

An Efficient Algorithm for Sparse Representations with ℓ^p Data Fidelity Term

Paul Rodríguez and Brendt Wohlberg

Abstract—Basis Pursuit (BP) and Basis Pursuit Denoising (BPDN), well established techniques for computing sparse representations, minimize an ℓ^2 data fidelity term, subject to an ℓ^1 sparsity constraint or regularization term, by mapping the problem to a linear or quadratic program. BPDN with an ℓ^1 data fidelity term has recently been proposed, also implemented via a mapping to a linear program. We introduce an alternative approach via an Iteratively Reweighted Least Squares algorithm, providing computational advantages and greater flexibility in the choice of data fidelity term norm.

Index Terms—Image restoration, inverse problem, regularization, total variation.

I. INTRODUCTION

A sparse representation is an adaptive signal decomposition consisting of a linear combination of *atoms* from an overcomplete *dictionary*, where the coefficients of the linear combination are optimized according to some sparsity criterion. Applications of these representations include EEG (electroencephalography) and MEG (magnetoencephalography) estimation [1], time-frequency analysis [2], spectrum estimation [3], denoising [4], image coding [5], and cartoon/texture decomposition of images [6].

One of the most well-known methods for computing such a sparse representation is Basis Pursuit Denoising (BPDN) [4], which consists of the minimization

$$\min_{\mathbf{u}} \frac{1}{2} \|\Phi \mathbf{u} - \mathbf{b}\|_2^2 + \lambda \|\mathbf{u}\|_1, \quad (1)$$

where $\|\Phi \mathbf{u} - \mathbf{b}\|_2^2$ and $\|\mathbf{u}\|_1$ are known as the fidelity term and the sparsity term respectively, \mathbf{b} is the signal to be decomposed, Φ is the (overcomplete) dictionary matrix, λ is a weighting factor controlling the relative importance of the data fidelity and sparsity terms, and \mathbf{u} is the sparse representation. This optimization problem is mapped to a quadratic program, which is solved via interior point methods. An alternative approach [1], [7] is to solve

$$\min_{\mathbf{u}} \frac{1}{2} \|\Phi \mathbf{u} - \mathbf{b}\|_2^2 + \frac{\lambda}{q} \|\mathbf{u}\|_q^q, \quad (2)$$

Paul Rodríguez is with Digital Signal Processing Group at the Pontificia Universidad Católica del Perú, Lima, Peru. Email: prodrig@pucp.edu.pe, Tel: +51 19 9339 5427

Brendt Wohlberg is with T-7 Mathematical Modeling and Analysis, Los Alamos National Laboratory, Los Alamos, NM 87545, USA. Email: brendt@t7.lanl.gov, Tel: +1 505 667 6886, Fax: +1 505 665 5757

This work was carried out under the auspices of the National Nuclear Security Administration of the U.S. Department of Energy at Los Alamos National Laboratory under Contract No. DE-AC52-06NA25396 and was partially supported by the NNSA's Laboratory Directed Research and Development Program.

where $q \leq 1$, via a form of the Iteratively Reweighted Least Squares (IRLS) [8] method.

Total Variation (TV) regularization methods [9] of denoising and image restoration are closely related to BPDN (and directly equivalent for 1-d signals). Recently, there has been significant interest in TV functionals with an ℓ^1 data fidelity term [10], [11], with advantages including superior denoising performance with speckle noise. Granai and Vandergheynst [12] have observed that these advantages are also applicable to sparse representations, and proposed a variant of BPDN with ℓ^1 data fidelity term

$$\min_{\mathbf{u}} \|\Phi \mathbf{u} - \mathbf{b}\|_1 + \lambda \|\mathbf{u}\|_1, \quad (3)$$

solved by mapping it to a linear program (as proposed by Fu *et. al.* [11]). While elegant, this approach is computationally expensive, since a BPDN problem in M unknowns, with Φ , an $N \times M$ dictionary matrix (generally overcomplete, i.e.: $M > N$), is mapped to a linear program in $2(N + M)$ unknowns.

