

EFFICIENT RECONSTRUCTION OF BLOCK-SPARSE SIGNALS

Joel Goodman

Keith Forsythe, Benjamin Miller

Naval Research Laboratory
4555 Overlook Ave. SW
Washington, DC 20375

MIT Lincoln Laboratory
244 Wood Street
Lexington, MA 02420

ABSTRACT

In many sparse reconstruction problems, M observations are used to estimate K components in an N dimensional basis, where $N > M \gg K$. The exact basis vectors, however, are not known a priori and must be chosen from an $M \times N$ matrix. Such under-determined problems can be solved using an ℓ_2 optimization with an ℓ_1 penalty on the sparsity of the solution. There are practical applications in which multiple measurements can be grouped together, so that $K \times P$ data must be estimated from $M \times P$ observations, where the ℓ_1 sparsity penalty is taken with respect to the vector formed using the ℓ_2 norms of the rows of the data matrix. In this paper we develop a computationally efficient block partitioned homotopy method for reconstructing $K \times P$ data from $M \times P$ observations using a grouped sparsity constraint, and compare its performance to other block reconstruction algorithms.

1. INTRODUCTION

Sparse reconstruction is essential to compressed sensing (CS) applications, where it has been shown that it is possible to recover a K -sparse signal of length N from $O(K \log N)$ compressive measurements [1–3].

In applications where measurements are obtained from multiple sensors [4], it is advantageous to group the sensor measurements together to provide robustness against impairments such as noise and signal fading. The formulation for multichannel measurements

$$\min_X \|\text{vec}(Y) - \Phi \text{vec}(X)\|_2^2 + \mu \sum_{i=1}^N \|e_i^T X\|_2 \quad (1)$$

uses an ℓ_1 sparsity constraint vector grouped across the ℓ_2 norms of the rows of the data matrix, where $\Phi \in \mathbb{R}^{MP \times NP}$ is the dictionary matrix, the matrix $Y \in \mathbb{R}^{M \times P}$ represents M observations from each of P sensor channels, $X \in \mathbb{R}^{N \times P}$ is the sparse data matrix, $\text{vec}(\cdot)$ is an operator which stacks the

rows of a matrix to form a column vector and $e_i \in \mathbb{R}^{N \times 1}$ is a column vector of 0s with a 1 in the i th position.

To solve (1) using block homotopy processing, we formulate the subgradients with respect to blocks (rows) of the matrix X , and the regularization parameter μ is systematically reduced, tracking the solution $X(\mu)$, until a new ℓ_2 normed row of the solution is about to turn nonzero. Reduction in μ continues, activating more and more rows of $X(\mu)$. We will demonstrate that an approximate form of block homotopy outperforms greedy block reconstruction (with arbitrary Φ in (1)) for roughly the same computational complexity, while solving (1) for all sparsity levels of X . The rest of this paper is organized as follows. In Section 2, we extend the homotopy technique in [5] to block partitioning. In Section 3, we demonstrate the performance of block homotopy continuation with respect to other block reconstruction algorithms, and in Section 4 we provide a brief summary.

2. BLOCK PARTITIONED HOMOTOPY

We return to the original optimization problem in (1) with a slight variation in notation,

$$L(x) \stackrel{\text{def}}{=} \|y - \Phi x\|_2^2 + \mu \sum_{k=1}^N \|x(k)\|_2, \quad (2)$$

where $x(k)$ is the k th row of X , and $y \in \mathbb{R}^{MP \times 1}$ and $x \in \mathbb{R}^{NP \times 1}$ are $\text{vec}(Y)$ and $\text{vec}(X)$, respectively. Complex vectors can be handled in the form (2) by grouping the real and imaginary components together after observing that complex ℓ_2 norms can be written as real ℓ_2 norms in these components.

