LA-UR- 07-0078

Approved for public release;
distribution is unlimited.

Title: SPARSE REPRESENTATIONS WITH DATA FIDELITY TERM VIA AN ITERATIVELY REWEIGHTED LEAST SQUARES ALGORITHM

Author(s): BRENDT WOHLBERG
PAUL RODRIGUEZ

Intended for: PROCEEDINGS:
IEEE INTERNATIONAL CONFERENCE ON IMAGE PROCESSING

Los Alamos
NATIONAL LABORATORY
EST. 1943

Sparse Representations with ℓ^1 Data Fidelity Term via an Iteratively Reweighted Least Squares Algorithm

Paul Rodríguez and Brendt Wohlberg

Abstract— **Basis Pursuit and Basis Pursuit Denoising, well established techniques for computing sparse representations, minimize an ℓ^2 data fidelity term subject to an ℓ^1 sparsity constraint or regularization term on the solution by mapping the problem to a linear or quadratic program. Basis Pursuit Denoising 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 greater flexibility in the choice of data fidelity term norm, and computational advantages in certain circumstances.**

I. Introduction

Recently, several authors have considered sparse approximation and image restoration problems with respect to error measures other than the usual ℓ^2 norm (Euclidean distance). Specifically, problems which include a ℓ^1 data fidelity term have attracted great attention.

In the case of sparse approximations, the Basis Pursuit (BP) and Basis Pursuit denoising (BPDN) paradigm, introduced in [1], consist in the following minimization problems:

$$\begin{array}{ll} \text{BP} & \min \left\|\mathbf{u}\right\|_1 \ \text{ subject to } \Phi\mathbf{u} = \mathbf{b} \\ \text{BPDN} & \min \ \frac{1}{2}\left\|\Phi\mathbf{u} - \mathbf{b}\right\|_2^2 + \lambda\left\|\mathbf{u}\right\|_1 \end{array}$$

where Φ is a $n \times p$ matrix form using the basis vectors from an (in general) overcomplete dictionary i.e. $p \gg n$, and it is assumed that rank$(\Phi) = n$. The BP and BPDN problems are solved using linear or quadratic programming via interior point methods (see [1]) or by iteratively solving a weighted ℓ^2 version of the original problem (see [2], [3]). The Affine Scaling Transformation (AST) algorithm ([4], [5]) was proposed to minimize a modified BP and BPDN cost functional; these functionals use the ℓ^p norm to measure the sparsity of the representation sought in the overcomplete dictionary Φ

$$\begin{array}{ll} \text{AST-BP} & \min \left\|\mathbf{u}\right\|_p^p \ \text{ subject to } \Phi\mathbf{u} = \mathbf{b} \\ \text{AST-BPDN} & \min \ \frac{1}{2}\left\|\Phi\mathbf{u} - \mathbf{b}\right\|_2^2 + \frac{\lambda}{p}\left\|\mathbf{u}\right\|_p^p \end{array}$$

and it can be summarized as:

$$\mathbf{u}^{(k)} = W^{(k)^2}\Phi^T\left(\Phi W^{(k)^2}\Phi^T + \lambda I\right)^{-1}\mathbf{b} \qquad (1)$$

where $W^{(k)} = \text{diag}\left(\left|\mathbf{u}^{(k-1)}\right|^{1-\frac{p}{2}}\right)$. Note that setting $p = 1$ and $\lambda = 0$ will solve the BP problem, whereas if $\lambda > 0$ will solve the BPDN problem. More recently, [6] studies the BPDN problem with an ℓ^1 data fidelity term

$$\text{BPDN } \ell^1 - \text{Fidelity} \qquad \min \ \left\|\Phi\mathbf{u} - \mathbf{b}\right\|_1 + \lambda\left\|\mathbf{u}\right\|_1$$

and shows the equivalence of the above minimization problem with Linear Programming; as an application, denoising in the presence of impulse noise is considered for (only) 1D signals.

In the case of image restoration, [7] considers the minimization of a cost functional with an ℓ^1 norm for both the fidelity and regularization terms. More specifically, for Total Variation (TV), the inclusion of the ℓ^1 data fidelity term [8], [9] has a number of advantages, including superior denoising performance with salt and pepper (speckle) noise [10]; there are a fair amount of algorithms [11], [9], [10], [12], [13], [14] to tackle this problem.

In this paper we propose a simple yet flexible and computationally efficient algorithm to solve the generalized BP problem

$$\text{Generalized BPDN} \qquad \min \ \frac{1}{q}\left\|\Phi\mathbf{u} - \mathbf{b}\right\|_q^q + \frac{\lambda}{p}\left\|\mathbf{u}\right\|_p^p$$

this algorithm, called Iteratively Reweighted Norm for BPDN (or IRN-BPDN), which is an extension to the AST algorithm [4], [5], is closely related to the Iteratively Reweighted Norm for Total Variation (or IRN-TV, see [11]).

