Limited Angle Tomography of Sparse Images from Noisy Data using TLS MUSIC Algorithm

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

Abstract— **The limited angle tomography problem is to reconstruct an image from a set of its projections (Radon transform) over a limited range of angles. It has applications in medical imaging and synthetic aperture radar. From the Radon projection-slice theorem, this problem is equivalent to image reconstruction from partial Fourier data. This paper presents an algorithm for solving this problem for sparse images. The fraction of nonzero pixels must be one-fourth the fraction of Fourier data available or less; their locations are unknown; and there is no support constraint. The problem requires only solution of two Toeplitz-mosaic-Toeplitz (TMT) linear systems of equations and a 2D FFT. For noisy data, the minimum singular vector of a TMT matrix is required; this can be computed using inverse power method (TLS). Numerical examples illustrate the new algorithms.**

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

I. INTRODUCTION

A. Problem Statement

The limited angle tomography problem is to reconstruct an image $x(y,z)$ from its projections $p(t,\theta)$

$$p(t,\theta)=\int\int x(y,z)\delta(t-y\cos\theta-z\sin\theta)dy\,dz \quad (1)$$

over a narrow range of angles θ. Using the projection-slice theorem [1], this is equivalent to reconstructing the image from the partial Fourier data in a bowtie-shaped region. This has applications to medical imaging and non-destructive evaluation where projection data is available over a narrow range of angles, due to physical constraints.

In synthetic aperture radar, the projection data are also bandlimited in t and the image is complex. Then Fourier data is only known in a segment of an annulus in the Fourier plane [2]. Multistatic, spotlight-mode, and strip-map synthetic aperture radar (and holography) can all be formulated in this way.

In practice, the problem is discretized. If the image is bandlimited and has finite extent then the continuous 2D Fourier transform can be replaced with the 2D discrete Fourier transform. This should not be confused with assuming that the image has compact support *within* the 2D DFT. This is not only often untrue, but leads to an extremely ill-posed reconstruction problem. By using different sampling rates along different projection angles, rectangular 2D sampling can be used without interpolation.

This paper solves the following problem: Reconstruct a sparse image from its 2D DFT known only in a trapezoidal region of the 2D DFT Fourier plane. The trapezoidal region closely approximates an annular segment, and includes the bowtie region as a special case. The image is known to be sparse: the number of nonzero pixels does not exceed one-fourth the number of known DFT values. The locations of nonzero pixels are unknown, no compact support constraint *within* the 2D DFT is available, and the image is allowed to have complex-valued pixels.

B. Relevant Previous Approaches

Most approaches to limited angle tomography assume the image has compact support *within* the 2D DFT, so that the number of nonzero pixels does not exceed the number of known 2D DFT values in the bowtie-shaped region of the Fourier plane. Then the problem becomes a linear system of equations, which can be solved using conjugate gradient or alternating projections that alternately impose the known Fourier values and compact support constraints. However, this linear system of equations is extremely ill-posed. Hence some regularization is necessary, such as Tikhonov-regularized least-squares or the edge-preserving regularization used in [3]. However, any regularization introduces bias into the reconstruction, so that the solution is incorrect even in the noiseless case. We will not attempt to survey literature on limited angle tomography.

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 2D 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

This paper presents a new algorithm that uses deterministic side information that the image is sparse. This does not introduce bias, so the solution is correct in the absence of noise. The algorithm is very fast, requiring only solution of two Toeplitz-mosaic-Toeplitz (TMT) systems of equations (for noisy data, computation of of a TMT minimum singular vector).

First, the null vector (noiseless data) or minimum singular vector (noisy data) of a TMT matrix constructed from the Fourier data is computed. Second, this vector is unwrapped into a 2D signal with trapezoidal support, and its 2D DFT computed. The smallest magnitudes of this 2D DFT provide the locations of nonzero pixels. Since nonzero pixels can only occur at integer coordinates, there is some error correction capability for low noise levels. Finally, the nonzero pixel values are computed by solving another overdetermined linear system of equations.