The homotopy approach solves the optimization problem based on (2) for large μ and then tracks the solution as μ is decreased. Initially, μ is chosen to be $2\|\Phi^T y\|_2$, which yields an all-zero solution. Solutions are tracked by solving the subgradient equation for a minimum. Recall that the subdifferential at \tilde{x} of a convex function $f(x)$ is the set of subgradient vectors ξ that satisfy

$$f(x) - f(\tilde{x}) \geq (x - \tilde{x})^T \xi. \quad (3)$$

When $f(x)$ is differentiable at \tilde{x} , the subgradient consists of the derivative vector. When \tilde{x} minimizes $f(x)$, 0 is a subgra-

This work is sponsored by the Department of the Air Force under Air Force Contract FA8721-05-C-0002. Opinions, interpretations, conclusions and recommendations are those of the author and are not necessarily endorsed by the United States Government.

dient of $f(x)$ at \tilde{x} . The derivative of $\|x\|_2$ is $x/\|x\|_2$ away from zero. With $f(\cdot) = \|\cdot\|_2$ and $\tilde{x} = 0$, we observe that $\|x\|_2 \geq x^T \xi$, where $\|\xi\|_2 \leq 1$.

If $f(x)$ is the sum of convex functions $f = \sum_k f_k(x)$, each of which has subgradients ξ_k , then $\xi = \sum_k \xi_k$ is a subgradient of f . Using this observation, the subdifferential of $L(x)$ contains the set

$$\{2(\Phi^T \Phi x - \Phi^T y) + \mu [u_1^T \cdots u_N^T]^T\}, \quad (4)$$

where

$$u_k \stackrel{\text{def}}{=} \begin{cases} \frac{x(k)}{\|x(k)\|_2} & x(k) \neq 0 \\ \xi_k : \|\xi_k\|_2 \leq 1 & x(k) = 0 \end{cases}. \quad (5)$$

Any ξ_k satisfying the constraints can be used.

Let \mathcal{C}_{off} represent a subset of the positive integers less than or equal to N such that $k \in \mathcal{C}_{\text{off}}$ implies $x(k) = 0$. Let \mathcal{C}_{on} represent the complement of \mathcal{C}_{off} in the same set of positive integers. Without loss of generality, assume that the components are ordered so that

$$x = \begin{pmatrix} x_{\text{on}} \\ x_{\text{off}} \end{pmatrix} = \begin{pmatrix} x_{\text{on}} \\ 0 \end{pmatrix}, \quad (6)$$

where the subscripts refer to the sets \mathcal{C}_{on} and \mathcal{C}_{off} . Partition the matrices accordingly, letting

$$\begin{pmatrix} A & B \\ B^T & C \end{pmatrix} = 2\Phi^T \Phi \quad (7)$$

and

$$z = \begin{pmatrix} z_{\text{on}} \\ z_{\text{off}} \end{pmatrix} = 2\Phi^T y. \quad (8)$$

Then zero is a subgradient if

$$\begin{pmatrix} A & B \\ B^T & C \end{pmatrix} \begin{pmatrix} x_{\text{on}} \\ 0 \end{pmatrix} + \mu \begin{pmatrix} u_{\text{on}} \\ u_{\text{off}} \end{pmatrix} = \begin{pmatrix} z_{\text{on}} \\ z_{\text{off}} \end{pmatrix}. \quad (9)$$

The block components u_k with $k \in \mathcal{C}_{\text{on}}$ are required to have unit ℓ_2 norm, even if the corresponding $x(k)$ vanishes. This condition holds for the initial nonzero solution and will hold throughout the homotopy. We can unravel the matrix equation to get

$$\begin{aligned} x_{\text{on}} &= A^{-1}(z_{\text{on}} - \mu u_{\text{on}}) \\ u_{\text{off}} &= \frac{1}{\mu}(z_{\text{off}} - B^T A^{-1} z_{\text{on}}) + B^T A^{-1} u_{\text{on}}. \end{aligned} \quad (10)$$

For sufficiently large μ , (2) has the all-zero solution. In fact, for real ρ_k , observe that

$$\min_{\|\rho\|_2=1} \sum_k |\rho_k| = \left(\max_{\sum_k |\rho_k|=1} \|\rho\|_2 \right)^{-1} = 1 \quad (11)$$

since $\|\cdot\|_2$ is a convex function and thus assumes its maximum at the vertices of the simplex $\{\rho : \rho_k \geq 0 \text{ and } \sum \rho_k = 1\}$.

