7-2009

Automatic hard thresholding for sparse signal reconstruction from NDE measurements

Aleksandar Dogandžić
Iowa State University, ald@iastate.edu

Kun Qiu
Iowa State University

Automatic hard thresholding for sparse signal reconstruction from NDE measurements

Abstract
We propose an automatic hard thresholding (AHT) method for sparse-signal reconstruction. The measurements follow an underdetermined linear model, where the regression-coefficient vector is modeled as a superposition of an unknown deterministic sparse-signal component and a zero-mean white Gaussian component with unknown variance. Our method demands no prior knowledge about signal sparsity. Our AHT scheme approximately maximizes a generalized maximum likelihood (GML) criterion, providing an approximate GML estimate of the signal sparsity level and an empirical Bayesian estimate of the regression coefficients. We apply the proposed method to reconstruct images from sparse computerized tomography projections and compare it with existing approaches.

Keywords
nondestructive testing, numerical analysis, matrix algebra, nondestructive evaluation, QNDE

Disciplines
Materials Science and Engineering | Structures and Materials

This article appeared in *AIP Conference Proceedings* 1211 (2010): 806–813 and may be found at http://dx.doi.org/10.1063/1.3362486.

AUTOMATIC HARD THRESHOLDING FOR SPARSE SIGNAL RECONSTRUCTION FROM NDE MEASUREMENTS

Aleksandar Dogandžić and Kun Qiu

Citation: AIP Conf. Proc. **1211**, 806 (2010); doi: 10.1063/1.3362486
View online: http://dx.doi.org/10.1063/1.3362486
View Table of Contents: http://proceedings.aip.org/dbt/dbt.jsp?KEY=APCPCS&Volume=1211&Issue=1
Published by the American Institute of Physics.

AUTOMATIC HARD THRESHOLDING FOR SPARSE SIGNAL RECONSTRUCTION FROM NDE MEASUREMENTS

Aleksandar Dogandžić and Kun Qiu

Iowa State University, Center for Nondestructive Evaluation,
1915 Scholl Road, Ames, IA 50011, USA

ABSTRACT. We propose an automatic hard thresholding (AHT) method for sparse-signal reconstruction. The measurements follow an underdetermined linear model, where the regression-coefficient vector is modeled as a superposition of an unknown deterministic sparse-signal component and a zero-mean white Gaussian component with unknown variance. Our method demands no prior knowledge about signal sparsity. Our AHT scheme approximately maximizes a generalized maximum likelihood (GML) criterion, providing an approximate GML estimate of the signal sparsity level and an empirical Bayesian estimate of the regression coefficients. We apply the proposed method to reconstruct images from sparse computerized tomography projections and compare it with existing approaches.

Keywords: Compressive Sampling, Model Selection, Sparse Signal Reconstruction
PACS: 02.50.Tt Inference methods

INTRODUCTION

Sparse signal processing methods have been developed and applied to biomagnetic imaging, spectral estimation, and compressive sampling [1]–[7]. However, most sparse-signal reconstruction schemes require *tuning* [6]. Iterative thresholding algorithms have attracted significant attention due to their simplicity, speed, theoretical performance guarantees, and ability to tackle large-scale reconstruction problems [6]–[7].

The goals of this paper are: to introduce sparse signal processing to the nondestructive evaluation (NDE) community, illustrate its applicability to NDE problems, and propose an *automatic* iterative thresholding sparse-signal reconstruction scheme that demands no prior knowledge about signal sparsity and no convergence thresholds.

We introduce the notation used in this paper:

- $\mathcal{N}(\boldsymbol{y}\,;\,\boldsymbol{\mu}, \Sigma)$ denotes the multivariate probability density function (pdf) of a real-valued Gaussian random vector $\boldsymbol{y}$ with mean vector $\boldsymbol{\mu}$ and covariance matrix Σ;
- $|\cdot|$, $\|\cdot\|_{\ell_p}$, and "^{T}" denote the determinant, ℓ_p norm, and transpose, respectively;
- $\lfloor x \rfloor$ is the largest integer smaller than or equal to x.