Here, we propose a more computationally efficient algorithm, motivated by our Iteratively Reweighted Norm (IRN) [13], [14] approach for ℓ^1 -TV (and which may also be considered a generalization of the AST/FOCUSS algorithms for BPDN [1], [7]), capable of solving the more general form of BPDN

$$\min_{\mathbf{u}} \frac{1}{p} \|\Phi \mathbf{u} - \mathbf{b}\|_p^p + \frac{\lambda}{q} \|\mathbf{u}\|_q^q, \quad (4)$$

which includes both the standard BPDN (see (1) and (2)) and ℓ^1 data fidelity term BPDN (see (3)) as special cases.

II. IRN-BPDN ALGORITHM

A. Previous Related Work

The IRN approach is closely related to the Iteratively Reweighted Least Squares (IRLS) method [15], [8], [16], [17], [18]. Similar ideas have also been applied [19], [7] to solving the standard BP and BPDN problems [4] for sparse representations.

IRLS minimizes the ℓ^p norm

$$F(\mathbf{u}) = \frac{1}{p} \left\| \Phi \mathbf{u} - \mathbf{b} \right\|_p^p \quad (5)$$

for $p \leq 2$ by approximating it, within an iterative scheme, by a weighted ℓ^2 norm. At iteration k the solution $\mathbf{u}^{(k)}$ is the minimizer of $\frac{1}{2} \|W^{(k)1/2} (\Phi \mathbf{u} - \mathbf{b})\|_2^2$, with weighting matrix $W^{(k)} = \text{diag} (|\Phi \mathbf{u}^{(k)} - \mathbf{b}|^{p-2})$, and the iteration

$$\mathbf{u}^{(k+1)} = \left(\Phi^T W^{(k)} \Phi \right)^{-1} \Phi^T W^{(k)} \mathbf{b},$$

which minimizes the weighted version of (5) using the weights derived from the previous iteration, converges to the minimizer of $F(\mathbf{u})$ [18].

When $p < 2$, the definition of the weighting matrix $W^{(k)}$ must be modified to avoid the possibility of division by zero. For $p = 1$, it may be shown [17] that the choice

$$W_{n,n}^{(k)} = \begin{cases} |r_n^{(k)}|^{-1} & \text{if } |r_n^{(k)}| \geq \epsilon \\ \epsilon^{-1} & \text{if } |r_n^{(k)}| < \epsilon \end{cases},$$

where $\mathbf{r}^{(k)} = \Phi\mathbf{u}^{(k)} - \mathbf{b}$, and ϵ is a small positive number, guarantees global convergence to the minimizer of $\sum_n \rho_\epsilon(r_n)$, where

$$\rho_\epsilon(r_n) = \begin{cases} \epsilon^{-1} r_n^2 & \text{if } |r_n| \leq \epsilon \\ 2|r_n| - \epsilon & \text{if } |r_n| > \epsilon \end{cases}$$

is the Huber function [20].

B. Fidelity Term

The data fidelity term of the generalized BPDN functional (4) is the same as the term that the IRLS functional (5) seeks to minimize. In order to replace the ℓ^p norm by a ℓ^2 norm, we define the quadratic functional

$$Q_F^{(k)}(\mathbf{u}) = \frac{1}{2} \left\| W_F^{(k)1/2} (\Phi\mathbf{u} - \mathbf{b}) \right\|_2^2 + \left(1 - \frac{p}{2}\right) F(\mathbf{u}^{(k)}), \quad (6)$$

where $\mathbf{u}^{(k)}$ is a constant representing the solution of the previous iteration, $F(\cdot)$ is defined in (5), and

$$W_F^{(k)} = \text{diag} \left(\tau_{F,\epsilon_F} (A\mathbf{u}^{(k)} - \mathbf{b}) \right). \quad (7)$$

Following a common strategy in IRLS type algorithms [8], the function

$$\tau_{F,\epsilon_F}(x) = \begin{cases} |x|^{p-2} & \text{if } |x| > \epsilon_F \\ \epsilon_F^{p-2} & \text{if } |x| \leq \epsilon_F \end{cases}, \quad (8)$$

is defined (for some small ϵ_F) to avoid numerical problems when $p < 2$ and $A\mathbf{u}^{(k)} - \mathbf{b}$ has zero-valued components.