Thus, letting $\rho_k = x_k/\|x\|_2$, $\|x\|_1 \geq \|x\|_2$. Similarly, letting $\rho_k = \|x(k)\|_2/\|x\|_2$, we have $\sum_k \|x(k)\|_2 \geq \|x\|_2$. Consequently, the nonnegative, convex function

$$L_{\text{low}} \stackrel{\text{def}}{=} \|y - \Phi x\|_2^2 + \mu \|x\|_2 \quad (12)$$

is a lower bound on $L(x)$. If we show L_{low} is minimized at $x = 0$, it follows that $x = 0$ also minimizes $L(x)$. Furthermore, this minimum is unique since L_{low} is strictly convex except potentially on rays from the origin where $\|x\|_2$ is not strictly convex. However, restricted to any ray, L_{low} has a unique minimum and, hence, $x = 0$ is the unique minimum of L .

To minimize L_{low} , take the gradient away from $x = 0$ and take the inner product with x , getting

$$(x^T \nabla L_{\text{low}})(x) = 2x^T (\Phi^T \Phi x - \Phi^T y) + \mu x^T u, \quad (13)$$

where $u = x/\|x\|_2$. This is strictly positive if

$$\mu x^T u = \mu \|x\|_2 > 2|x^T \Phi^T y|. \quad (14)$$

Since $|x^T \Phi^T y| \leq \|x\|_2 \|\Phi^T y\|_2$, we have a strictly positive inner product when $\mu > 2\|\Phi^T y\|_2$. In this case, a nonzero x cannot minimize L_{low} , so $x = 0$ provides the minimum.

Starting with $x = 0$ and $\mu = 2\|\Phi^T y\|_2 = \|z_{\text{off}}\|_2 = \|z\|_2$, it is clear that (10) is solved with a valid $u_{\text{off}} = u$: $\|u_k\|_2 \leq 1$. In fact, this is the case as long as

$$\mu \geq \max_k \|z(k)\|_2. \quad (15)$$

When μ is reduced until equality holds in (15), we can incorporate the index k achieving equality into the set \mathcal{C}_{on} . Equation (10) still holds since $u_k = z(k)/\mu$. Then μ can be reduced further using (10). The sets \mathcal{C}_{on} and \mathcal{C}_{off} are modified in the following manner.

1. If reducing μ causes a block component u_k with $k \in \mathcal{C}_{\text{off}}$ to achieve unity norm, then enter the k^{th} block into \mathcal{C}_{on} and reduce \mathcal{C}_{off} accordingly.
2. If reducing μ causes the k^{th} block component of x_{on} to vanish, then add the k^{th} block to \mathcal{C}_{off} and modify \mathcal{C}_{on} accordingly.

2.1. A Simplified Approximation

The steps evaluated below (for the case when the block size is greater than unity) assumes $u_{\text{on}}(\mu)$ is piecewise constant between branches. Only rule 1 applies during the homotopy. We justify this by noting that rule 2 involves finding a zeroed block component of x_{on} using the first equations of (10), assuming a constant u_{on} .

The new nonzero component u_k at each branch is determined by (10); once activated, this block component remains fixed throughout the algorithm. Defining E_k to be a matrix

whose orthonormal columns span the entries associated with the k^{th} block, and

$$\alpha \stackrel{\text{def}}{=} z_{\text{off}} - B^T A^{-1} z_{\text{on}}, \quad \beta \stackrel{\text{def}}{=} B^T A^{-1} u_{\text{on}},$$

the norm crossing of rule 1 is expressed, in the k^{th} block component of u_{off} , by the equality

$$\|\mu\beta_k + \alpha_k\|_2^2 = \|E_k^T(\mu\beta + \alpha)\|_2^2 = \mu^2 \|E_k^T u_{\text{off}}\|_2^2 = \mu^2 \|u_k\|_2^2 = \mu^2,$$

where $\alpha_k \stackrel{\text{def}}{=} E_k^T \alpha$ and $\beta_k \stackrel{\text{def}}{=} E_k^T \beta$. This quadratic in μ is solved for each block component using

$$\mu = \frac{-2\alpha_k^T \beta_k \pm \sqrt{4(\alpha_k^T \beta_k)^2 - 4\|\alpha_k\|_2^2(\|\beta_k\|_2^2 - 1)}}{2(\|\beta_k\|_2^2 - 1)}. \quad (16)$$