CP1211, *Review of Quantitative Nondestructive Evaluation Vol.29,*
edited by D. O. Thompson and D. E. Chimenti

MEASUREMENT MODEL

We model a $N \times 1$ real-valued measurement vector $\boldsymbol{y}$ as

$$\boldsymbol{y} = H\,\boldsymbol{z} \tag{1a}$$

where H is a known $N \times m$ full-rank *sensing matrix* with $N < m$, and $\boldsymbol{z}$ is an $m \times 1$ multivariate Gaussian vector with pdf

$$p(\boldsymbol{z}\,|\,\boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{z}\,|\,\mathcal{T}_A\{\boldsymbol{s}_A\}, \delta^2\,I_m). \tag{1b}$$

Here,

- A and $\boldsymbol{s}_A$ are a size-r index set defining the *sparse signal component* and the corresponding $r \times 1$ *vector of sparse-signal coefficients*, respectively; the size (cardinality) r of A is *unknown*. We refer to r as the *sparsity level* of the signal.
- δ^2 is an unknown *variance-component parameter*.
- $\mathcal{T}_A\{\boldsymbol{s}_A\}$ creates an $m \times 1$ vector by placing the elements of $\boldsymbol{s}_A$ at appropriate positions defined by the index set A and setting the remaining elements to zero; e.g. $\mathcal{T}_A\{\boldsymbol{s}_A\} = [s_1, s_2, 0, s_4, 0]^T$ for $m = 5$, $A = \{1, 2, 4\}$, and $\boldsymbol{s}_A = [s_1, s_2, s_4]^T$.

Define the set of all model parameters:

$$\boldsymbol{\theta} = (A, \boldsymbol{s}_A, \delta^2) \tag{2}$$

The likelihood function of $\boldsymbol{\theta}$ is obtained by *integrating* $\boldsymbol{z}$ *out* from the model (1):

$$p(\boldsymbol{y}\,|\,\boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{y}\,|\,H_A\,\boldsymbol{s}_A, \delta^2\,H\,H^T) \tag{3}$$

where H_A denotes the *restriction* of the sensing matrix H to an index set A, e.g. if $A = \{1, 2, 5\}$, then $H_A = [\boldsymbol{h}_1\,\boldsymbol{h}_2\,\boldsymbol{h}_5]$, where $\boldsymbol{h}_i$ is the ith column of H. To obtain (3), we have used the fact that $H\,\mathcal{T}_A\{\boldsymbol{s}_A\} = H_A\,\boldsymbol{s}_A$.

PARAMETER ESTIMATION ASSUMING KNOWN SPARSITY LEVEL r

For a known sparsity level r, the ML estimate of the model parameters $\boldsymbol{\theta}$ is

$$\widehat{\boldsymbol{\theta}}(r) = \arg \max_{\boldsymbol{\theta}, \dim(A)=r} \ln p(\boldsymbol{y}\,|\,\boldsymbol{\theta}) \tag{4}$$

where $\dim(A)$ denotes the size of the index set A. We treat $\boldsymbol{z}$ as the *missing (unobserved) data* and derive an *expectation-conditional maximization either (ECME)* iteration for estimating $\boldsymbol{\theta}$. An ECME algorithm maximizes *either* the expected complete-data log-likelihood function (where the expectation is computed with respect to the conditional distribution of the unobserved data given the observed measurements) *or* the actual observed-data log-likelihood, see [8, ch. 5.7].

An ECME Step for Estimating $\boldsymbol{\theta}$

Assume that a current estimate of the unknown parameters $\boldsymbol{\theta}^{(p)} = (A^{(p)}, \boldsymbol{s}^{(p)}_{A^{(p)}}, (\delta^2)^{(p)})$ is available, where p denotes the iteration index. First, select the indices of r largest-magnitude elements of

$$\boldsymbol{z}^{(p+1)} = \mathcal{T}_{A^{(p)}}\{\boldsymbol{s}^{(p)}_{A^{(p)}}\} + H^T\,(H\,H^T)^{-1}\left(\boldsymbol{y} - H_{A^{(p)}}\,\boldsymbol{s}^{(p)}_{A^{(p)}}\right) \tag{5a}$$

and use them to construct $A^{(p+1)}$; then, construct $\boldsymbol{s}_{A^{(p+1)}}^{(p+1)}$ using the values of these r largest elements; finally, update an estimate of the variance component δ^2 as

$$(\delta^2)^{(p+1)} = \frac{1}{N}\,(\boldsymbol{y} - H_{A^{(p+1)}}\,\boldsymbol{s}_{A^{(p+1)}}^{(p+1)})^T\,(H\,H^T)^{-1}\,(\boldsymbol{y} - H_{A^{(p+1)}}\,\boldsymbol{s}_{A^{(p+1)}}^{(p+1)}) \tag{5b}$$

