A Non-Iterative Procedure for Computing Sparse and Sparsifiable Solutions to Slightly Underdetermined Linear Systems of Equations

Andrew E. Yagle
Department of EECS, The University of Michigan, Ann Arbor, MI 48109-2122

Abstract—**The problem of computing sparse (mostly zero) solutions to underdetermined linear systems of equations has received much attention recently, due to its applications to compressed sensing. Under mild assumptions, the sparsest solution has minimum-L1-norm, and can be computed using linear programming. In some applications (valid deconvolution, singular linear transformations), the linear system is underdetermined by a relatively small amount, and a simpler solution is desirable. This paper presents a closed-form solution for computing the K-sparse solution to an M-by-N underdetermined linear system of equations, if N exceeds (K+1)(N-M+1). A numerical example and program illustrates the new algorithm.**

Keywords— **Sparse reconstruction**
Phone: 734-763-9810. Fax: 734-763-1503.
Email: aey@eecs.umich.edu. **EDICS: 2-REST.**

I. INTRODUCTION

A. Problem Statement

The goal is to solve underdetermined linear system

$$y = \tilde{H}x = \tilde{H}W'Wx = Hz; H = \tilde{H}W'; z = Wx \quad (1)$$

- $\tilde{H}$ and $H=\tilde{H}W'$ are rank=M and M×N,N>M;
- W is an N×N unitary transformation matrix that sparsifies the unknown column N-vector x;
- z has at most K out of N non-zero elements;
- $y, \tilde{H}, W, H, K, M, N$ are known; x, z unknown.
- N≥(K+1)(N-M+1).

Examples of the sparsifying transformation W include the wavelet transform and gradient edge detectors. Of course, if the unknown x is itself known to be sparse, as in X-ray crystallography, then $W = I$.

The condition N≥(K+1)(N-M+1) makes this problem inappropriate for compressed sensing. Three applications satisfying this condition are:

- Valid (no signal support information is used) deconvolution problems. The valid convolution of:
  - Unknown signal $x[n]$ of length M with known
  - Impulse response (1-D) or point-spread (2-D) function $h[n]$ of length L gives the observed data:
  - $y[n] = h[n] * x[n]$ of length M+L-1. However, only M-L+1 of M+L-1 values of $y[n]$ are known.
- Linear transformation inverse problems $y = Hx$ in which H is effectively rank-deficient.
- Reconstruction problems $y = Hx$ in which some of the data y are missing or garbled.

We note that if H is substantially underdetermined $N >> M$, but each row is known to be bandlimited to $\pi\frac{M}{N}$, then our non-iterative algorithm [1] can be used. We have also used related but different approaches to X-ray crystallography [2] and to the limited-angle tomography formulation of SAR [3].

B. Relevant Previous Approaches

Use of sparseness as side information in signal reconstruction goes back at least as far as 1979 [4]. Most early work considered the deconvolution of sparse 1D spike trains arising in reflections from layered media in seismic exploration, although [5] considered 1D reconstruction from bandlimited data. The first use of sparseness in 2D (image) reconstruction of which we are aware is [6]. All of this early work minimized the ℓ_1 norm (sum of absolute values) of the signal, using linear programming. The idea was that the ℓ_1 norm solution lies on a vertex of the simplex and so is sparse. This idea has recently been put on a firmer theoretical ground in [7] and other recent papers. More recent work using this approach is [8]-[10]. Using the ℓ_1 norm for the signal and ℓ_2 for the error is called LASSO. The problem with this approach is the amount of computation required by linear programming for image reconstruction.

Another way of formulating the linear sparse reconstruction problem is as a matrix "subset selection" problem [11]-[14]. The forward greedy algorithm successively selects the matrix column closest (in the mean-square sense) to the residual error resulting from the previous matrix column selections. The backward greedy algorithm starts with a general solution and successively removes the matrix column that increases the mean-square residual error the least. The latter algorithm has been shown to give the correct answer if the noise level is sufficiently small [13]. However, again the problem is the amount of computation required here by subset selection.

Still another recent approach is to include a thresholding constraint in a Landweber-like iterative algorithm [15]-[18], often arising from statistical image priors that implicitly (but not explicitly) maximize sparsity. This is a straightforward approach, and unlike the above methods requires reasonable computation for 2-D problems. However, even in the absence of noise, convergence to an optimally-sparse solution is not in general guaranteed for these algorithms.