The component crossing unity norm first (largest μ) is activated. Pseudocode for the approximation is given in Algorithm 1. This method differs from another approximation to group LASSO [6] in its solution path as well as its final result.

Algorithm 1 APPROXIMATEBLOCKHOMOTOPY

Input: $\Phi \in \mathbb{R}^{MP \times NP}$, $y \in \mathbb{R}^{MP \times 1}$, i_{max} , ε_t

$\hat{k} \leftarrow \arg\max_k \|E_k^T \Phi^T y\|_2$, $\hat{\mu} \leftarrow \|E_{\hat{k}}^T \Phi^T y\|_2$, $x \leftarrow 0$

$u_{\hat{k}} \leftarrow E_{\hat{k}}^T \Phi^T y / \|E_{\hat{k}}^T \Phi^T y\|_2$, $i \leftarrow 1$

$\mathcal{C}_{\text{on}} \leftarrow \{\hat{k}\}$, $\mathcal{C}_{\text{off}} \leftarrow \{m \in \mathbb{Z} | 0 < m \leq N, m \neq \hat{k}\}$

while $\|y - \Phi x\|_2^2 + \hat{\mu} \|x\|_1 > \varepsilon_t$ **and** $i \leq i_{\text{max}}$ **do**

$A \leftarrow 2\Phi_{\text{on}}^T \Phi_{\text{on}}$, $B \leftarrow 2\Phi_{\text{on}}^T \Phi_{\text{off}}$

$z_{\text{on}} \leftarrow 2\Phi_{\text{on}}^T y$, $z_{\text{off}} \leftarrow 2\Phi_{\text{off}}^T y$

$\alpha \leftarrow z_{\text{off}} - B^T A^{-1} z_{\text{on}}$, $\beta \leftarrow B^T A^{-1} u_{\text{on}}$

for all $k \in \mathcal{C}_{\text{off}}$ **do**

$\mu_k \leftarrow \sup\{\mu | 0 < \mu < \hat{\mu}, \|\alpha_k + \beta_k \mu\|_2^2 = 1\}$

end for

$\hat{k} \leftarrow \arg\max_k \mu_k$, $\hat{\mu} \leftarrow \mu_{\hat{k}}$, $x_{\text{on}} \leftarrow A^{-1}(z_{\text{on}} - \hat{\mu} u_{\text{on}})$

$u_{\hat{k}} \leftarrow (\alpha_{\hat{k}} / \hat{\mu} + \beta_{\hat{k}}) / \|\alpha_{\hat{k}} / \hat{\mu} + \beta_{\hat{k}}\|_2$, $i \leftarrow i + 1$

$\mathcal{C}_{\text{on}} \leftarrow \mathcal{C}_{\text{on}} \cup \{\hat{k}\}$, $\mathcal{C}_{\text{off}} \leftarrow \mathcal{C}_{\text{off}} \setminus \{\hat{k}\}$

end while

return \mathcal{C}_{on}

3. PERFORMANCE

Monte Carlo simulations were run comparing the detection (dictionary column identification) performance of exact and approximate block partitioned homotopy processing, as well as S-OMP [7]. The exact homotopy algorithm was implemented using `ode45` in Matlab. The signals used were sparse in a Fourier basis, where $\Phi = \Theta W_{\text{inv}} \in \mathbb{C}^{100 \times 500}$, with $W_{\text{inv}} \in \mathbb{C}^{500 \times 500}$ representing an inverse DFT matrix, and $\Theta \in \mathbb{R}^{100 \times 500}$ a matrix whose elements were drawn from a zero mean unit variance Gaussian distribution. The simulations consisted of 1000 runs at sparsity levels ranging from