II. Noiseless Solution Procedure

A. Problem Formulation

We make the following assumptions:

- x_{ij} is defined on an M×M rectangular lattice;
- We have fM^2 of M^2 values X_{mn} of the DFT of x_{ij} in a trapezoidal region of the Fourier plane;
- f=fraction of the 2D DFT values available. This includes both a limited range of projection angles θ and bandlimitation in t of projections $p(t,\theta)$;
- $K = fM^2/4 - 1$ or fewer of the x_{ij} are nonzero;
- The locations $\{(i_n, j_n), 1 \le n \le K\}$ of the nonzero image pixels are all unknown;
- The known 2D DFT values X_{mn} include an additive zero-mean white Gaussian noise random field.

We start our indexing at $n = 1$ instead of $n = 0$ for convenience in matrix indexing. The (M×M)-point 2D DFT X_{mn} of x_{ij} for $1 \le m, n \le M$ is

$$X_{mn} = \sum_{i=1}^{M}\sum_{j=1}^{M} x_{ij} e^{-j\frac{2\pi}{M^2}[(i-1)(m-1)+(j-1)(n-1)]}. \quad (2)$$

B. DFT-Based Derivation of Algorithm

Let s_{ij} be the 2D function having 2D DFT S_{mn}:

- S_{mn} has compact support of K+1 pixels shaped similarly to the trapezoidal support of the known X_{mn} but 1/4 the size (1/2 in linear dimension);
- s_{ij} is an indicator function for nonzero x_{ij}:

$$\begin{cases} s_{ij} = 0 & \text{if } x_{ij} \neq 0; \\ s_{ij} \neq 0 & \text{if } x_{ij} = 0 \end{cases}. \quad (3)$$

Then ($**$ denotes 2D convolution)

$$x_{ij}s_{ij} = 0 \to X_{mn} ** S_{mn} = 0. \quad (4)$$

Since there are only K nonzero values of x_{ij}, s_{ij} is uniquely determined, to an (irrelevant) scale factor. The support of its 2D DFT S_{mn} looks like this:

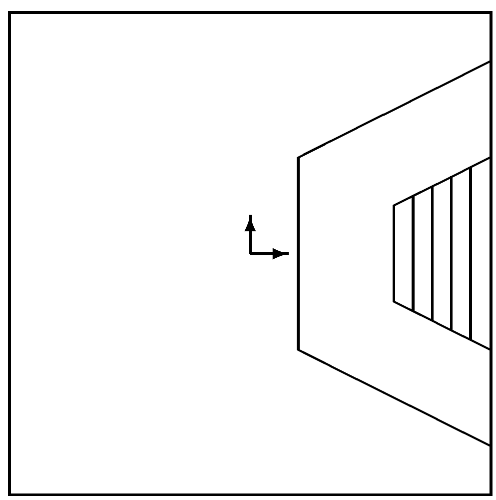

Fig. 1: Supports of S_{mn} and known X_{mn}.

The second of (4) can be written as a Toeplitz-mosaic-Toeplitz (TMT) linear system of equations for the unknown S_{mn}. Its structure looks like this:

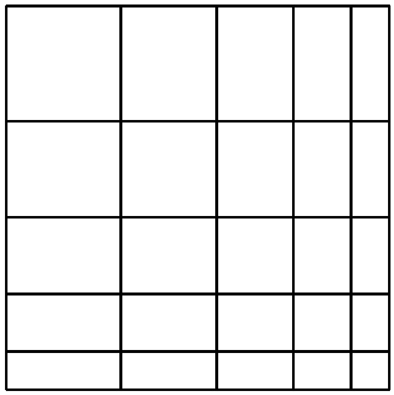

Fig. 2: A TMT structure. Each submatrix is itself Toeplitz, in block Toeplitz structure.

Each submatrix is itself Toeplitz and there is an overall block Toeplitz structure. This TMT structure is derived as a submatrix of a circulant-block-circulant matrix in the alternate derivation to follow.

