Double Overrelaxation Thresholding Methods for Sparse Signal Reconstruction

Kun Qiu and Aleksandar Dogandžić
ECpE Department, Iowa State University
3119 Coover Hall, Ames, IA 50011, email: {kqiu,ald}@iastate.edu

***Abstract*—We propose double overrelaxation (DORE) and automatic double overrelaxation (ADORE) thresholding methods for sparse signal reconstruction. The measurements follow an underdetermined linear model, where the regression-coefficient vector is a sum of an unknown deterministic sparse signal component and a zero-mean white Gaussian component with an unknown variance. We first introduce an expectation-conditional maximization either (ECME) algorithm for estimating the sparse signal and variance of the random signal component and then describe our DORE thresholding scheme. The DORE scheme interleaves two successive overrelaxation steps and ECME steps, with goal to accelerate the convergence of the ECME algorithm. Both ECME and DORE algorithms aim at finding the maximum likelihood (ML) estimates of the unknown parameters assuming that the signal sparsity level is *known*. If the signal sparsity level is *unknown*, we propose an *unconstrained sparsity selection (USS)* criterion and show that, under certain conditions, maximizing the USS objective function with respect to the signal sparsity level is equivalent to finding the sparsest solution of the underlying underdetermined linear system. Our ADORE scheme demands no prior knowledge about the signal sparsity level and estimates this level by applying a golden-section search to maximize the USS objective function. We employ the proposed methods to reconstruct images from sparse tomographic projections and compare them with existing approaches that are feasible for large-scale data. Our numerical examples show that DORE is significantly faster than the ECME and related iterative hard thresholding (IHT) algorithms.**

***Index Terms*—Compressive sampling, sparse signal reconstruction, iterative hard thresholding, overrelaxation, model selection.**

I. Introduction

Sparse signal processing methods have attracted considerable attention recently, and have been developed and applied to biomagnetic imaging, spectral estimation, and compressive sampling, see [1]–[5] and references therein. However, most existing sparse signal reconstruction schemes require *tuning* [4], where the reconstruction performance depends crucially on the choice of the tuning parameters. Iterative hard thresholding (IHT) and normalized iterative hard thresholding (NIHT) sparse signal reconstruction algorithms in [6] and [7] (see also [8]) have attracted significant attention due to their low complexity, theoretical performance guarantees, and applicability to large-scale problems. However, these methods converge slowly to a fixed point, demanding a large number of iterations, and require knowledge of the signal sparsity level, which is a tuning parameter.

This work was supported by the National Science Foundation under Grant CCF-0545571.

In this paper, we develop

- a probabilistic framework for generalizing iterative hard thresholding (IHT) algorithms and interpreting them as expectation-conditional maximization either (ECME) iterations,
- a method for *accelerating* the convergence of the ECME and IHT algorithms, and
- a model selection criterion and an *automatic* reconstruction scheme that estimates the signal sparsity level from the data.

We introduce our two-stage hierarchical probabilistic model in Section II and review an expectation-conditional maximization either (ECME) algorithm for maximizing the marginal likelihood function of the unknown parameters (Section II-A); the IHT iteration in [6] is a special case of the ECME iteration in Section II-A, see [9]. In Section III, we describe our double overrelaxation (DORE) thresholding method for accelerating the convergence of the ECME iteration. In Section IV, we introduce an unconstrained sparsity selection (USS) criterion and show that, under certain conditions, USS is equivalent to finding the sparsest solution of the underlying underdetermined linear system. In Section IV-A, we combine this objective function with the DORE iteration to develop an *automatic double overrelaxation thresholding* method (ADORE) that estimates the signal sparsity level. Numerical simulations in Section V compare reconstruction performances of the proposed and existing methods.

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;
- $\|\cdot\|_{\ell_p}$, and "^{T}" denote the ℓ_p norm and transpose, respectively;
- I_n and $\mathbf{0}_{n\times 1}$ are the identity matrix of size n and the $n \times 1$ vector of zeros, respectively;
- $\lceil x \rceil$ is the smallest integer larger than or equal to x;
- $\dim(A)$ denotes the size of a set A;
- $\mathrm{supp}(\boldsymbol{x})$ returns the support set of a vector $\boldsymbol{x}$, i.e. the index set corresponding to the nonzero elements of $\boldsymbol{x}$, e.g. $\mathrm{supp}([0,1,-5,0,3,0]^T) = \{2,3,5\}$;
- the thresholding operator $\mathcal{T}_r(\boldsymbol{x})$ keeps the r largest-magnitude elements of a vector $\boldsymbol{x}$ intact and sets the rest to zero, e.g. $\mathcal{T}_2([0,1,-5,0,3,0]^T) = [0,0,-5,0,3,0]^T$.