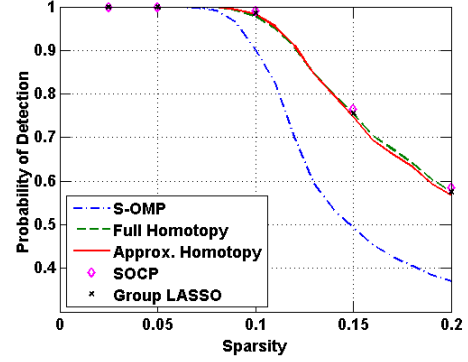


Fig. 1. Comparison of detection performance using various sparse reconstruction techniques where the signals are down-sampled by a factor of 5. In all cases the SNR is 30 dB.

1% to 20% in 1% steps. In all cases, signals were blocked in groups of $P = 5$ (e.g., a 5-sensor array). Approximate block homotopy matched the performance of the exact algorithm and outperformed S-OMP, as illustrated in Fig. 1.

We also compared the performance of approximate block partitioned homotopy to that of SOCP using

$$\min_x \sum_{k=1}^P \|x(k)\|_2 \quad \text{s.t.} \quad \|y - \Phi x\|_2 \leq \epsilon, \quad (17)$$

and group LASSO with the block shrinkage function

$$S_{\mu}(x(k)) = \begin{cases} x(k) - \frac{\mu}{2} \frac{x(k)}{\|x(k)\|_2} & \text{if } \|x(k)\|_2 > \frac{\mu}{2} \\ 0 & \text{if } \|x(k)\|_2 \leq \frac{\mu}{2} \end{cases}, \quad (18)$$

replacing the shrinkage function in [8]. In the cases simulated, the identification performance of block partitioned homotopy processing was virtually identical to that of SOCP and group LASSO for the discrete sparsity values of 2.5%, 5%, 10%, 15% and 20%; however, there was a performance advantage to block partitioned homotopy processing discussed in Section 3.2. Group CoSaMP [9] was not used in the comparison since it requires $3 \times$ as many columns, and thus did not detect signals beyond 7% sparsity.

3.1. Array Processing Application Example

We simulated a CS receiver with $P = 4$ antennas operating in an environment with RF emitters, including hoppers, in the 2–6 GHz frequency range. The 4-antenna linear array was spaced in 2.5 cm increments, with the response of each antenna constant (unity gain) in azimuth. Compressive measurements in a sensor-frequency basis were obtained via random sampling, with $\Phi \in \mathbb{R}^{MP \times NP}$, where $M = 64$ and $N = 512$ (see [4] for a full description of the simulated scenario). As shown in Fig. 2, approximate block homotopy detection performance was noticeably better than that of S-OMP at all SNRs and sparsity levels measured.

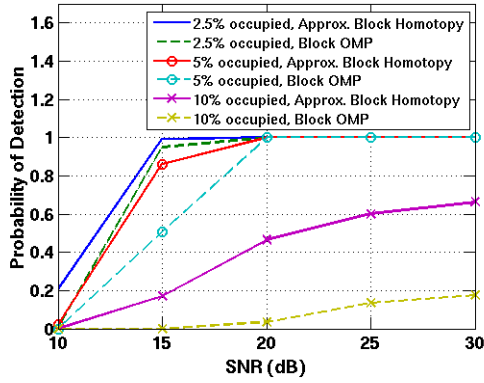


Fig. 2. Comparison of detection performance using approximate block partitioned homotopy continuation and S-OMP in an array processing application. Each of 4 channels is randomly down-sampled by a factor of 8.

