SYNTHESIS AND ANALYSIS PRIOR ALGORITHMS FOR JOINT-SPARSE RECOVERY

A. Majumdar and R. K. Ward

Department of Electrical and Computer Engineering, University of British Columbia
{angshulm, rababw}@ece.ubc.ca

ABSTRACT

This paper proposes a *Majorization-Minimization* approach for solving the synthesis and analysis prior joint-sparse multiple measurement vector reconstruction problem. The proposed synthesis prior algorithm yielded the same results as the Spectral Projected Gradient (SPG) method. The analysis prior algorithm is the first to be proposed for this problem. It yielded considerably better results than the proposed synthesis prior algorithm. For problems of a given size, the run times for our proposed algorithms are fixed; unlike SPG where the reconstruction time also depends on the support size of the vectors.

Index Terms— Compressed Sensing, Multiple Measurement Vector, Convex Optimization

1. INTRODUCTION

In Compressed Sensing (CS) the problem is to recover a sparse vector from its random low dimensional Fourier projections that may be corrupted by noise.

$$y_{n\times 1} = H_{n\times N} x_{N\times 1} + \eta_{n\times 1},\ n < N \tag{1}$$

where x is the high dimensional s-sparse vector to be recovered, H is the random/Fourier projection, y is the lower dimensional projection of x and η is the noise (assumed to be Normally distributed).

An extension of the Single Measurement Vector (SMV) problem represented by (1) is the Multiple Measurement Vector (MMV) problem. For MMV, a set of vectors with a common sparse support are to be recovered. By common support, it is meant that all the vectors have non-zero coefficients at the same positions. The MMV model is,

$$y_j = Hx_j + \eta_j,\ j = 1...\text{r} \tag{2}$$

where y_j is the lower dimensional projection corresponding to x_j (all x_j's have a common support as mentioned earlier). The rest of the symbols have the same meanings as in (1).

It is possible to represent (2) in the following compact form,

$$Y = HX + N \tag{3}$$

where $Y = [y_1 | ... | y_r]$, $X = [x_1 | ... | x_r]$ and $N = [\eta_1 | ... | \eta_r]$.

Since the unknown vectors (x's) have a common sparse support, the matrix X will be row-sparse. The MMV problem has been studied in the past [1-5]. Optimization based recovery methods recover X by solving the following problem

$$\min_X \|X\|_{m,p}^p \text{ subject to } \|Y - HX\|_F^2 \le \varepsilon \tag{4}$$

where $\|X\|_{m,p}^p = \sum_{j=1}^{n} \|X^{j\to}\|_m^p$ ($X^{j\to}$ is the vector whose entries form the j^{th} row of X), $\|.\|_F$ denotes the Frobenius norm (l_2-norm of all the elements) of the matrix and ε is a parameter dependent on the variance of noise.

In [2], the values m=2 and p≤1 were proposed. Since values of p. such as p<1, make the problem non-convex, m=2 and p=1 are used [6]. Thus the inverse problem (3) is solved via,

$$\min_X \|X\|_{2,1} \text{ subject to } \|Y - HX\|_F^2 \le \varepsilon \tag{5}$$

The choice of such values for the norms can be understood intuitively. The l_2-norm over the rows ($X^{j\to}$'s) enforces non-zero values on all the elements of the row vector whereas the summation over the l_2-norm ($\sum_{j=1}^{r} \|X^{j\to}\|_2$) enforces row-sparsity, i.e. the selection of few rows.

MMV problems arise in varied areas of applied signal processing like communication [7], Seismic Imaging [8] and MRI [9, 10]. For example, in multi-echo T1/T2 weighted MR imaging, the same anatomical cross section is imaged by varying certain parameters in order to acquire images with differing tissue contrasts. It has been argued in [9, 10], that such multi-echo images have a common sparse support in the wavelet domain.

It is possible to recover the signal by (5) when the signal itself is sparse. However in MRI, the image to be recovered is not sparse, but has a sparse representation in some transform domains (e.g. finite difference). In such a case the synthesis prior formulation (5) is not applicable. Once needs to employ the following analysis prior formulation,

$$\min_X \|AX\|_{2,1} \text{ subject to } \|Y - HX\|_F^2 \le \varepsilon \tag{6}$$

where A is the sparsifying transform.

The Spectral Projected Gradient L1 algorithm [6] solves the synthesis prior joint-sparse MMV recovery problem (5). There is no existing algorithm to solve the analysis prior problem (6). In this paper we propose algorithms for solving the synthesis prior (5) and the analysis prior problems using the Majorization-Minimization approach. These algorithms are derived in Section 2 below and in Section 3, the detailed experimental results are shown for simulated data.