yielding $\boldsymbol{\theta}^{(p+1)} = (A^{(p+1)}, \boldsymbol{s}_{A^{(p+1)}}^{(p+1)}, (\delta^2)^{(p+1)})$. For example, for $r = 3$ and if the first, second, and fifth elements of $\boldsymbol{z}^{(p+1)}$ have the largest magnitudes, we select $A^{(p+1)} = \{1, 2, 5\}$, $\boldsymbol{s}_{A^{(p+1)}}^{(p+1)} = [z_1^{(p+1)}, z_2^{(p+1)}, z_5^{(p+1)}]^T$, and $H_{A^{(p+1)}} = [\boldsymbol{h}_1\,\boldsymbol{h}_2\,\boldsymbol{h}_5]$, where $z_i^{(p+1)}$ denotes the ith element of $\boldsymbol{z}^{(p+1)}$. To simplify the notation, we omit the dependence of $\boldsymbol{z}^{(p+1)}$ and $(\delta^2)^{(p+1)}$ on r. Here, the updates of $A^{(p+1)}$ and $\boldsymbol{s}_{A^{(p+1)}}^{(p+1)}$ correspond to expectation-maximization (EM) steps whereas the update of $(\delta^2)^{(p+1)}$ in (5b) corresponds to the conditional maximization (CM) step. The above ECME iteration guarantees

$$\ln p(\boldsymbol{y} \,|\, \boldsymbol{\theta})\big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^{(p+1)}} \geq \ln p(\boldsymbol{y} \,|\, \boldsymbol{\theta})\big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^{(p)}}. \tag{6}$$

Note that $(H\,H^T)^{-1}$ can be pre-computed before the iteration starts; hence, this ECME iteration *does not* require matrix inversion. Interestingly, the above EM updates of $A^{(p+1)}$ and $\boldsymbol{s}_{A^{(p+1)}}^{(p+1)}$ are closely related to the *hard-thresholding step* in [7, eq. (11)]. Indeed, if the sensing matrix H is orthonormal, i.e.

$$H\,H^T = I_N \tag{7}$$

iteratively updating $A^{(p+1)}$ and $\boldsymbol{s}_{A^{(p+1)}}^{(p+1)}$ using the above EM steps is equivalent to the *iterative hard thresholding (IHT) algorithm* in [7, eq. (10)].

Fisher Information Matrix for $\boldsymbol{s}_A$ and δ^2 Assuming Known A

Using the well-known result in [9, eq. (3.32) on p. 48], we obtain the block-partitioned Fisher information matrix for the signal coefficients $\boldsymbol{s}_A$ and variance component δ^2 assuming known sparse-signal coefficient index set A:

$$\mathcal{I}(\boldsymbol{\theta}) = \frac{1}{\delta^2}\,\begin{bmatrix} H_A^T\,(H\,H^T)^{-1}\,H_A & \mathbf{0}_{r\times 1} \\ \mathbf{0}_{r\times 1}^T & \frac{1}{2}\,N/\delta^2 \end{bmatrix}. \tag{8}$$

GML RULE FOR SELECTING THE SPARSITY LEVEL r

Consider now the selection of the sparsity level r using the *GML rule* [10, p. 223]: maximize

$$\mathrm{GML}(r) = \mathrm{GL}(\boldsymbol{\theta})\big|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}(r)} \tag{9a}$$

with respect to r, where

$$\mathrm{GL}(\boldsymbol{\theta}) \;\;=\;\; \ln p(\boldsymbol{y} \,|\, \boldsymbol{\theta}) - \tfrac{1}{2}\,\ln|\mathcal{I}(\boldsymbol{\theta})| \tag{9b}$$

$\widehat{\boldsymbol{\theta}}(r)$ is the ML estimate of $\boldsymbol{\theta}$ *for a given* r in (4), and $\mathcal{I}(\boldsymbol{\theta})$ is the Fisher information matrix for $\boldsymbol{s}_A$ and δ^2 in (8). Here, the second term in (9b) *penalizes* the growth of r.

Direct application of the above GML rule may be computationally expensive. In particular, there are approximately N candidates of the sparsity level r where, for each r, we need to find $\widehat{\boldsymbol{\theta}}(r)$ by running a full ECME iteration. In the following, we outline our automatic hard thresholding (AHT) scheme for approximate maximization of (9a).