S_{mn} are the elements of the null vector of this TMT matrix constructed from the known X_{mn}. Then an inverse 2D DFT computes s_{ij}, which is zero at the locations of the nonzero x_{ij}:

$$s_{i_n,j_n} = 0, \quad 1 \leq n \leq K. \tag{5}$$

Once the locations of the nonzero x_{ij} are known, their values can be computed by solving a second TMT linear system of equations. This system is actually Toeplitz-block-Toeplitz (all blocks have same size).

C. Matrix Derivation of Algorithm

We now present a second derivation of the algorithm, for reasons noted below. We can write

$$X_{m_1-m_2,n_1-n_2} = \sum_{n=1}^{K} x_{i_n,j_n} e^{-j\frac{2\pi}{M^2}[(i_n-1)(m_1-m_2-1)+(j_n-1)(n_1-n_2-1)]}. \tag{6}$$

The $(M^2 \times M^2)$ circulant-block-circulant matrix [C] having for its first row $X_{m,n}$ can be factored as

$$[C] = D^H \cdot \text{DIAG}[x_{11} \cdots x_{MM}]D \tag{7}$$

where D is the Kronecker product of the DFT matrix having $(n,k)^{th}$ element $e^{-j\frac{2\pi}{M}(n-1)(k-1)}$ with itself.

Only K of the M^2 diagonal values x_{ij} are nonzero, by assumption. Let F be the submatrix of D in which all rows but those corresponding to $\{(i_n, j_n), 1 \leq n \leq K\}$ have been deleted, and all columns but the K+1 corresponding to the support of s_{ij} (defined above) have been deleted. Then the (K+1)×(K+1) TMT submatrix $[X]$ of $[C]$ can be factored as

$$[X] = F^H \cdot \text{DIAG}[x_{(i_n,j_n)}]F. \tag{8}$$

This (K+1)×(K+1) matrix clearly has rank K, so it has a null vector $\vec{a} = [a_1 \ldots a_K]'$. Postmultiplying (8) by this null vector $\vec{a}$ gives

$$[X]\vec{a} = F^H \cdot \text{DIAG}[x_{i_n,j_n}]F\vec{a} = 0 \rightarrow F\vec{a} = 0. \tag{9}$$

Since F is a submatrix of the 2D DFT matrix D, we can compute $D\vec{a}$ using a 2D FFT and see which values are zero. The rows of D corresponding to those values are the rows of F, and this identifies the locations $\{i_n, j_n\}$ of nonzero x_{ij}.

Although this derivation is more complicated than the first one, it makes two important points:

- The *conditioning* of the problem is determined primarily by the condition number of the matrix F. If a compact support constraint is used, F is extremely ill-conditioned. But if the nonzero values $x_{(i_n,j_n)}$ are (roughly) evenly spaced throughout the $M^2 \times M^2$ region, F will be fairly-well conditioned;
- If there are in fact fewer than K nonzero values of x_{i_n,j_n}, the second derivation shows that the TMT matrix $[X]$ is rank-deficient by more than one; its rank is the actual number of nonzero $x_{(i_n,j_n)}$.

III. Noisy Data Solution Procedure

A. Other Spectral Estimation Algorithms

Most spectral estimation algorithms, such as Pisarenko method, MUSIC, and ESPRIT, operate not on the data but on the autocorrelation function estimated from the data. This has the advantage that additive white noise tends to be concentrated in the subspace spanned by the singular vectors associated with the minimum singular values, since the autocorrelation of zero-mean white noise is an impulse.

However, all autocorrelation-based methods are inappropriate here, for the following three reasons:

- Only a small number of data points are available, not a long time series of data;
- Estimation of autocorrelation from data, which is always inexact due to end effects, is impractical;

• In practice, the additive noise is often neither white nor uncorrelated with the data.

Hence an approach that operates directly on the data, rather than on the autocorrelation, is necessary.