The constant (with respect to \mathbf{u}) term $(1 - \frac{p}{2}) F(\mathbf{u}^{(k)})$ is added in (6) so that, neglecting numerical precision issues in (6) and (7),

$$F(\mathbf{u}^{(k)}) = Q_F^{(k)}(\mathbf{u}^{(k)}) \quad (9)$$

as $\epsilon_F \rightarrow 0$. In other words, the weighted ℓ^2 norm tends to the original ℓ^p norm fidelity term at $\mathbf{u} = \mathbf{u}^{(k)}$. The bound (see the appendix of [18])

$$F(\mathbf{u}) < Q_F^{(k)}(\mathbf{u}) \quad \forall \mathbf{u} \neq \mathbf{u}^{(k)} \quad p \leq 2, \quad (10)$$

and the Fréchet derivatives for $F(\mathbf{u})$ and $Q_F^{(k)}(\mathbf{u})$

$$\begin{aligned} \nabla_{\mathbf{u}} F(\mathbf{u}) &= \Phi^T (\Phi\mathbf{u} - \mathbf{b})^{p-1} \\ \nabla_{\mathbf{u}} Q_F^{(k)}(\mathbf{u}) &= \Phi^T W_F^{(k)} (\Phi\mathbf{u} - \mathbf{b}). \end{aligned}$$

play an important role in the convergence proof in Section II-E. Observe also that

$$\nabla_{\mathbf{u}} F(\mathbf{u})|_{\mathbf{u}=\mathbf{u}^{(k)}} = \nabla_{\mathbf{u}} Q_F^{(k)}(\mathbf{u})|_{\mathbf{u}=\mathbf{u}^{(k)}} \quad (11)$$

when $\epsilon_F \rightarrow 0$, and note that the original fidelity term in (5) and its quadratic version in (6) have the same value and tangent direction at $\mathbf{u} = \mathbf{u}^{(k)}$.

C. Sparsity Term

The sparsity term in (4)

$$S(\mathbf{u}) = \frac{1}{q} \|\mathbf{u}\|_q^q \quad (12)$$

is handled similarly. We define the quadratic functional

$$Q_S^{(k)}(\mathbf{u}) = \frac{1}{2} \left\| W_S^{(k)1/2} \mathbf{u} \right\|_2^2 + \left(1 - \frac{q}{2}\right) S(\mathbf{u}^{(k)}), \quad (13)$$

where $\mathbf{u}^{(k)}$ is a constant representing the solution of the previous iteration, and

$$W_S^{(k)} = \text{diag} \left(\tau_{S,\epsilon_S}(\mathbf{u}^{(k)}) \right). \quad (14)$$

Following the strategy described in [7] τ_{S,ϵ_S} is defined (for some small ϵ_S) as

$$\tau_{S,\epsilon_S}(x) = \begin{cases} |x|^{q-2} & \text{if } |x| > \epsilon_S \\ 0 & \text{if } |x| \leq \epsilon_S \end{cases}, \quad (15)$$

where the choice $\tau_{S,\epsilon_S}(x) = 0$ for $|x| \leq \epsilon_S$ will be further discussed in Section II-D.

Note that the constant (with respect to \mathbf{u}) term $(1 - \frac{q}{2}) S(\mathbf{u}^{(k)})$ is added in (13) to ensure that as $\epsilon_S \rightarrow 0$

$$S(\mathbf{u}^{(k)}) = Q_S^{(k)}(\mathbf{u}^{(k)}), \quad (16)$$

and the bound

$$S(\mathbf{u}) < Q_S^{(k)}(\mathbf{u}) \quad \forall \mathbf{u} \neq \mathbf{u}^{(k)} \quad q \leq 2; \quad (17)$$

is easily proven, following a similar approach as described in the appendix of [18]). It is straightforward to compute the the Fréchet derivatives for $S(\mathbf{u})$ and $Q_S^{(k)}(\mathbf{u})$