BASIC INGREDIENTS OF THE AHT ALGORITHM

We propose to interleave *expansion and compression steps* for updating estimates of r and *ECME steps* for updating estimates of $\boldsymbol{\theta}$. Here, the expansion and compression steps aim at improving (9b) whereas the ECME steps aim at approximating $\widehat{\boldsymbol{\theta}}(r)$ for a given r by increasing the likelihood function (3), see (6). We used a similar expansion/compression approach for GML maximization in [5]; however, the variance-component model in [5] is different from the statistical model (1) that we employ here.

Denote the iteration index by p and assume that $\boldsymbol{\theta}^{(p)} = (A^{(p)}, \boldsymbol{s}^{(p)}_{A^{(p)}}, (\delta^2)^{(p)})$ is available from *Iteration* p, where $\dim(A^{(p)}) = r^{(p)}$. Now, our AHT scheme consists of two types of iteration steps:

- an expansion/ECME step, which *increases* the sparsity level estimate by one, i.e. assigns $r^{(p+1)} = r^{(p)} + 1$ and performs one ECME step [where we replace r in the ECME step described on the previous page with $r^{(p+1)}$];
- a compression/ECME step, which *decreases* the sparsity level estimate by one, i.e. assigns $r^{(p+1)} = r^{(p)} - 1$ and applies one ECME step.

Determining the Best Parameter and Missing-data Estimates

In all iteration steps and all cycles of the AHT algorithm, we keep track of the *best estimate* $\boldsymbol{\theta}^\star = (A^\star, \boldsymbol{s}^\star_{A^\star}, (\delta^2)^\star)$ of the parameters $\boldsymbol{\theta}$ that yields the largest $\mathrm{GL}(\boldsymbol{\theta})$ in (9b) and the corresponding estimate of the missing-data signal-coefficient vector $\boldsymbol{z}^\star$. In particular, after each *Iteration* p yielding an estimate $\boldsymbol{\theta}^{(p)}$, we check the condition

$$\mathrm{GL}(\boldsymbol{\theta}^{(p)}) > \mathrm{GL}\big(\boldsymbol{\theta}^\star\big). \tag{10}$$

If (10) holds, set the new best parameter and missing-data estimates to $\boldsymbol{\theta}^\star = \boldsymbol{\theta}^{(p)}$ and $\boldsymbol{z}^\star = \boldsymbol{z}^{(p)}$; otherwise, keep $\boldsymbol{\theta}^\star$ and $\boldsymbol{z}^\star$ intact.

One Cycle of the AHT Algorithm

One AHT cycle consists of parameter initialization, the expansion sequence, and the compression sequence:

(Parameter initialization) Choose the initial estimates of the sparsity level $r^{(0)}$ and missing-data $\boldsymbol{z}^{(0)}$, select the $r^{(0)}$ largest-magnitude elements of $\boldsymbol{z}^{(0)}$ and use them to construct $A^{(0)}$ and $\boldsymbol{s}^{(0)}_{A^{(0)}}$.

(Expansion sequence) Apply the sequence of expansion/ECME steps, starting with $p = 0$ and ending with *Iteration* $p+1$ in which the *stopping condition*

$$\mathrm{GL}(\boldsymbol{\theta}^{(p+1)}) < \min\Big\{\mathrm{GL}(\boldsymbol{\theta}^{(p+1-L)}), \frac{1}{L}\sum_{i=0}^{L-1}\mathrm{GL}(\boldsymbol{\theta}^{(p-i)})\Big\} \tag{11}$$

is first satisfied, where L denotes the length of a *moving-average window*; here,

$$\begin{aligned}\mathrm{GL}(\boldsymbol{\theta}^{(p)}) \;\;=\;\; & -\tfrac{1}{2}\,(N-r-2)\,\ln[(\delta^2)^{(p)}] - \tfrac{1}{2}\,\ln|H^T_{A^{(p)}}\,(H\,H^T)^{-1}\,H_{A^{(p)}}| \\ & -\tfrac{1}{2}\,\ln|2\,\pi\,e\,H\,H^T| - \tfrac{1}{2}\,\ln(\tfrac{1}{2}\,N)\end{aligned} \tag{12}$$

follows by using (5b) (with $p+1$ replaced by p) and (8).