The approach used in this paper has been termed MUSIC since it is conceptually similar to MUSIC, but differs from MUSIC in these five (minor) ways:

• It operates on the data, not its autocorrelation;
• Noise is dealt with not by exploiting its (approximately) impulsive autocorrelation added to data autocorrelation, but by perturbing data directly;
• The finite number of possible locations introduces error correction to nearest for small noise levels;
• The Fourier data is in an offset trapezoidal region;
• Space and wavenumber have been exchanged.

A simple likelihood function argument shows that if the noise is an additive zero-mean white Gaussian random field, then the likelihood is maximized when the given Fourier data X_{mn} are perturbed as little as possible (in the mean square or Frobenius norm sense) to make the TMT matrix drop its rank. Two major approaches are known for this problem.

The first is an iterative algorithm that alternates between the following two constraints:

• Computing the nearest (in Frobenius norm) lower rank matrix using the singular value decomposition, by subtracting the outer product of the minimum singular vectors times the minimum singular value;
• Computing the nearest (in Frobenius norm) Toeplitz matrix by averaging along the diagonals.

The other is structured total least squares, which iteratively perturbs the matrix closer to singular, while averaging diagonals to preserve structure.

Both of these approaches have been applied successfully to other problems. However, they have two problems rendering them inappropriate here:

• The TMT matrix size makes repeated computation of its singular value decomposition impractical;
• The Frobenius norm (sum of squared magnitudes of matrix elements) weights lower frequency components more than higher frequency components, since they occur more often in the TMT matrix (e.g., along main diagonals of Toeplitz submatrices).

B. Background for Noisy Data Algorithm

Consider (9) with the noiseless $\{X_{mn}\}$ replaced with noisy $\{X_{mn}\}$. Then $[X]\vec{a} \neq 0$. What happens now is that $F\vec{a}$ computes the 2D DFT of (unwrapped) $\vec{a}$, "filters" it with $\{x_{ij}\}$, and then F^H computes the inverse 2D DFT of the result. In the noiseless case, only K of the $\{x_{ij}\}$ are nonzero, and choosing the 2D DFT of $\vec{a}$ to be zero at those nonzero locations makes $[X]\vec{a}$=0. In the noisy case, all of the $\{x_{ij}\}$ are nonzero, so there is no way to make $[X]\vec{a}$=0.

However, if the noise level is not too high, K of the $\{x_{ij}\}$ will be larger than the remaining M-K values. Heuristically, $[X]\vec{a}$ will be minimized by choosing the DFT of $\vec{a}$ to be zero at the locations of the largest $\{x_{ij}\}$ values, and nonzero elsewhere. Hence computing the value of $\vec{a}$ that minimizes $[X]\vec{a}$ (in the mean square norm) can be expected to pick out the locations of the largest $\{x_{ij}\}$, which are assumed to be the true nonzero $\{x_{ij}\}$. This value of $\vec{a}$ is the singular vector of $[X]$ associated with the minimum singular value of $[X]$. In the sequel, we refer to this as the "minimum singular vector" of the matrix $[X]$.

This will not work perfectly, of course, since the DFT of $\vec{a}$ has varying nonzero values. Hence the minimum of $[X]\vec{a}$ will be attained by a weighting of $\{x_{ij}\}$. But it can be expected that the DFT of $\vec{a}$ will be significantly smaller at the K locations of the true nonzero $\{x_{ij}\}$. Note that choosing $\vec{a}$ to be the minimum singular vector of $[X]$ also minimizes the perturbation (in Frobenius norm) of $[X]$ that makes $[X]$ drop rank, without regard to struture. Hence this is the *non-structured* total least squares (TLS) solution. TLS has been used effectively with Prony's method; we refer to our method as TLS MUSIC.

Note that the SVD of $[X]$ need not be computed– only its minimum singular vector is required. This can be computed fairly quickly using a few iterations of the inverse power method. Each iteration requires solving a linear system of equations with structured (Toeplitz or TBT) matrix $[X]$. Also note that there are only M^2 possible locations of nonzero values of $\{x_{ij}\}$, so error correction to the nearest possible location can happen if the noise is sufficiently small.