C. New Approach of This Paper

All of the above approaches are iterative, requiring many iterations, are computationally intensive, or both. The approach of this paper requires only:

- The right nullspace of H and solution to (1);
- Solution of N×(K+1)(N-M+1) linear system;
- Rank-one factored (K+1)×(N-M+1) matrix;
- An N-point FFT to find nonzero locations of z;
- Solution of an N×(K+1) linear system to find z;
- Computation of $x = W'z$ (if x is not sparse).

Note that all of these operations are straightforward linear algebra and can be implemented in a non-iterative way. Of course, iterative methods such as conjugate gradient may be used in some steps, but the overall algorithm is non-iterative, and can even be considered to be a closed-form computation.

II. Derivation of New Algorithm

Let G be an N×(N-M) matrix of (not necessarily orthogonal) vectors spanning the right nullspace of H. Let $\tilde{z}$ be the minimum ℓ_2 norm solution to (1):

$$HG = 0; G = [g^1|\dots|g^{N-M}]; \tilde{z} = H'(HH')^{-1}y. \quad (2)$$

Then the desired solution z can be written as

$$z = \tilde{z} + \sum_{i=1}^{N-M} e_i g^i = \sum_{i=0}^{N-M} e_i g^i \quad (3)$$

for some unknown constants $\{e_i\}$. We define $e_0 g^0 = \tilde{z}$ for notational convenience. Computing the N-point discrete Fourier transform of each column and using the results as first rows of circulant matrices gives

$$\hat{Z} = \sum_{i=0}^{N-M} e_i \hat{G}^i \quad (4)$$

where $\hat{g}^i_k = \sum_{n=0}^{N-1} g^i_n e^{-j2\pi nk/N}, 0 \le k \le N-1$ and

$$\hat{Z} = \begin{bmatrix} \hat{z}_0 & \cdots & \hat{z}_{N-1} \\ \ddots & \ddots & \ddots \\ \hat{z}^*_{N-1} & \cdots & \hat{z}_0 \end{bmatrix} \quad (5)$$

$$\hat{G}^i = \begin{bmatrix} \hat{g}^i_0 & \cdots & \hat{g}^i_{N-1} \\ \ddots & \ddots & \ddots \\ {\hat{g}^i_{N-1}}^* & \cdots & \hat{g}^i_0 \end{bmatrix} \quad (6)$$

using the conjugate symmetry relation $\hat{g}^i_{N-k} = {\hat{g}^i_k}^*$.

The eigenvalues of each circulant matrix $\hat{G}^i$ are the elements $\{g^i_n\}$ of g^i. Because only K elements of z are nonzero, $\hat{Z}$ has rank K, and there exists a vector v of length K+1 such that

$$\hat{Z}\begin{bmatrix} v \\ 0 \end{bmatrix} = \sum_{i=0}^{N-M} e_i \hat{G}^i \begin{bmatrix} v \\ 0 \end{bmatrix} = 0 \quad (7)$$

Let $\hat{\tilde{G}}^i$ be the matrix of the first K+1 columns of $\hat{G}^i$. Each $\hat{\tilde{G}}^i$ is Toeplitz. Rewrite this equation as

$$\begin{bmatrix} \hat{\tilde{G}}^0 & \dots & \hat{\tilde{G}}^{N-M} \end{bmatrix} \begin{bmatrix} e_0 v \\ \vdots \\ e_{N-M} v \end{bmatrix} = 0 \quad (8)$$

This is a system of N equations in a total of (N-M+1) unknowns $\{e_i\}$ and (K+1) unknowns $\{v_i\}$.

- If N>(N-M+1)+(K+1) then this is an overdetermined system of quadratic equations, which by Bezout's theorem almost surely has only the actual solution assumed to exist. But its solution is difficult.
- If N>(N-M+1)·(K+1) then this is an overdetermined system of *linear* equations. Solution is easy.

We can solve for the (N-M+1)(K+1) unknowns $\{e_i v_j\}$. Arranging these into an (N-M+1)×(K+1) matrix and computing its rank-one decomposition yields $\{e_i\}$ and $\{v_j\}$ to a (irrelevant) scale factor.

Next, make the following definitions:

$$\{z_{i_n}, 1 \le n \le K\} = \{z_i : z_i \ne 0, 1 \le i \le N\} \quad (9)$$

$$[F_{ik}] = [e^{-j2\pi(i-1)(k-1)/N}]_{ik}, 1 \le i,k \le N \quad (10)$$

$$[\tilde{F}_{ik}] = [e^{-j2\pi(i-1)(n_k-1)/N}]_{ik}, 1 \le k \le K \quad (11)$$

- $\{z_{i_n}, 1 \le n \le K\}$ are the nonzero values of $\{z_i\}$;
- F is the DFT matrix implementing $\hat{Z} = Fz$;
- $\tilde{F}$ is a tall N×K submatrix of F.