3.2. Computational Complexity

The computationally taxing operations of approximate block partitioned homotopy includes the inversion of the matrix A from (10). In block partitioned form at iteration k , the matrix A is updated from iteration $k - 1$ as

$$A_k = \begin{bmatrix} A_{k-1} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}, \quad (19)$$

where A_{ij} for $i, j \in \{1, 2\}$ corresponds to the elements from $2\Phi^T\Phi$ added to A when $x(k)$ is added to \mathcal{C}_{on} . The inverse of this matrix in block partitioned form is given by

$$A_k^{-1} = \begin{bmatrix} A_{k-1}^{-1} + A_{k-1}^{-1}A_{12}S^{-1}A_{21}A_{k-1}^{-1} & A_{k-1}^{-1}A_{12}S^{-1} \\ -S^{-1}A_{21}A_{k-1}^{-1} & S^{-1} \end{bmatrix}, \quad (20)$$

where $S = A_{22} - A_{21}A_{11}^{-1}A_{12}$ [10]. The dominant operation in (20) is the multiplication $A_{k-1}^{-1}A_{12}S^{-1}A_{21}A_{k-1}^{-1}$, which requires $O(P^3k^2)$ operations, and summing through K iterations yields $O(P^3K^3)$. Computing $z = \Phi^Ty$ requires $O(MNP^2)$ operations, while the computation of $2\Phi^T\Phi$ requires $O(MNKP^3)$ operations to obtain the matrices A and B . In some cases, however, there may be sufficient memory to store $2\Phi^T\Phi$ for all realizations of Φ .

S-OMP's complexity is dominated by front end correlations requiring $O(P^2KMN)$ operations after K iterations, given efficient rank-1 updates are used to compute the projector used in reconstruction. SOCP has complexity $O((KN)^3)$ [11], and in general, group LASSO approaches to solving (1) require iteratively searching for a μ or ϵ to achieve the desired level of sparsity.

4. SUMMARY

In this paper we present block partitioned homotopy processing for multichannel sparse signal reconstruction. We ex-

tend the homotopy continuation [5] to basis pursuit optimization with a block partitioned sparsity constraint that consists of the ℓ_2 norm of each row of the signal matrix combined in an ℓ_1 fashion. Block partitioned homotopy processing outperformed S-OMP in identifying the columns vectors of the dictionary spanned by signal, and did so with roughly the same computational complexity as S-OMP. Block partitioned homotopy processing performance was on par with SOCP, but with significantly lower computational complexity.

5. REFERENCES

- [1] E. Candes, J. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Comm. Pure Appl. Math.*, vol. 59, no. 8, pp. 1207–1233, 2006.
- [2] R. Baraniuk, "Compressive sensing," *IEEE Signal Process. Mag.*, vol. 24, no. 4, pp. 118–120, 2007.
- [3] D. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.
- [4] B. Miller, J. Goodman, K. Forsythe, J. Sun, and V. Goyal, "A multi-sensor compressed sensing receiver: Performance bounds and simulated results," *Proc. 43rd Asilomar Conf. Signals, Syst. and Comput.*, pp. 1571–1575, 2009.
- [5] D. Malioutov, M. Cetin, and A. Wilsky, "Homotopy continuation for sparse signal representation," in *Proc. ICASSP*, 2005, vol. 5.
- [6] M. Yuan and Y. Lin, "Model selection and estimation in regression with grouped variables," Tech. Rep. 1095, University of Wisconsin, Nov 2004.
- [7] J. Tropp, A.C. Gilbert, and M.J. Strauss, "Simultaneous sparse approximation via greedy pursuit," *Proc. ICASSP*, Mar 2005.
- [8] I. Daubechies, M. Defrise, and C. De Mol, "An iterative thresholding algorithm for linear inverse problems with sparsity constraint," *Comm. Pure Appl. Math.*, vol. LVII, pp. 1413–1457, 2004.
- [9] D. Needell and J. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," *Applied and Computational Harmonic Analysis*, vol. 26, no. 3, pp. 301–321, 2009.
- [10] G. Golub and C. Van Loan, *Matrix Computations*, Johns Hopkins University Press, Baltimore, Maryland, 1983.
- [11] D. Malioutov, M. Cetin, and A. Wilsky, "Sparse signal reconstruction perspective for source localization with sensor arrays," *IEEE Trans. Signal Process.*, pp. 3010–3022, 2005.