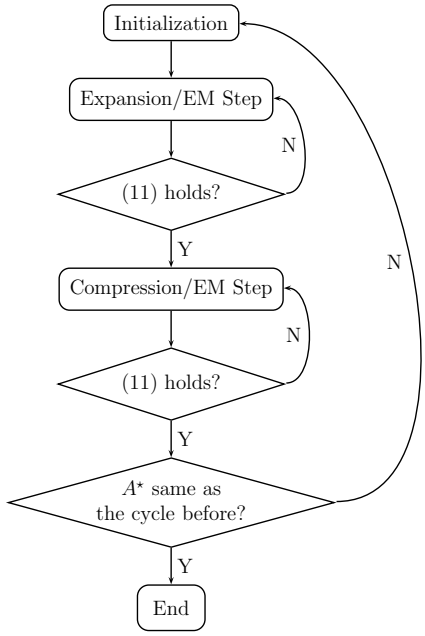

FIGURE 1. A flow diagram of the AHT algorithm.

(Compression sequence) Upon completion of the expansion sequence, apply the sequence of compression/ECME steps. This sequence is terminated using the stopping criterion (11).

THE AHT ALGORITHM

The flow diagram in Fig. 1 summarizes the full AHT algorithm, which consists of multiple AHT cycles:

- **Initialization**. Choose the initial sparsity level r_{init} and missing-data signal-coefficient vector $\boldsymbol{z}_{\text{init}}$.
- **Cycle 1**. Choose $r^{(0)} = r_{\text{init}}$ and $\boldsymbol{z}^{(0)} = \boldsymbol{z}_{\text{init}}$ and continue as described in the previous section, when discussing one AHT cycle.
- **Cycle $\boldsymbol{c} > \mathbf{1}$**. Reset the iteration index to $p = 0$ and choose

$$r^{(0)} = r^\star, \quad \boldsymbol{z}^{(0)} = \boldsymbol{z}_{\text{init}}$$

  where $r^\star$ is determined from $\boldsymbol{\theta}^\star$ (the best estimate of $\boldsymbol{\theta}$ found in all iterations of all past cycles). Continue as described in the previous section, when discussing one AHT cycle.
- If $A^\star$ (the best sparse-signal index set obtained in all iterations of all past cycles) has not changed between two consecutive cycles, finish the AHT algorithm and report the best parameter estimates $\boldsymbol{\theta}^\star$.

An Approximate AHT Scheme Assuming Orthonormal Sensing Matrix and RIP

For orthonormal sensing matrices H satisfying (7) and the restricted isometry property (RIP) [11], we approximate (12) as

$$\mathrm{GL}_{\mathrm{app}}(\boldsymbol{\theta}^{(p)}) \approx -\tfrac{1}{2}\,(N-r-2)\,\ln[(\delta^2)^{(p)}] - \tfrac{1}{2}\,r\,\ln\big(\frac{N}{m}\big) - \tfrac{1}{2}\,N\,\ln(2\,\pi\,e) - \tfrac{1}{2}\,\ln(\tfrac{1}{2}\,N). \quad (13)$$

Then, our approximate AHT scheme consists of replacing (12) with (13) throughout the AHT iteration.

NUMERICAL EXAMPLES: 2D TOMOGRAPHIC RECONSTRUCTION

We take tomographic projections of the Shepp-Logan phantom of size $m = 128^2$ with a 'defect,' shown in Fig. 2 (left). The elements of $\boldsymbol{y}$ are 2-D DFT coefficients of the image in Fig. 2 (left) sampled over a star-shaped domain, as illustrated in Fig. 2 (right); see also [3]. The sensing matrix is chosen as [2]

$$H = \Phi\,\Psi$$

with $N \times m$ *sampling matrix* Φ and $m \times m$ orthonormal sparsifying matrix Ψ constructed using selected rows of 2-D DFT matrix (yielding the corresponding 2-D DFT coefficients of the phantom image that are within the star-shaped domain) and Haar wavelet transform, respectively. Here, the rows of H are orthonormal, satisfying (7). The Haar wavelet transform coefficients of the image in Fig. 2 (left) are sparse, with $r \approx 0.1\,m$. In the example in Fig. 2 (right), the samples are taken along 30 radial lines in the frequency plane, each containing 128 samples, which yields $N/m \approx 0.22$.

Prior to application of the AHT algorithm, the measurements $\boldsymbol{y}$ have been scaled so that [see also (7)]

$$\frac{1}{N}\,\boldsymbol{y}^T\,(H\,H^T)^{-1}\,\boldsymbol{y} = \frac{1}{N}\|\boldsymbol{y}\|_{\ell_2}^2 = 1$$

which ensures that $(\delta^2)^{(p)} \le 1$ in all AHT iteration steps.