2. INFORMAL DERIVATION OF ALGORITHMS

The Majorization-Minimization (MM) approach [11, 12] is employed to derive solution to the following problems,

Synthesis prior: $\min_X \|X\|_{2,1}$ subject to $\|Y-HX\|_F^2 \le \varepsilon$

Analysis prior: $\min_X \|AX\|_{2,1}$ subject to $\|Y-HX\|_F^2 \le \varepsilon$

Instead of solving the aforesaid constrained problems, we propose solving their unconstrained counterparts,

$$\min_X J_1(X), \text{ where } J_1(X) = \frac{1}{2}\|Y-HX\|_F^2 + \lambda\|X\|_{2,1} \quad (7)$$

$$\min_X J_2(X), \text{where } J_2(X) = \frac{1}{2}\|Y-HX\|_F^2 + \lambda\|AX\|_{2,1} \quad (8)$$

The constrained and the unconstrained formulations are equivalent for proper choice of the Lagrangian λ. Unfortunately for most practical problems it is not possible to determine λ explicitly by analytical means. Therefore, instead of 'guessing' λ, given the value of ε, we will reach the solution of the constrained problem by iteratively solving a series of unconstrained problems with decreasing values of λ. Such cooling techniques are successful [13-15] since the Pareto curve for the said problem is smooth [13].

2.1. Majorization Minimization

Owing to the limitations in space, we do not derive every step of the algorithm since they can be found in previous works which will be duly referred. Following the MM technique outlined in [14] (and successfully used in [17, 18]), in each iteration (i), the aforesaid minimization problems on J_1 and J_2 can be substituted by the following,

$$\min_X G_1^{(i)}(X),\ G_1^{(i)}(X) = \frac{1}{2}\|B^{(i)}-X\|_F^2 + \frac{\lambda}{\alpha}\|X\|_{2,1} \quad (9)$$

$$\min_X G_2^{(i)}(X),\ G_2^{(i)}(X) = \frac{1}{2}\|B^{(i)}-X\|_F^2 + \frac{\lambda}{\alpha}\|AX\|_{2,1} \quad (10)$$

$$B^{(i)} = X^{(i)} + \frac{1}{\alpha}H^T(Y-HX^{(i)}) \quad (11)$$

where α is greater than the maximum eigenvalue of H^TH. The update (11) is known as the Landmeyer iterations.

2.2. Solving the synthesis prior problem

For the synthesis prior problem, we need to solve (9) at each iteration. Taking the derivative of $G_1^{(i)}(X)$ we get,

$$\frac{dG_1^{(i)}(X)}{dX} = X - B^{(i)} + \frac{\lambda}{\alpha}\Lambda signum(X) \quad (12)$$

where $\Lambda = diag(\|X^{(i)j\rightarrow}\|_2^{-1})|X^{(i)}|$.

Setting the derivative to zero and re-arranging, we get,

$$B = X + \frac{\lambda}{\alpha}\Lambda signum(X) \quad (13)$$

This can be solved by the following soft-thresholding,

$$X^{(i+1)} = signum(B^{(i)})\max(0, |B^{(i)}| - \frac{\lambda}{\alpha}\Lambda) \quad (14)$$

Equations (11) and (14) suggest a compact solution for the unconstrained synthesis prior problem. This is given in the following algorithm.

Initialize: $X^{(0)} = 0$
Repeat until convergence:
$B^{(i)} = X^{(i)} + \frac{1}{\alpha}H^T(Y-HX^{(i)})$
$X^{(i+1)} = signum(B^{(i)})\max(0, \|B^{(i)}\| - \frac{\lambda}{\alpha}\Lambda)$

2.3. Solving the analysis prior problem

Solving the analysis prior problem requires minimization of (10) in each iteration. Taking the derivative of $G_2^{(i)}(X)$ we get,

$$\frac{dG_2^{(i)}(X)}{dX} = X - B^{(i)} + \frac{\lambda}{\alpha}A^T\Omega AX \quad (15)$$

where $\Omega = diag(\|W^{(i)j\rightarrow}\|_2^{-1})$ and $W_{M\times r} = A_{M\times N}X_{N\times r}$.