II. IRN-BPDN Algorithm

A. *Fidelity term*

$$W_F = \text{diag}\left(\frac{2}{q}f_F(\Phi\mathbf{u} - \mathbf{b})\right),$$

where f_F is defined (for some small ϵ_F) as

$$f_F(x) = \begin{cases} |x|^{1-\frac{q}{2}} & \text{if } |x| > \epsilon_F \\ \epsilon_F^{1-\frac{q}{2}} & \text{if } |x| \le \epsilon_F, \end{cases}$$

Paul Rodríguez is with T-7 Mathematical Modeling and Analysis, Los Alamos National Laboratory, Los Alamos, NM 87545, USA. Email: prodrig@t7.lanl.gov, Tel: (505) 606 1483 , Fax: (505) 665 5757

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: (505) 667 6886, Fax: (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.

Algorithm 1 IRN-BPDN algorithm

Initialize

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

Iterate

$$W_F^{(k)} = \mathrm{diag}\left(\frac{2}{q} f_F(\Phi \mathbf{u}^{(k-1)} - \mathbf{b})\right)$$

$$W_R^{(k)} = \mathrm{diag}\left(\frac{2}{p} f_R\left(\mathbf{u}^{(k-1)}\right)\right)$$

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

$$\mathbf{u}^{(k)} = \left(W_R^{(k)}\right)^2 \Phi^T \boldsymbol{\chi}^{(k)}$$

B. Sparsity term

See [4], [5] for details.

$$W_R = \mathrm{diag}\left(\frac{2}{p} f_R(\mathbf{u})\right),$$

where f_R is defined (for some small ϵ_R) as

$$f_R(x) = \begin{cases} |x|^{1-\frac{p}{2}} & \text{if } |x| > \epsilon_R \\ 0 & \text{if } |x| \leq \epsilon_R \end{cases}$$

C. Derivation

$$\frac{1}{q} \|\Phi \mathbf{u} - \mathbf{b}\|_q^q + \frac{\lambda}{p} \|\mathbf{u}\|_p^p$$

$$\frac{1}{2} \left\|W_F{}^{-1} \Phi \mathbf{u} - W_F{}^{-1} \mathbf{b}\right\|_2^2 + \frac{\lambda}{2} \left\|W_R{}^{-1} \mathbf{u}\right\|_2^2$$

Set $\mathbf{u} = W_R \boldsymbol{\nu}$

$$\frac{1}{2} \left\|W_F{}^{-1} \Phi W_R \boldsymbol{\nu} - W_F{}^{-1} \mathbf{b}\right\|_2^2 + \frac{\lambda}{2} \|\boldsymbol{\nu}\|_2^2$$

gradient

$$W_R \Phi^T W_F{}^{-2} \Phi W_R \boldsymbol{\nu} - W_R \Phi^T W_F{}^{-2} \mathbf{b} + \lambda \boldsymbol{\nu} = 0$$

set $\boldsymbol{\nu} = W_R \Phi^T \boldsymbol{\chi}$

$$W_R \Phi^T W_F{}^{-2} \Phi W_R \boldsymbol{\nu} - W_R \Phi^T W_F{}^{-2} \mathbf{b} + \lambda W_R \Phi^T \boldsymbol{\chi} = 0$$

$$W_R \Phi^T W_F{}^{-2} \left(\Phi W_R{}^2 \Phi^T \boldsymbol{\chi} - \mathbf{b} + \lambda W_F{}^2 \boldsymbol{\chi}\right) = 0$$

solve

$$\boldsymbol{\chi} = \left(\Phi W_R{}^2 \Phi^T + \lambda W_F{}^2\right)^{-1} \mathbf{b}$$

Replace for $\boldsymbol{\nu}$ and $\mathbf{u}$

III. Results

Granai and Vandergheynst [6] have previously discussed the advantages of BPDN with an ℓ^1 data fidelity term. Here were provide additional evidence based on examples computed using the proposed algorithm. A simple cubic phase cosine image is displayed in Figure 1, and the same image after addition of 5% speckle noise is displayed in Figure 2. This simple example is intended to illustrate the advantages of BPDN when an appropriate dictionary is available, in this case the DCT. Figures 3(a), 3(b), and 3(c) display denoising results for standard ℓ^2 BPDN, ℓ^1-TV, 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 former has a superior visual quality.