Our performance metric is the peak signal-to-noise ratio (PSNR) of a signal estimate $\widehat{\boldsymbol{s}}$:

$$\text{PSNR (dB)} = 20\log_{10}([\Psi\,\boldsymbol{s}]_{\mathrm{MAX}} - [\Psi\,\boldsymbol{s}]_{\mathrm{MIN}}) - 10\,\log_{10}(\|\widehat{\boldsymbol{s}} - \boldsymbol{s}\|_{\ell_2}^2/m)$$

where $[\Psi\,\boldsymbol{s}]_{\mathrm{MIN}}$ and $[\Psi\,\boldsymbol{s}]_{\mathrm{MAX}}$ denote the smallest and largest elements of $\Psi\,\boldsymbol{s}$. In the following, we compare the reconstruction performances of

- the traditional filtered backprojection where we set the unobserved DFT coefficients to zero and then take inverse DFT, see [3];
- the standard and debiased Barzilai-Borwein gradient-projection for sparse reconstruction method in [4, Sec. III.B] (labeled GPSR and GPSR-DB, respectively) with convergence threshold `tolP` $= 10^{-5}$ and regularization parameter $\tau = 0.001\,\|H^T\,\boldsymbol{y}\|_{\ell_\infty}$ (where `tolP` and τ have been manually tuned to achieve good performance), see `http://www.lx.it.pt/~mtf/GPSR`;
- our approximate AHT method that utilizes $\mathrm{GL}_{\mathrm{app}}(\boldsymbol{\theta}^{(p)})$ in (13), with $\boldsymbol{z}_{\mathrm{init}} = \widehat{\boldsymbol{s}}_{\mathrm{GPSR}}, r_{\mathrm{init}} = \lfloor N/[2\,\ln(m/N)]\rfloor$, and $L = 100$, where $\widehat{\boldsymbol{s}}_{\mathrm{GPSR}}$ is the GPSR estimate of the signal coefficients $\boldsymbol{s}$ and the choice of r_{init} is motivated by the asymptotic results in [12, Sec. 7.6.2]; AHT estimates the sparse signal $\boldsymbol{s}$ as $\widehat{\boldsymbol{s}}_{\mathrm{AHT}} = \mathcal{T}_{A^\star}\{\boldsymbol{s}^\star_{A^\star}\}$.

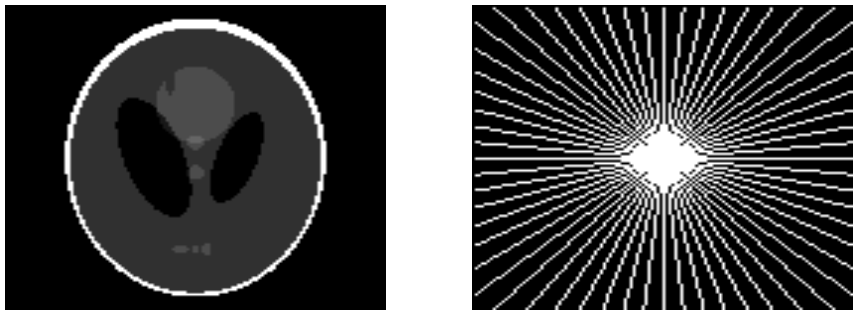

FIGURE 2. (Left) A size-128^2 Shepp-Logan phantom with a 'defect' and (right) a typical star-shaped sampling domain in the frequency plane.

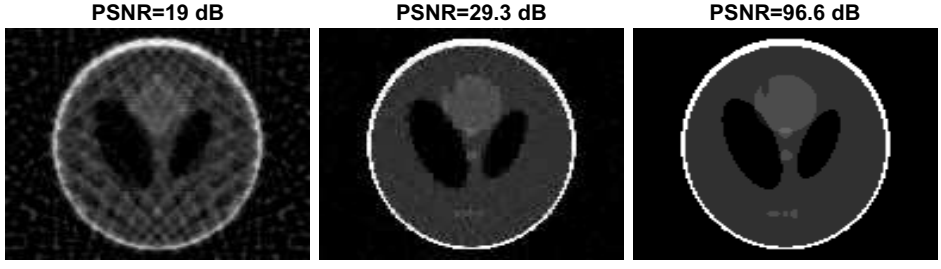

FIGURE 3. (Left) Filtered backprojection, (middle) GPSR-DB, and (right) AHT reconstructions for the sampling scheme in Fig. 2 (right).