Setting the gradient to zero we get,

$$(I + \frac{\lambda}{\alpha}A^T\Omega A)X = B^{(i)} \quad (16)$$

It is not possible to solve (16) directly as the sparsifying transform (A) in most cases is available as a fast operator and not as an explicit matrix. The derivation of the solution to (16) is similar to the derivation of analysis prior sparse optimization [11] and analysis prior group-sparse optimization [16]. Thus we skip the detailed derivations; and show the final update equations,

$$Z^{(i+1)} = (\frac{\alpha}{\lambda}\Omega^{-1} + cI)^{-1}(cZ^{(i)} + A(B^{(i)} - A^TZ^{(i)})) \quad (16)$$

$$X^{(i+1)} = B^{(i)} - A^TZ^{(i)} \quad (17)$$

where c is greater than the maximum eigenvalue of A^TA.

This leads to the following algorithm for solving the analysis prior joint-sparse optimization problem.

Initialize: $X^{(0)} = 0$
Repeat until convergence:
$B^{(i)} = X^{(i)} + \frac{1}{\alpha}H^T(Y-HX^{(i)})$
$Z^{(i+1)} = (\frac{\alpha}{\lambda}\Omega^{-1} + cI)^{-1}(cZ^{(i)} + A(B^{(i)} - A^TZ^{(i)}))$
$X^{(i+1)} = B^{(i)} - A^TZ^{(i)}$

2.3. Solving the constrained problem via cooling

We have derived algorithms to solve the unconstrained problems. As mentioned before, the constrained and the unconstrained forms are equivalent for proper choice of ε and λ. However, there is no analytical relationship between them in general. When faced with a similar situation, we employed the cooling technique following previous studies [13].

The cooling technique solves the constrained problem in two loops. The outer loop decreases the value of λ. The inner loop solves the unconstrained problem for a specific value λ. As λ is progressively decreased, the solution of the unconstrained problem reaches the desired solution. Such a cooling technique works because the pareto curve between the objective function and the constraint is smooth. The cooling algorithm for the synthesis and analysis prior are:

Synthesis Prior Algorithm

Initialize: $X^{(0)} = 0$; $\lambda < \max(P^T x)$
Choose a decrease factor (*DecFac*) for cooling λ
Outer Loop: While[1] $\| y - Hx \|_F^2 \geq \varepsilon$
Inner Loop: While[2] $\dfrac{J^{(i)} - J^{(i+1)}}{J^{(i)} + J^{(i+1)}} \geq Tol$
$J^{(i)} = \| Y - HX^{(i)} \|_F^2 + \lambda \| X^{(i)} \|_{2,1}$
Compute: $B^{(i)} = X^{(i)} + \dfrac{1}{\alpha} H^T (Y - HX^{(i)})$
Compute: $X^{(i+1)} = signum(B^{(i)}) \max(0, \left|B^{(i)}\right| - \dfrac{\lambda}{\alpha} \Lambda)$
$J^{(i+1)} = \| Y - HX^{(i)} \|_F^2 + \lambda \| X^{(i)} \|_{2,1}$
End While[2] (inner loop ends)
$\lambda = \lambda \times DecFac$
End While[1] (outer loop ends)

Analysis Prior Algorithm

Initialize: $X^{(0)} = 0$; $\lambda < \max(P^T x)$
Choose a decrease factor (*DecFac*) for cooling λ
Outer Loop: While[1] $\| y - Hx \|_F^2 \geq \varepsilon$
Inner Loop: While[2] $\dfrac{J^{(i)} - J^{(i+1)}}{J^{(i)} + J^{(i+1)}} \geq Tol$
$J^{(i)} = \| Y - HX^{(i)} \|_F^2 + \lambda \| AX^{(i)} \|_{2,1}$
Compute: $B^{(i)} = X^{(i)} + \dfrac{1}{\alpha} H^T (Y - HX^{(i)})$
Update: $Z^{(i+1)} = (\dfrac{\alpha}{\lambda} \Omega^{-1} + cI)^{-1} (cZ^{(i)} + A(B^{(i)} - A^T Z^{(i)}))$
Update: $X^{(i+1)} = B^{(i)} - A^T Z^{(i)}$
$J^{(i+1)} = \| Y - HX^{(i)} \|_F^2 + \lambda \| AX^{(i)} \|_{2,1}$
End While[2] (inner loop ends)
$\lambda = \lambda \times DecFac$
End While[1] (outer loop ends)

3. EXPERIMENTAL RESULTS

We followed the experimental procedure as in [1]. The experiments were carried out on AMD64 machine with 4GB RAM running Matlab 2009a.