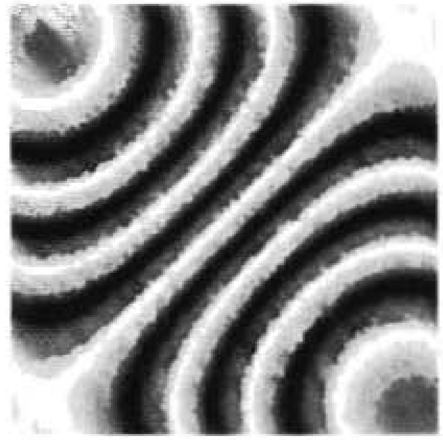


Fig. 1. 128×128 Cubic phase image: $I(x, y) = \cos(x^3 - y^3 - 9x - 9y)$

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

IV. Conclusions

The proposed algorithm provides a flexible and computationally efficient means of solving the generalized BPDN problem, including the ℓ^1 BPDN problem.

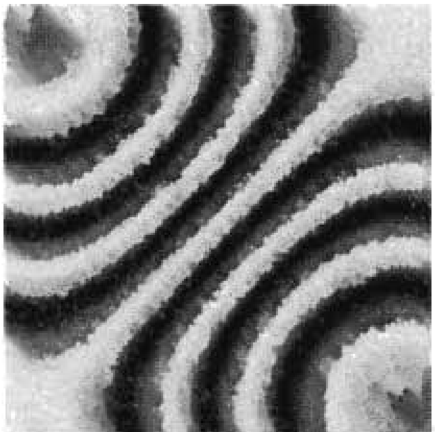
Note

This is an early draft version of this paper, submitted for an LA-UR number at this stage due to an approaching submission deadline.

References

[1] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Journal on Scientific Computing*, vol. 20, no. 1, pp. 33–61, 1999.

[2] I. F. Gorodnitsky and B. D. 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.



(a) Denoised image via (standard) BPDN using an overcomplete DCT dictionary. SNR: 16.02dB.

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

(c) Denoised image via BPDN with ℓ^1 data fidelity term using an overcomplete DCT dictionary. SNR: 25.62

Fig. 3. Comparison of 3 denoising strategies: 3(a) (standard) BPDN, 3(b) ℓ^1-TV and 3(c) BPDN with ℓ^1 data fidelity term. Note that even though the ℓ^1-TV result has a slightly higher SNR than BPDN with ℓ^1 data fidelity term, the former has a superior visual quality.

[3] 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.

[4] 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.

[5] B. D. Rao, K. Engan, S. F. Cotter, J. Palmer, and K. Kreutz-Delgado, "Subset selection in noise based on diversity measure minimization," *IEEE Transactions On Signal Processing*, vol. 51, pp. 760–770, 2003.

[6] L. Granai and P. Vandergheynst, "Sparse Approximation by Linear Programming using an L1 Data-Fidelity Term," in *Proc. of Workshop on Signal Processing with Adaptative Sparse Structured Representations*, Parallel Computing in Electrical Engineering, 2005.

[7] H. Fu, M. K. Ng, M. Nikolova, and J. L. Barlow, "Efficient minimization methods of mixed l2-l1 and l1-l1 norms for image restoration," *SIAM J. Sci. Comput.*, vol. 27, no. 6, pp. 1881–1902, 2006.

[8] M. Nikolova, "Minimizers of cost-functions involving nonsmooth data-fidelity terms. application to the processing of outliers," *SIAM Journal on Numerical Analysis*, vol. 40, no. 3, pp. 965–994, 2002.

[9] T. F. Chan and S. Esedoglu, "Aspects of total variation regularized l^1 function approximation," *SIAM Journal on Applied Mathematics*, vol. 65, no. 5, pp. 1817–1837, 2005.

[10] M. Nikolova, "A variational approach to remove outliers and impulse noise," *Journal of Mathematical Imaging and Vision*, vol. 20, no. 1-2, pp. 99–120, 2004. 4th Conference on Mathematics and Image Analysis; September 10-13, 2002; Paris, France.

[11] P. Rodríguez and B. Wohlberg, "An iteratively reweighted norm algorithm for total variation regularization," in *Proc. of the Asilomar Conference on Signals, Systems, and Computers*, 2006.

[12] J. Darbon and M. Sigelle, "A fast and exact algorithm for total variation minimization," *Lecture Notes in Computer Science*, vol. 3522/2005, pp. 351–359, 2005.

[13] D. Goldfarb and W. Yin, "Second-order cone programming methods for total variation-based image restoration," *SIAM J. Sci. Comput.*, vol. 27, no. 2, pp. 622–645, 2005.

[14] J.-F. Aujol, G. Gilboa, T. Chan, and S. Osher, "Structure-texture image decomposition - modeling, algorithms, and parameter selection," *International Journal of Computer Vision*, vol. 67, no. 1, pp. 111–136, 2006.