C. Noisy Data Algorithm

1. Assemble the given Fourier data X_{mn} in a trapezoidal region into the TMT matrix $[X]$;
2. Compute the minimum singular vector $\vec{a}$ of $[X]$ using a few iterations of the inverse power method;
3. Unwrap the minimum singular vector $\vec{a}$ into a trapezoidal region and compute its 2D DFT;
4. Identify the locations of the K smallest (in magnitude) values of this 2D DFT. These are the estimated locations of the nonzero values of $\{x_{ij}\}$;
5. Compute the nonzero values of $\{x_{ij}\}$ by solving an overdetermined linear system of equations.

IV. Micro-Examples

These examples illustrate the algorithm. Larger numerical simulations are given in the next section.

A. Noiseless Algorithm Micro-Example

The goal is to reconstruct x_{ij} from its (4×4)2D DFT values X_{mn} in the triangular region

$$\begin{bmatrix} * & * & * & * & -5-j \\ * & * & * & -8j & 1+7j \\ * & * & 9+3j & -4 & -1-j \\ * & * & * & 2+2j & 5-5j \\ * & * & * & * & -5-j \end{bmatrix} \quad (10)$$

where $*$ denotes an unknown value and the origin is at the center. Since this is a (4×4)2D DFT, the values in the top and bottom rows are identical, as are the values in the first and last columns. x_{ij} is sparse in that only 3 of its 16 values are nonzero.

The trapezoidal support of S_{mn} is

$$S_{mn} = \begin{bmatrix} b & 0 & 0 & 0 \\ c & a & 0 & 0 \\ d & 0 & 0 & 0 \end{bmatrix} \quad (11)$$

where again the origin is at the center and the top and bottom rows, and left and right columns, are identical. Flipping S_{mn} left-right and up-down, complex conjugating it, and sliding it over the known X_{mn} values yields the TMT linear system

$$\begin{bmatrix} -1-j & 1+7j & -5-j & -8j \\ 5-5j & -1-j & 1+7j & -4 \\ -5-j & 5-5j & -1-j & 2+2j \\ 2+2j & -4 & -8j & 9+3j \end{bmatrix} \begin{bmatrix} d^* \\ c^* \\ b^* \\ a^* \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \quad (12)$$

which has the solution

$$\begin{bmatrix} d^* \\ c^* \\ b^* \\ a^* \end{bmatrix} = \begin{bmatrix} -j \\ 1-j \\ 1-2j \\ 2 \end{bmatrix}. \quad (13)$$

The 2D DFT using Matlab notation of S_{mn} is

$$DFT\left\{ \begin{bmatrix} 1+2j & 0 & 0 & 0 \\ 1+j & 2 & 0 & 0 \\ j & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \right\} =$$

$$\begin{bmatrix} 4+4j & 2+2j & 4j & 2+6j \\ 2-2j & 0 & 2+2j & 4 \\ -2+2j & 4j & 2+2j & 0 \\ 4j & 2+2j & 0 & -2+2j \end{bmatrix}. \quad (14)$$

The three zeros indicate the locations of the nonzero x_{ij}. Solving a 3×3 linear system yields

$$x_{ij} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 2+j & 0 & 0 \\ 0 & 0 & 0 & 3+j \\ 0 & 0 & 4+j & 0 \end{bmatrix}. \quad (15)$$

References

[1] S.R. Deans, *The Radon Transform and Some of its Applications*, 1983.

[2] D.C. Munson, J.D. O'Brien, W.K. Jenkins, "A tomographic formulation of spotlight-mode synthetic aperture radar," *Proc. of the IEEE* 71, 917-925, 1983.

[3] A.H. Delaney and Y. Bresler, "Edge-preserving regularization for limited angle tomography," *IEEE Trans. Image Proc.*

[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. Image Proc.* 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.