$$\begin{aligned} \nabla_{\mathbf{u}} S(\mathbf{u}) &= \mathbf{u}^{q-1} \\ \nabla_{\mathbf{u}} Q_S^{(k)}(\mathbf{u}) &= W_S^{(k)} \mathbf{u}, \end{aligned}$$

and note that

$$\nabla_{\mathbf{u}} S(\mathbf{u})|_{\mathbf{u}=\mathbf{u}^{(k)}} = \nabla_{\mathbf{u}} Q_S^{(k)}(\mathbf{u})|_{\mathbf{u}=\mathbf{u}^{(k)}} \quad (18)$$

when $\epsilon_S \rightarrow 0$. As for the fidelity term, it is important to note that the original sparsity term (12) and its quadratic version (13) have the same value and tangent direction at $\mathbf{u} = \mathbf{u}^{(k)}$.

D. Algorithm Derivation

For improved readability, this derivation focuses on the ℓ^1 BPDN case, but the general case of ℓ^p BPDN is a trivial extension. Combining the terms described in Sections II-B and II-C gives the functional (compare it to (3))

$$T^{(k)}(\mathbf{u}) = \frac{1}{2} \left\| W_F^{(k)1/2} (\Phi\mathbf{u} - \mathbf{b}) \right\|_2^2 + \frac{\lambda}{2} \left\| W_S^{(k)1/2} \mathbf{u} \right\|_2^2 + C(\mathbf{u}^{(k)}), \quad (19)$$

where $C(\mathbf{u}^{(k)})$ combines the constants, with respect to \mathbf{u} , in (6) and (13).

Initialize

$$\mathbf{u}^{(0)} = \Phi^T (\Phi\Phi^T + \lambda I)^{-1} \mathbf{b}$$

Iterate $k = 0, 1, \dots$

$$W_F^{(k)} = \left[\text{diag} \left(\tau_{F, \epsilon_F} (\Phi \mathbf{u}^{(k)} - \mathbf{b}) \right) \right]^{-1}$$

$$W_R^{(k)} = \left[\text{diag} \left(\tau_{S, \epsilon_S} (\mathbf{u}^{(k)}) \right) \right]^{-1}$$

$$\boldsymbol{\chi}^{(k)} = (\Phi W_S^{(k)} \Phi^T + \lambda W_F^{(k)})^{-1} \mathbf{b}$$

$$\mathbf{u}^{(k+1)} = W_R^{(k)} \Phi^T \boldsymbol{\chi}^{(k)}$$

Algorithm 1: IRN-BPDN algorithm

The first step is to move the weighting (diagonal) matrix $W_S^{(k)}$ from the sparsity term into the fidelity term; this can be accomplished by setting $\mathbf{u} = W_S^{(k)1/2} \boldsymbol{\nu}$, giving (we neglect the constant term $C(\mathbf{u}^{(k)})$)

$$T^{(k)}(\boldsymbol{\nu}) = \frac{1}{2} \left\| W_F^{(k)1/2} \Phi W_S^{(k)-1/2} \boldsymbol{\nu} - W_F^{(k)1/2} \mathbf{b} \right\|_2^2 + \frac{\lambda}{2} \|\boldsymbol{\nu}\|_2^2. \quad (20)$$

It is important to note that the expressions involving $W_S^{(k)}$ raised to a negative power (in particular $W_S^{(k)-1/2}$ or $W_S^{(k)-1}$) do not generate a division by zero (see (14), (15)) since the ℓ^q -norm in the sparsity term is restricted to cases with $q \leq 2$. Computing the gradient (Fréchet derivative) of (20) and setting it to zero gives

$$W_S^{(k)-1/2} \Phi^T W_F^{(k)} \Phi W_S^{(k)-1/2} \boldsymbol{\nu} - W_S^{(k)-1/2} \Phi^T W_F^{(k)} \mathbf{b} + \lambda \boldsymbol{\nu} = 0.$$

Now, setting $\boldsymbol{\nu} = W_S^{(k)-1/2} \Phi^T \boldsymbol{\chi}$, and factoring out $W_S^{(k)-1/2} \Phi^T W_S^{(k)}$ gives

$$W_S^{(k)-1/2} \Phi^T W_F^{(k)} \left(\Phi W_S^{(k)-1} \Phi^T \boldsymbol{\chi} - \mathbf{b} + \lambda W_F^{(k)-1} \boldsymbol{\chi} \right) = 0.$$

Finally, we find the minimum to (19) by solving

$$\boldsymbol{\chi} = \left(\Phi W_S^{(k)-1} \Phi^T + \lambda W_F^{(k)-1} \right)^{-1} \mathbf{b} \quad (21)$$

and then substituting for $\boldsymbol{\nu}$ and \mathbf{u} .