Then we have the (N×K)(K×N) factorization

$$\hat{Z} = F\text{diag}[z_i]F^H = \tilde{F}\text{diag}[z_{i_n}]\tilde{F}^H \tag{12}$$

Extending $\tilde{F}$ by any other column of F gives

$$0 = \hat{Z}\begin{bmatrix} v \\ 0 \end{bmatrix} = \tilde{F}\text{diag}[z_{i_n}\ 0]\tilde{F}^H\begin{bmatrix} v \\ 0 \end{bmatrix} \tag{13}$$

Since $\tilde{F}$ is a tall matrix and $z_{i_n} \neq 0$, we have

$$\tilde{F}^H\begin{bmatrix} v \\ 0 \end{bmatrix} = 0 \rightarrow F^H\begin{bmatrix} v \\ 0 \end{bmatrix} = 0 \,\text{for}\, i \in \{i_n\} \tag{14}$$

This shows that an inverse N-point DFT of a zero-padded v is zero at the locations $\{i_n\}$ of nonzero z_i. The conditioning of the location problem is determined by the condition number of $\tilde{F}$. Random distribution of $\{i_n\}$ yields good conditioning; clumping of $\{i_n\}$ and large gaps yields poor conditioning. The DFT of null vector of Toeplitz-structured matrix can be seen as a deterministic version of MUSIC, without computing the autocorrelation first.

III. Numerical Example

A 30×30 block image x_{ij} was sparsified by cyclic convolution with the corner detector:

$$z_{ij} = \sum_{m=0}^{29}\sum_{n=0}^{29} x_{mn}h_{i-m,j-n}, h_{ij} = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \tag{15}$$

z_{ij} was then unwrapped by rows to a column vector z_n and multiplied by a random 832×900 matrix $\tilde{H}$:

$$y_i = \sum_{n=1}^{900} \tilde{H}_{in}z_n, \quad 1 \leq i \leq 832 \tag{16}$$

Given knowledge of $\tilde{H}$ and that the sparsified image z_{ij} has only 12 out of 900 nonzero elements, the goal is to reconstruct x_{ij} from y_i (832 observations).

We have $K = 12, M = 832, N = 900$. Since

$$N = 900 > 897 = (12 + 1)(900 - 832 + 1) \tag{17}$$

we can use the method of this paper. The 900×897 matrix is indeed singular, with singular values

$$\sigma_1 = 389; \sigma_{896} = 0.080; \sigma_{897} = 1.8 \times 10^{-14} \tag{18}$$

The null vector of this matrix was arranged into a 13×69 rank-one matrix having singular values

$$\sigma_1 = 1.00; \sigma_2 = 6 \times 10^{-14} \tag{19}$$

The discrete Fourier transform of order=900 of the length=13 factor of the rank-one decomposition was computed. Their magnitudes were ordered from smallest to largest. The 12^{th} and 13^{th} smallest were

$$p_{12} = 1.23 \times 10^{-13}; p_{13} = 0.0085 \tag{20}$$

a sharp threshold confirming the presence of only 12 nonzero elements of z_n, which was then rewrapped to the sparsified image z_{ij}.

The corner detector was then deconvolved from z_{ij} to obtain x_{ij}. The corner detector frequency response is zero along the 2-D frequency axes, so side information x_{i1}=x_{1j}=0 was used to fill in those values.

Figure #1 shows the minimum-norm-least-squares solution to $y = \tilde{H}z$. This is clearly insufficient for determining the nonzero elements of z_{ij}. Figure #2 shows the reconstruction of the indicator function for nonzero z_{ij}, found by setting the 12 smallest values to one and the other 900–12 values to zero. As noted above, there is a sharp threshold. Figure #3 shows the reconstruction of the original image x_{ij}.

This demonstrates that the algorithm works with good conditioning and sharp thresholds throughout. The Matlab code used to generate this example is given at the end of the paper in an Appendix.

IV. Conclusion

We have presented a non-iterative algorithm for reconstructing sparse solutions to slightly underdetermined systems of linear equations. The algorithm requires only simple linear algebra operations. A numerical example demonstrates the algorithm in reconstruction of a simple, sparsifiable image. If the underdetermination is not small, but matrix rows are bandlimited, our algorithm of [1] can be used.

References

[1] A.E. Yagle, "A Non-Iterative Procedure for Sparse Solutions to Linear Equations with Bandlimited Rows," 2008.

