}$ is the number of components of $\boldsymbol{\Phi}$ that do not belong to the support (wrong prior information).

and sparsity $\eta = 64$. In Fig. 4(a), we observe that, independently on the used value of p, a higher value of known positions φ leads to an increase in the percentage of correct reconstructions from the same amount of measurements ℓ, even when a lower value of p already provides an increase with respect to $p = 1$ (the advantage of using the prior information does not vanish when p is decreased to improve reconstruction). Also, Fig. 4(b) and (c) shows that as φ increases, both the number of iterations and the time to convergence decrease, for all tested values of p. Note in Fig. 4(c) that decreasing p has the effect of increasing the time required to reconstruct the signals from the same amount of measurements, but the used prior information reduces both the number of iterations and the convergence time.

Finally, it is important to evaluate the sensitivity of the proposed method with respect to the parameter τ^{p-2}, used in (17). We conducted a complete set of experiments in which we applied Algorithm 1 with τ^{2-p} ranging from 10^{-12} to 10^{2} (although τ should be less than 1, according to Section III, we also included the range $1 \leq \tau^{2-p} \leq 10^{2}$ for illustration). These experiments show that a large range of values can be used with equivalent results.

As an example, Fig. 5 shows the percentage of correct reconstructions, the total number of required iterations, and the normalized computation time, as functions of the tested parameter when applying Algorithm 1 with $p = 0.01$; the remaining parameters were kept as in Fig. 4. As expected, τ^{2-p} greater than 1 does not allow Algorithm 1 to reconstruct the signals, as shown in the right side of Fig. 5(a). On the other hand, too low values of the same parameter also lead to a reduction in the percentage of correct reconstructions, due to the reduced stability of the resulting linear system in (14). Note, however, that a large range of values of τ^{2-p}, with orders of magnitude between approximately 10^{-6} and 10^{-2}, allow the reconstruction of the same percentage of signals with equivalent computation times and numbers of iterations.

VI. Conclusion

This paper proposes a signal reconstruction scheme based on the iteratively reweighted least squares (IRLS) method for compressive sensing with prior information. The proposed method, related to the definition of the weights matrix used at each iteration, allows the efficient use of information on the support of the sparse domain of the underlying signal.

The simulation results show that by using prior information, a reduction occurs in the number of linear measurements required to attain a prespecified percentage of correct reconstructions. This reduction is directly related to the number of known positions, so that if φ positions

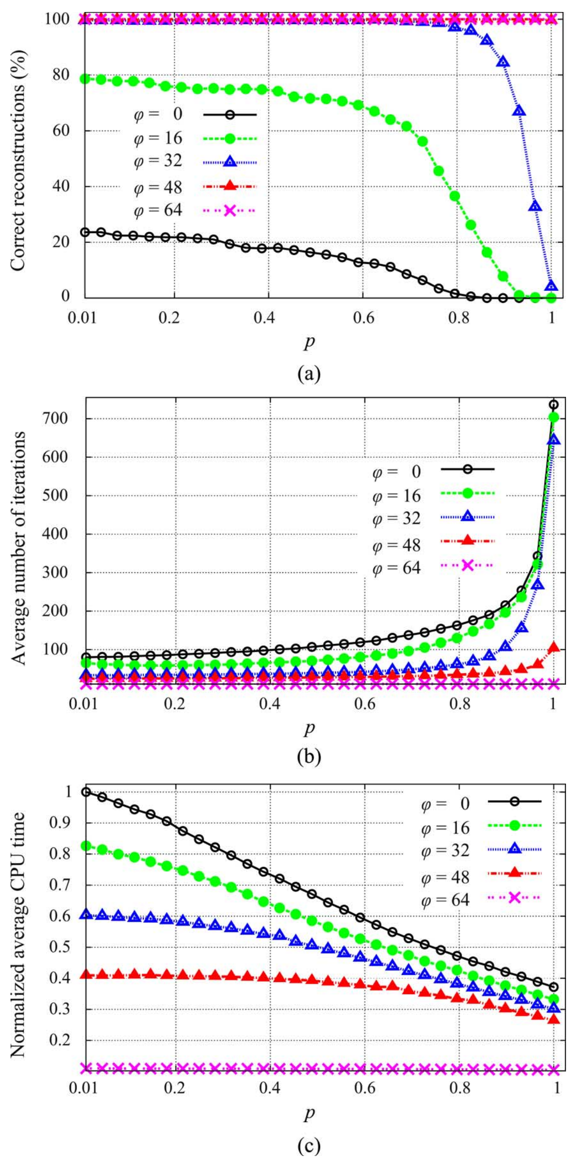

Fig. 4. Numerical results obtained using Algorithm 1, as a function of $\boldsymbol{p}$ and for different values of known positions $(\boldsymbol{\varphi})$ in the support region of the sparse domain: (a) Percentage of correct reconstructions; (b) Average number of iterations; (c) Average time to convergence. All 500 test signals are of length $\boldsymbol{N = 1024}$ and sparsity $\boldsymbol{\eta = 64}$ in an arbitrary, randomly determined transformed domain. In all cases, reconstruction is based on $\boldsymbol{\ell = 2.5\eta = 160}$ measurements.

are known, the number of required measurements is also reduced by φ with respect to the IRLS scheme.

A reduction in the number of iterations and computation time required for convergence is also verified when prior information is added to the reconstruction procedure. This result was consistent, independently on the number of linear measurements used for reconstruction.

An important observation regarding the IRLS with prior information is that the reduction in the magnitudes of both the number of required measurements and the computational cost when using prior information occurs for all tested values of p in the ℓ_p minimization. Hence, a further reduction can be attained by reducing p while at the same time using prior information. Also, our experimentations show that if the prior information is not perfectly reliable, meaning that some components of the support are misplaced, there is still an improvement in performance with respect to no prior information or to less correct positions being available. The final performance depends on the difference between the numbers or correct and wrong locations.