It is interesting to note that (19) may be rewritten as

$$T^{(k)}(\mathbf{u}) = \frac{1}{2} \left\| W^{(k)1/2} (\tilde{\Phi} \mathbf{u} - \tilde{\mathbf{b}}) \right\|_2^2 + C(\mathbf{u}^{(k)}), \quad (22)$$

where

$$W^{(k)} = \begin{pmatrix} W_F^{(k)} & 0 \\ 0 & W_S^{(k)} \end{pmatrix}, \tilde{\Phi} = \begin{pmatrix} \Phi \\ \sqrt{\lambda} I \end{pmatrix}, \text{ and } \tilde{\mathbf{b}} = \begin{pmatrix} \mathbf{b} \\ 0 \end{pmatrix},$$

which has the same form as a standard IRLS problem, but differs in the computation of the weighting matrix.

The IRN-BPDN algorithm is summarized in Algorithm 1. The initial solution is the minimum ℓ^2 norm solution obtained by setting the weighting matrices to identity matrices.

E. Convergence of the IRN-BPDN algorithm

Here we briefly sketch the proof of global convergence of the IRN-BPDN algorithm. We first note that from (9) and (16) it is easy to check that $T(\mathbf{u}^{(k)}) = T^{(k)}(\mathbf{u}^{(k)})$, where $T(\mathbf{u}) = \frac{1}{p} \|\Phi \mathbf{u} - \mathbf{b}\|_p^p + \frac{\lambda}{q} \|\mathbf{u}\|_q^q$ (see (4)), $T^{(k)}(\mathbf{u})$ is defined in (19) and $\mathbf{u}^{(k)}$ is the vector used to compute the weights $W_F^{(k)}$ and $W_S^{(k)}$. Moreover, from (10) and (17) we have that

$$T(\mathbf{u}) \leq T^{(k)}(\mathbf{u}) \quad \forall \mathbf{u} \quad p, q \leq 2$$

with equality only for $\mathbf{u} = \mathbf{u}^{(k)}$. It is also easy to check (see (11) and (18)) that

$$\nabla_{\mathbf{u}} T(\mathbf{u})|_{\mathbf{u}=\mathbf{u}^{(k)}} = \nabla_{\mathbf{u}} T^{(k)}(\mathbf{u})|_{\mathbf{u}=\mathbf{u}^{(k)}}. \quad (23)$$

Furthermore the Hessian of $T^{(k)}(\mathbf{u})$

$$\nabla_{\mathbf{u}}^2 T^{(k)}(\mathbf{u}) = \Phi^T W_F^{(k)} \Phi + \lambda W_S^{(k)}$$

is a positive defined matrix (i.e.: $\nabla_{\mathbf{u}}^2 T^{(k)}(\mathbf{u}) > 0$) as well as the linear system defined in (21): $\Phi W_S^{(k)-1} \Phi^T + \lambda W_F^{(k)-1} > 0$.

The quadratic functional $T^{(k)}(\mathbf{u})$ is tangent to $T(\mathbf{u})$ at $\mathbf{u} = \mathbf{u}^{(k)}$, where it is also an upper bound for $T(\mathbf{u})$, and it has a positive definite Hessian. Using these results (see [18]), it can be shown that the minimizer of $T^{(k)}(\mathbf{u})$, given by (21) with $\boldsymbol{\nu} = W_S^{(k)-1/2} \Phi^T \boldsymbol{\chi}$ and $\mathbf{u} = W_S^{(k)1/2} \boldsymbol{\nu}$ converges to the minimizer of (4) as we iterate over k .