The first experiment shows that our proposed algorithm is as good as the benchmark Spectral Projected Gradient method [6] for solving the synthesis prior problem. The length of the vectors to be recovered (N) is 60 and the length of the measured vectors (n) is 20. An i.i.d Gaussian matrix of size 20 x 60 was used for projection (H). The following graph for recovery rates was produced by varying the number of sparse rows from one to n=20. The number of input/output vectors (r) was taken to be 2, 5 and 10 for three sets of experiments. The results are shown for 1000 trials for each configuration. If the Normalized Mean Squared Error (NMSE) between the recovered and the groundtruth is less than 10^{-3}, the recovery is considered successful.

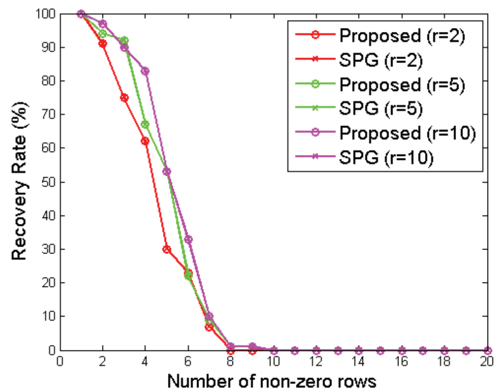

Fig. 1. Comparison of recovery rate for proposed algorithm and SPG for synthesis prior problem.

It is concluded from Fig. 1 that the recovery rates for our proposed algorithm and SPG are exactly the same. This is not surprising since both of them solve the same convex problem. Fig. 1 corroborates the findings in [1], that when the number of measurements increases (as r increases), the recovery improves, i.e. the curves shift to the right.

The average reconstruction times for SPG and our proposed method are shown in Fig. 2. (for r=5). The graph shows that the time taken by our proposed algorithm for problems of the same size is almost same. But for SPG, the time increases as the number of sparse rows increase.

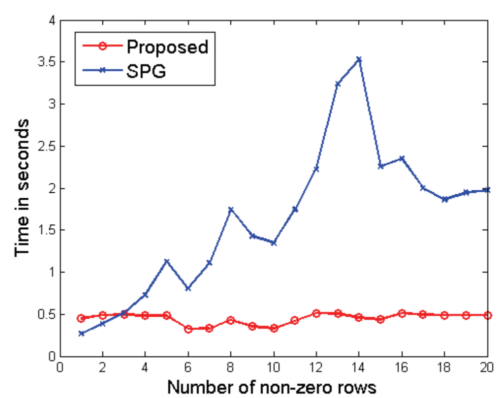

Fig. 2. Comparison of recovery time for proposed algorithm and SPG for synthesis prior problem.

The second experiment compares the synthesis prior algorithm with the analysis prior algorithm. The sparsifying transform (A) was chosen to be redundant Haar. The **measurement matrix** (H) for analysis prior algorithm is an i.i.d Gaussian matrix. For the synthesis prior algorithm, the projection matrix was the same Gaussian matrix post multiplied by the transpose of the sparsifying transform matrix (A^T). The length of the signal (N) to be recovered is 64. These signals were sparse in the redundant Haar transform domain. The number of input/measurement vectors (r) was fixed at 5.

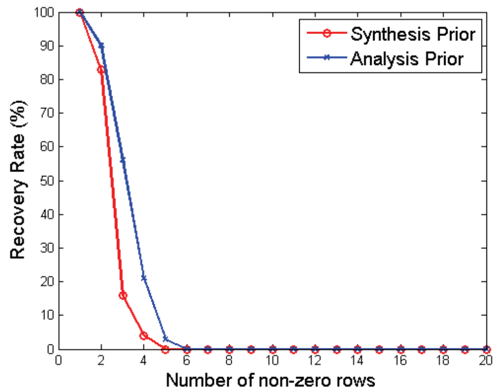

Fig. 3. Comparison of recovery rate for Synthesis and Analysis prior.

Fig. 3 shows that the recovery rate from the analysis prior algorithm is better than that from the synthesis prior algorithm. This observation is in line with the findings in [11]. We do not carry forth a similar experiment for an orthogonal sparsifying transform since the reconstruction accuracies are theoretically bound to be the same for analysis and synthesis prior methods.

Fig.4 shows the recovery time for the synthesis and analysis prior algorithms The analysis prior problem takes slightly more time. But the time required by both algorithms remains almost constant as the number of sparse rows varies.