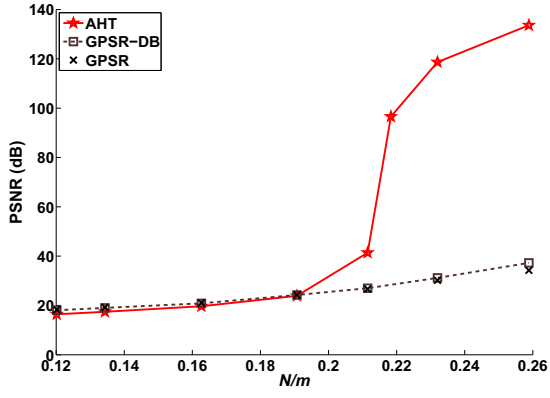

FIGURE 4. PSNR as a function of the normalized number of measurements N/m, where the number of measurements changes by varying the number of radial lines in the star-shaped sampling domain.

The reconstructions by the above methods (and corresponding PSNRs) are shown in Fig. 3 for the sampling pattern in Fig. 2 (right) with $N/m \approx 0.22$.

Fig. 4 shows the PSNR performances of the AHT, GPSR, and GPSR-DB methods as we vary the number of radial lines and, consequently, N/m. For $N/m \gtrsim 0.22$, AHT significantly outperforms the GPSR and GPSR-DB schemes. Observe the sharp phase transition

exhibited by AHT at $N \approx 0.22\,m$, which is close to the theoretical minimum observation number of $2\,r \approx 0.2\,m$.

CONCLUDING REMARKS

We developed an automatic hard thresholding method for reconstructing sparse signals from compressive samples and applied it to tomographic reconstruction from sparse projections.

ACKNOWLEDGMENT

This work was supported by the the National Science Foundation under Grant CCF-0545571 and NSF Industry-University Cooperative Research Program, Center for Nondestructive Evaluation (CNDE), Iowa State University.

REFERENCES

1. I.F. Gorodnitsky and B.D. Rao, "Sparse signal reconstruction from limited data using FOCUSS: A re-weighted minimum norm algorithm," *IEEE Trans. Signal Processing*, **45**, pp. 600–616 (1997).
2. E. Candès and J. Romberg, "Signal recovery from random projections," in *Computational Imaging III: Proc. SPIE-IS&T Electronic Imaging,* vol. 5674, C.A. Bouman and E.L. Miller (Eds.), San Jose, CA, Jan. 2005, pp. 76–86.
3. E.J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inform. Theory*, **52**, pp. 1207–1233 (2006).
4. M.A.T. Figueiredo, R.D. Nowak, and S.J. Wright, "Gradient projection for sparse reconstruction: application to compressed sensing and other inverse problems," *IEEE J. Select. Areas Signal Processing*, **1**, pp. 586–597 (2007).
5. A. Dogandžić and K. Qiu, "ExCoV: Expansion-compression variance-component based sparse-signal reconstruction from noisy measurements," in *Proc. 43rd Annu. Conf. Inform. Sci. Syst.*, Baltimore, MD, Mar. 2009, pp. 186–191.
6. A. Maleki and D.L. Donoho, "Freely available, optimally tuned iterative thresholding algorithms for compressed sensing," in *Proc. Signal Process. Adaptive Sparse Structured Representations (SPARS'09)*, Saint-Malo, France, Apr. 2009.
7. T. Blumensath and M.E. Davies, "Iterative hard thresholding for compressed sensing," *Appl. Comp. Harmonic Anal.*, vol. 27, pp. 265–274, Nov. 2009.
8. G.J. McLachlan and T. Krishnan, *The EM Algorithm and Extensions*, 2nd ed. New York: Wiley, 2008.
9. S.M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1993.
10. S.M. Kay, *Fundamentals of Statistical Signal Processing: Detection Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1998.
11. E.J. Candès and T. Tao, "Decoding by linear programming," *IEEE Trans. Inform. Theory*, **51**, pp. 4203–4215 (2005).
12. D.L. Donoho and J. Tanner "Counting faces of randomly projected polytopes when the projection radically lowers dimension," *J. Amer. Math. Soc.*, **22**, pp. 1–53, (2009).