III. COMPUTATIONAL RESULTS

From a computational point of view, the main advantage of the IRN-BPDN algorithm over a mapping into a linear program of the original problem (see (3)) is the size of the linear system to be solved: if the original BPDN problem has M unknowns, with a $N \times M$ dictionary matrix, the linear system to be solved (in the case of IRN-BPDN, the linear system is described in (21)) will have M (not N) unknowns, whereas, if (3) is mapped to a linear program, the linear system to be solved will have $2(N + M)$ unknowns.

Here we provide empirical evidence of the superior computational performance of the proposed algorithm, when compared to the linear programming approach [12]. The IRN-BPDN algorithm was implemented in Matlab (code available in [21]), as was the linear programming method, which utilized the SparseLab [22] Matlab toolbox, with some minor modifications to handle 2-dimensional datasets. The comparisons were run on a 3GHz Pentium4 machine.

We chose a cubic phase cosine image with sizes 16×16 , 32×32 , 64×64 and 128×128 (see Figure 1 for the in 128×128 case) and we add 5% speckle noise in each case (see Figure 2 for the in 128×128 case) and performed ℓ^1 -BPDN using a DCT dictionary with an overcompleteness factor of 4. Both procedures (IRN-BPDN and mapping into a linear program) give similar results from a quality (SNR) point of view; in general the solution given by the linear program have a better SNR for small problems than the solution found by the IRN-BPDN algorithm. The gap (in the SNR) between the

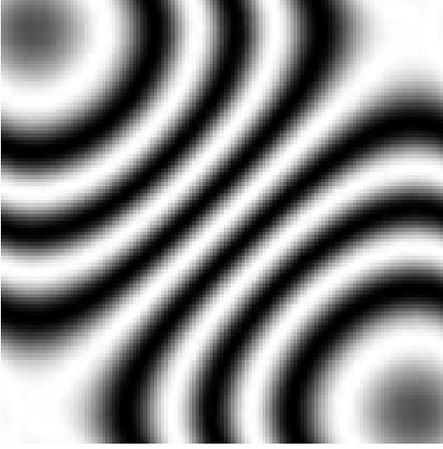


Fig. 1. 128×128 Cubic phase image.



Fig. 2. Cubic image with 5% speckle noise. SNR: 9.91dB.

solutions provided by both algorithms decreases as the size of the problem increases; this has been empirically confirmed by simulations run over 1-dimensional datasets and for 2-dimensional datasets we expect a similar behavior.

The time-performance of the IRN-BPDN is far superior to the procedure described in [12]: for a 16×16 image, IRN-BPDN requires 2.02 seconds to solve the ℓ^1 -BPDN problem, whereas the procedure described in [12] requires 28.09 seconds; for a 32×32 image IRN-BPDN requires 8.13 seconds while the procedure described in [12] requires 303.39 seconds. Note the difference in the scaling factor is due to the size of the linear system to be solved in each case. For input sizes of 64×64 and 128×128 IRN-BPDN takes 36.31 and 199.97 seconds respectively, while the implementation of the procedure described in [12] written by the authors was unable to finish: for a 64×64 input image and using a DCT dictionary with an overcompleteness factor of 4, [12] will generate $2 \cdot (64 \cdot 64 + 16 \cdot 64 \cdot 64) = 139264$ unknowns (557056 unknowns for the 128×128 case). This results are summarized in Table I.

TABLE I

TIME-PERFORMANCE COMPARISON BETWEEN IRN-BPDN AND ℓ^1 -BPDN IMPLEMENTED VIA MAPPING TO A LINEAR PROGRAM [12].

Image size	16×16	32×32	64×64	128×128
IRN-BPDN	2.0s	8.1s	36.3s	200.0s
LP ℓ^1 -BPDN	28.1s	303.4s	n/a	n/a

Figures 3(a), and 3(b) display denoising results for standard ℓ^1 -TV (included for the sake of quality assessment), and ℓ^1 BPDN via the proposed algorithm. Note that even though the ℓ^1 -TV result has a slightly higher SNR than BPDN with ℓ^1 data fidelity term, the latter has a superior visual quality.