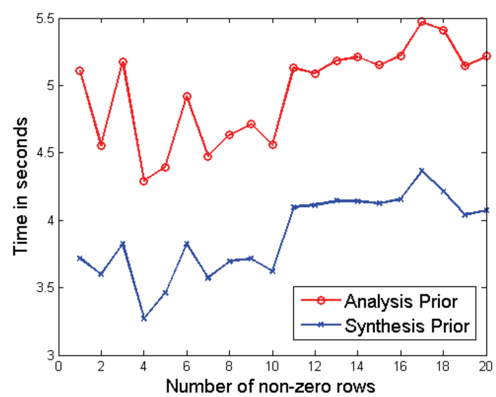

Fig. 4. Comparison of recovery time for Synthesis and Analysis prior algorithms

4. CONCLUSION

This paper proposed new algorithms for joint-sparse signal recovery. The multiple input vectors (to be recovered) have the same support, i.e. the ensemble of vectors is jointly sparse. The problem is to simultaneously recover these vectors from their lower dimensional projections.

The vectors to be recovered may be sparse or may be sparse is a transform domain. In the former case, the recovery is posed as a synthesis prior problem, whereas in the latter case it is posed as an analysis prior problem. This work proposes new algorithms to solve both problems. We compared our synthesis prior algorithm with the benchmark Spectral Projected Gradient (SPG) method. Our algorithm is as accurate as SPG. There was no existing algorithm to solve the analysis prior problem. Our algorithm is the first. It gives better results than the synthesis prior algorithm in situations where both are applicable. The execution speed of the proposed algorithms is only dependent on the size of the problem, unlike SPG, where the time increases as the size of the support for the vectors increase.

ACKNOWLEDGEMENT

This work was supported by NSERC, the Natural Sciences and Engineering Research Council of Canada and by QNRF, Qatar National Research Fund.

REFERENCES

[1] E. van den Berg and M. P. Friedlander, “Theoretical and empirical results for recovery from multiple measurements”, IEEE Trans. Info. Theory, Vol. 56 (5), pp. 2516-2527, 2010.

[2] S. F. Cotter, B. D. Rao, K. Engang, and K. Kreutz-Delgado, “Sparse solutions to linear inverse problems with multiple measurement vectors”. IEEE Trans. Sig. Proc., Vol. 53 (7), pp. 2477-2488, 2005.

[3] J. Chen and X. Huo, “Theoretical results on sparse represenations of multiple-measurement vectors”, IEEE Trans. Sig. Proc., Vol. 54 (12), pp. 4634-4643, 2006.

[4] J. A. Tropp, “Algorithms for simultaneous sparse approximation: Part II: Convex relaxation”, Signal Processing, Vol. 86, pp. 589,-602, 2006.

[5] M. Mishali and Y. C. Eldar, “Reduce and boost: Recovering arbitrary sets of jointly sparse vectors”, IEEE Trans. Sig. Proc., Vol. 56(10), pp. 4692-4702, 2008.

[6] http://www.cs.ubc.ca/labs/scl/spgl1/

[7] D. M. Malioutov, M. Cetin and A. S. Willsky, “A sparse signal reconstruction perspective for source localization with sensor arrays”, IEEE Trans. Sig. Proc., Vol 53 (8), pp. 3010-3022, 2005.

[8] F. J. Herrmann, “Compressive imaging by wavefield inversion with group sparsity”, SEG, 2009, Houston. Technical Report TR-2009-01.

[9] A. Majumdar and R. K. Ward, “Joint Reconstruction of Multi-echo MR Images Using Correlated Sparsity”, Magn. Reson. Imag. (accepted).

[10] A. Majumdar and R. K. Ward, “Accelerating Multi-Echo T2 Weighted MR Imaging: Analysis Prior Group Sparse Optimization”, J. Magn. Reson. (accepted).

[11] I. W. Selesnick and M. A. T. Figueiredo, “Signal restoration with overcomplete wavelet transforms: comparison of analysis and synthesis priors”, Proceedings of SPIE, Vol. 7446 (Wavelets XIII), 2009.

[12] http://cnx.org/content/m32168/latest/

[13] G. Hennenfent, E. van den Berg, M. P. Friedlander, and F. J. Herrmann, “New insights into one-norm solvers from the pareto curve”, Geophysics, Vol. 73(4), 2008.

[14] A. Majumdar and R. K. Ward, "Under-determined Non-Cartesian MR Reconstruction", MICCAI, pp. 513-520, 2010.

[15] A. Majumdar and R. K. Ward, “Compressive Color Imaging with Group Sparsity on Analysis Prior”, IEEE ICIP, pp. 1337-1340, 2010.