[2] A.E. Yagle, "New Atomicity-Exploiting Algorithms for Super-Resolution X-Ray Crystallography," 2008.

[3] A.E. Yagle, "Limited Angle Tomography of Sparse Images from Noisy Data using TLS MUSIC," 2008.

[4] H. Taylor, S. Banks, F. McCoy, "Deconvolution with the L_1 norm," *Geophysics* 44, 39-52, 1979.

[5] F. Santosa and W.W. Symes, "Linear inversion of bandlimited reflection seismograms," *SIAM J. Sci. Stat. Comp.* 7, 1307-1330, 1986.

[6] D.C. Dobson and F. Santosa, "Recovery of blocky images from noisy and blurred data," *SIAM J. Appl. Math* 56, 1181-1198, 1996.

[7] D. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ^1 minimization," *Proc. Nat. Acad. Sci.* 100(5), 2197-2202, 2003.

[8] K. Tsuda and G. Ratsch, "Image reconstruction by linear programming," *IEEE Trans. IP* 14, 737-744, 2005.

[9] E. Candes and J. Romberg, *Inverse Problems* June 2007.

[10] J. Haupt and R. Nowak, "Signal reconstruction from noisy random projections," *IEEE Trans. Info. Th.* 52, 4036-4048, 2006.

[11] G. Harikumar and Y. Bresler, "A new algorithm for computing sparse solutions to linear inverse problems," Proc. ICASSP 1996, 1331-1334.

[12] G. Harikumar, C. Couvreur, Y. Bresler, "Fast optimal and suboptimal algorithms for sparse solutions to linear inverse problems," Proc. ICASSP 1998, 1877-1880.

[13] C. Couvreur and Y. Bresler, "On the optimality of the backward greedy algorithm for the subset selection problem," *SIAM J. Matrix Anal. and Appl.* 21, 797-808, 2000.

[14] F.R. de Hoog and R.M.M. Mattheij, "Subset selection for matrices," *Linear Algebra and its Applications* 422, 349-359, 2007.

[15] B.K. Natarajan, "Sparse approximate solutions to linear systems," *SIAM J. Computers* 24, 227-234, 1995.

[16] M.A. Figueiredo and R.D. Nowak, "An EM algorithm for wavelet based image restoration," *IEEE Trans. Image Proc.* 12, 906-916, 2003.

[17] I. Daubechies, M. Defrise, C. de Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Comm. Pure Appl. Math* 57, 1413-1457, 2004.

[18] K.K. Herrity, A.C. Gilbert, J.A. Tropp, "Sparse approximation via iterative thresholding," Proc. ICASSP 2006.

V. APPENDIX

```
%GOAL: Reconstruct sparsifiable X from Y=H*X.
%SIZES: X and W:30X30(900). H:832X900. Y:832X1.
clear;rand(’seed’,0);X(30,30)=0;H=rand(832,900);
%Create the 30X30 sparsifiable block image X:
X(2:7,2:27)=ones(6,26);X(12:17,2:27)=ones(6,26);
X(2:27,2:7)=ones(26,6);X(22:27,2:27)=ones(6,26);
%Sparsify the image X to W with corner detector:
W=real(ifft2(fft2(X).*fft2([1 -1;-1 1],30,30)));
%Reconstruct sparse W from Y=H*W(:)(could be H*X):
W=W’;Y=H*W(:);N=null(H);N(:,69)=H\Y;A=[];FN=fft(N);
%Form 900X897 singular Toeplitz blocks matrix:
for I=1:69;A=[A toeplitz(FN(:,I),FN(1:13,I)’)];end
%Find null vector of A and its rank 1 decomposition:
[U S V]=svd(A);[U1 S1 V1]=svd(reshape(V(:,897),13,69));
%Compute DFT and note sharp threshold → #nonzero W:
P=abs(fft(U1(:,1),900));P1=sort(P);P1(12),P1(13)
%Make Q binary & compute values sparse W from data Y:
Q=reshape(flipud(P),30,30);Q(Q<.001)=0;Q=1-sign(Q);
Z1(900)=0;Q1=Q(:);I=find(Q1>0.001)+1;Z1(I)=H(:,I)\Y;
%Inverse filter corner detector & set values where 0:
FW=fft2(reshape(Z1,30,30))./fft2([1 -1;-1 1],30,30);
FW(1,:)=-sum(FW(2:30,:));FW=FW.’;FW(1,2:30)=-sum(FW(2:30,2:30));
FW(1,1)=sum(sum(X));WHAT=real(ifft2(FW));imagesc(WHAT)
```