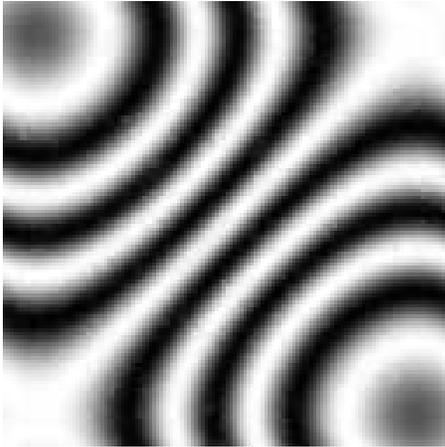
IV. CONCLUSIONS

As previously noted [12], [11], ℓ^1 BPDN (and related problems) provide superior performance to the corresponding ℓ^2 versions in certain applications. The proposed IRN-BPDN algorithm provides a flexible and computationally efficient

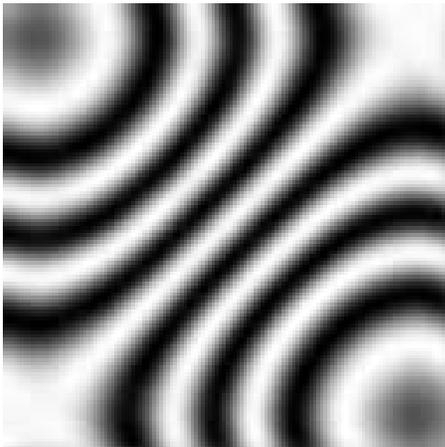
means of solving the generalized BPDN problem, including the ℓ^1 BPDN problem. The computational advantages of IRN-BPDN are such that this method may be applied to problem sizes that are impractical via the technique of mapping to a linear program.

REFERENCES

- [1] I. F. Gorodnitsky and B. D. Rao, "Sparse signal reconstruction from limited data using FOCUSS: a re-weighted minimum norm algorithm.," *IEEE Transactions on Signal Processing*, vol. 45, pp. 600–616, Mar. 1997.
- [2] S. G. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Transactions on Signal Processing*, vol. 41, pp. 3397–3415, Dec. 1993.
- [3] S. Chen and D. Donoho, "Application of basis pursuit in spectrum estimation," in *Proceedings ICASSP-98 (IEEE International Conference on Acoustics, Speech and Signal Processing)*, vol. 3, (Seattle, WA, USA), pp. 1865–1868, May 1998.
- [4] S. Chen, D. Donoho, and M. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. on Sci. Comp.*, vol. 20, no. 1, pp. 33–61, 1998.
- [5] L. Peotta, L. Granai, and P. Vanderghyest, "Very low bit rate image coding using redundant dictionaries," *Proceedings of SPIE—the international society for optical engineering*, vol. 5207, no. 1, pp. 228–239, 2003.
- [6] J. Starck, M. Elad, and D. Donoho, "Image decomposition via the combination of sparse representations and a variational approach," *IEEE Transactions On Image Processing*, vol. 14, pp. 1570–1582, Oct. 2005.
- [7] B. D. Rao and K. Kreutz-Delgado, "An affine scaling methodology for best basis selection," *IEEE Transactions On Signal Processing*, vol. 47, pp. 187–200, Jan. 1999.
- [8] J. A. Scales and A. Gersztenkorn, "Robust methods in inverse theory," *Inverse Problems*, vol. 4, pp. 1071–1091, Oct. 1988.
- [9] T. Chan, S. Esedoglu, F. Park, and A. Yip, "Recent developments in total variation image restoration," in *The Handbook of Mathematical Models in Computer Vision* (N. Paragios, Y. Chen, and O. Faugeras., eds.), Springer, 2005.
- [10] M. Nikolova, "A variational approach to remove outliers and impulse noise," *J. of Math. Imaging and Vision*, vol. 20, pp. 99–120, 2004.
- [11] H. Fu, M. K. Ng, M. Nikolova, and J. L. Barlow, "Efficient minimization methods of mixed ℓ_2 - ℓ_1 and ℓ_1 - ℓ_1 norms for image restoration," *SIAM J. Sci. Comput.*, vol. 27, no. 6, pp. 1881–1902, 2006.
- [12] L. Granai and P. Vanderghyest, "Sparse Approximation by Linear Programming using an L1 Data-Fidelity Term," in *Proc. of Workshop on Signal Processing with Adaptive Sparse Structured Representations*, Parallel Computing in Electrical Engineering, 2005.
- [13] P. Rodríguez and B. Wohlberg, "An iteratively weighted norm algorithm for total variation regularization," in *Proceedings of the 2006 Asilomar Conference on Signals, Systems, and Computers*, (Pacific Grove, CA, USA), pp. 892–896, Oct. 2006.