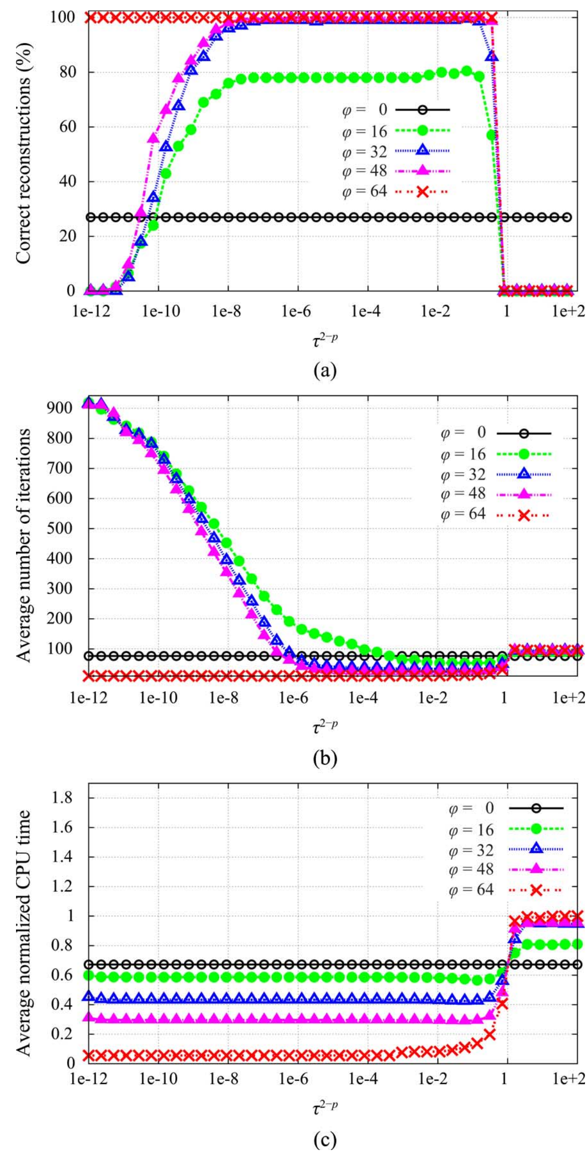

Fig. 5. Numerical results obtained using Algorithm 1, as a function of the used parameter $\boldsymbol{\tau}^{2-p}$ and for different values of known positions $(\boldsymbol{\varphi})$ in the support region of the sparse domain: (a) Percentage of correct reconstructions; (b) Average number of iterations; (c) Average time to convergence. All 500 test signals are of length $\boldsymbol{N = 1024}$ and sparsity $\boldsymbol{\eta = 64}$ in an arbitrary, randomly determined transformed domain. In all cases, reconstruction is based on $\boldsymbol{\ell = 2.5\eta = 160}$ measurements.

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their comments and suggestions, which represented an important contribution to this paper. In addition, the authors thank Dr. D. Williams and N. Gumataotao for their assistance with the NSF-supported cluster (NSF CNS-0709438), used in all the simulations here described.

REFERENCES

[1] R. von Borries, C. J. Miosso, and C. M. Potes, "Compressed sensing using prior information," in *Proc. 2nd IEEE Int. Workshop Computational Advances Multi-Sensor Adaptive Processing (CAMPSAP)*, Dec. 12–14, 2007, pp. 121–124.

[2] R. von Borries and C. J. Miosso, "Compressed sensing with prior information on the support region of the sparse domain," 2008, in preparation.

[3] R. Chartrand and W. Yin, "Iteratively reweighted algorithms for compressive sensing," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, Las Vegas, NV, Apr. 4, 2008, pp. 3869–3872.

[4] R. G. Baraniuk, "Compressive sensing lecture notes," *IEEE Signal Process. Mag.*, vol. 24, no. 4, pp. 118–121, 2007.

[5] E. J. Candès and M. B. Wakin, "An introduction to compressive sampling a sensing/sampling paradigm that goes against the common knowledge in data acquisition," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 21–30, 2008.

[6] M. F. Duarte, M. B. Wakin, and R. G. Baraniuk, "Fast reconstruction of piecewise smooth signals from incoherent projections," presented at the Structure et Parcimonie Pour la Représentation Adaptative de Signaux (SPARS), Rennes, France, Nov. 16–18, 2005.

[7] C. La and M. N. Do, "Signal reconstruction using sparse tree representations," presented at the SPIE Conf. Wavelet Applications Signal Image Processing XI, San Diego, CA, Jul. 2005.

[8] J. Garcia-Frias and I. Esnaola, "Exploiting prior knowledge in the recovery of signals from noisy random projections," in *Data Compression Conf. (DCC)*, 2007, pp. 333–342.

[9] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Trans. Image Process.*, vol. 15, pp. 3736–3745, 2006.

[10] M. Elad and M. Aharon, "Image denoising via learned dictionaries and sparse representation," in *IEEE Computer Soc. Conf. Computer Vision Pattern Recognition*, 2006, vol. 1, pp. 895–900.

[11] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.

[12] E. J. Candès and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?," *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5406–5425, Dec. 2006.

[13] B. D. Rao and K. Kreutz-Delgado, "An affine scaling methodology for best basis selection," *IEEE Trans. Signal Process.*, vol. 47, no. 1, pp. 187–200, Jan. 1999.

[14] C. D. Meyer, *Matrix Analysis and Applied Linear Algebra*. Philadelphia, PA: SIAM, 2000.