(a) Denoised image via ℓ^1 -TV. SNR: 25.77dB



(b) Denoised image (using the proposed algorithm) via BPDN with ℓ^1 data fidelity term using an over-complete DCT dictionary. SNR: 25.62dB

Fig. 3. ℓ^1 -TV 3(a) and 3(b) ℓ^1 BPDN. Note that even though the ℓ^1 -TV result has a slightly higher SNR than ℓ^1 -BPDN, the latter has a superior visual quality.

terminated systems (SparseLab).” Software library available from <http://sparselab.stanford.edu/>.

Paul Rodríguez received the BSc degree in electrical engineering from the Pontificia Universidad Católica del Perú, Lima, Peru, in 1997, and the MSc and PhD degrees in electrical engineering from the University of New Mexico, USA, in 2003 and 2005 respectively.

He was a postdoctoral research associate at Los Alamos National Laboratory, NM, USA, (August 2005 - August 2007). He is currently an Associate Professor with the Department of Electrical Engineering at Pontificia Universidad Católica del Perú, Lima, Peru. His research interests include AM-FM models, simd-algorithms, wavelets and adaptive signal decompositions, and inverse problems in signal and image processing.

- [14] B. Wohlberg and P. Rodríguez, “An iteratively reweighted norm algorithm for minimization of total variation functionals,” *IEEE Signal Processing Letters*, vol. 14, pp. 948–951, Dec. 2007.
- [15] A. E. Beaton and J. W. Tukey, “The fitting of power series, meaning polynomials, illustrated on band-spectroscopic data,” *Technometrics*, no. 16, pp. 147–185, 1974.
- [16] R. Wolke and H. Schwetlick, “Iteratively reweighted least squares: Algorithms, convergence analysis, and numerical comparisons,” *SIAM J. on Sci. and Stat. Comp.*, vol. 9, pp. 907–921, Sept. 1988.
- [17] S. A. Ruzinsky and E. T. Olsen, “ L_1 and L_∞ minimization via a variant of Karmarkar’s algorithm,” *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 37, pp. 245–253, 1989.
- [18] K. P. Bube and R. T. Langan, “Hybrid ℓ^1/ℓ^2 minimization with applications to tomography,” *Geophysics*, vol. 62, pp. 1183–1195, July-August 1997.
- [19] I. Gorodnitsky and B. Rao, “A new iterative weighted norm minimization algorithm and its applications,” in *IEEE Sixth SP Workshop on Statistical Signal and Array Processing, 1992*, (Victoria, BC, Canada), Oct. 1992.
- [20] P. J. Huber, “Robust regression: Asymptotics, conjectures and monte carlo,” *The Annals of Statistics*, vol. 1, no. 5, pp. 799–821, 1973.
- [21] P. Rodríguez and B. Wohlberg, “Numerical methods for inverse problems and adaptive decomposition (NUMIPAD).” Software library available from <http://numipad.sourceforge.net/>.
- [22] D. Donoho, V. Stodden, and Y. Tsaig, “Sparse solutions to underde-

Brendt Wohlberg received the BSc(Hons) degree in applied mathematics, and the MSc(Applied Science) and PhD degrees in electrical engineering from the University of Cape Town, South Africa, in 1990, 1993 and 1996 respectively.

He is currently a technical staff member in the Mathematical Modeling and Analysis Group (T-7) at Los Alamos National Laboratory, Los Alamos, NM, USA. His research interests include image coding, pattern recognition, wavelets and adaptive signal decompositions, and inverse problems in signal and image processing.