978-1-4244-7417-2/10/$26.00 ©2010 IEEE

II. Probabilistic Measurement Model and the ECME Algorithm

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

$$\boldsymbol{y} = H\,\boldsymbol{z} \tag{2.1a}$$

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

$$p_{\boldsymbol{z}\,|\,\boldsymbol{\theta}}(\boldsymbol{z}\,|\,\boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{z}\,;\,\boldsymbol{s}, \sigma^2\,I_m) \tag{2.1b}$$

$\boldsymbol{s}$ is an *unknown* $m \times 1$ sparse signal vector containing at most r nonzero elements ($r < N$), and σ^2 is an unknown *variance-component parameter*; we refer to r as the *sparsity level* of the signal and to the signal vector $\boldsymbol{s}$ as being r-*sparse*. Note that $\|\boldsymbol{s}\|_{\ell_0} = \dim(\mathrm{supp}(\boldsymbol{s}))$ counts the number of nonzero elements in $\boldsymbol{s}$; we refer to $\|\boldsymbol{s}\|_{\ell_0}$ as the *support size* of $\boldsymbol{s}$. Therefore, the support size $\|\boldsymbol{s}\|_{\ell_0}$ of the r-sparse vector $\boldsymbol{s}$ is upper-bounded by the sparsity level r. The set of unknown parameters is

$$\boldsymbol{\theta} = (\boldsymbol{s}, \sigma^2) \in \Theta_r \tag{2.2}$$

with the parameter space

$$\Theta_r = \mathcal{S}_r \times [0, +\infty) = \{\boldsymbol{s} \in \mathcal{R}^m : \|\boldsymbol{s}\|_{\ell_0} \le r\} \times [0, +\infty). \tag{2.3}$$

The marginal likelihood function of $\boldsymbol{\theta}$ is obtained by *integrating* $\boldsymbol{z}$ *out* [see (2.1)]:

$$p_{\boldsymbol{y}\,|\,\boldsymbol{\theta}}(\boldsymbol{y}\,|\,\boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{y}\,;\,H\,\boldsymbol{s}, \sigma^2\,H\,H^T). \tag{2.4a}$$

For a given sparsity level r, the maximum likelihood (ML) estimate of $\boldsymbol{\theta}$ is

$$\widehat{\boldsymbol{\theta}}_{\mathrm{ML}}(r) = \big(\widehat{\boldsymbol{s}}_{\mathrm{ML}}(r), \widehat{\sigma}^2_{\mathrm{ML}}(r)\big) = \arg\max_{\boldsymbol{\theta} \in \Theta_r} p_{\boldsymbol{y}\,|\,\boldsymbol{\theta}}\big(\boldsymbol{y}\,|\,\boldsymbol{\theta}\big). \tag{2.4b}$$

Obtaining the exact ML estimate $\widehat{\boldsymbol{\theta}}_{\mathrm{ML}}(r)$ in (2.4b) requires a combinatorial search and is therefore infeasible in practice. We now review the computationally feasible approach in [9] that aims at maximizing (2.4a) with respect to $\boldsymbol{\theta} \in \Theta_r$ and circumvents the combinatorial search.

In Sections II-A and III, we present the ECME and DORE iterations for estimating $\boldsymbol{\theta}$ assuming that r is known; therefore, in these sections we simplify the notation and omit the dependence of the estimates of $\boldsymbol{\theta}$ on the sparsity level r.

A. ECME Algorithm For Known Sparsity Level r

We treat $\boldsymbol{z}$ as the *missing (unobserved) data* and present an *expectation-conditional maximization either (ECME)* algorithm for approximately finding the ML estimate in (2.4b), assuming a fixed sparsity level r. 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 [12, Ch. 5.7] and [13].

Assume that the parameter set estimate $\boldsymbol{\theta}^{(p)} = \big(\boldsymbol{s}^{(p)}, (\sigma^2)^{(p)}\big)$ is available, where p denotes the iteration index. *Iteration* $p+1$ proceeds as (see also [9]):

- Update the sparse signal estimate using the expectation-maximization (EM) step, i.e. the expectation (E) step:

$$\begin{aligned}\boldsymbol{z}^{(p+1)} &= \mathrm{E}_{\boldsymbol{z}\,|\,\boldsymbol{y},\boldsymbol{\theta}}[\boldsymbol{z}\,|\,\boldsymbol{y}, \boldsymbol{\theta}^{(p)}] \\ &= \boldsymbol{s}^{(p)} + H^T\big(H\,H^T\big)^{-1}[\boldsymbol{y} - H\,\boldsymbol{s}^{(p)}]\end{aligned} \tag{2.5a}$$

  followed by the maximization (M) step, which simplifies to

$$\boldsymbol{s}^{(p+1)} = \arg\min_{\boldsymbol{s} \in \mathcal{S}_r} \|\boldsymbol{z}^{(p+1)} - \boldsymbol{s}\|^2_{\ell_2} = \mathcal{T}_r(\boldsymbol{z}^{(p+1)}) \tag{2.5b}$$

  and

- update the estimate of the variance component σ^2 using the following conditional maximization (CM) step:

$$(\sigma^2)^{(p+1)} = \frac{[\boldsymbol{y} - H\,\boldsymbol{s}^{(p+1)}]^T\,(H\,H^T)^{-1}\,[\boldsymbol{y} - H\,\boldsymbol{s}^{(p+1)}]}{N}. \tag{2.5c}$$

Here, $\mathrm{E}_{\boldsymbol{z}\,|\,\boldsymbol{y},\boldsymbol{\theta}}[\boldsymbol{z}\,|\,\boldsymbol{y},\boldsymbol{\theta}]$ denotes the mean of the pdf $p_{\boldsymbol{z}\,|\,\boldsymbol{y},\boldsymbol{\theta}}(\boldsymbol{z}\,|\,\boldsymbol{y},\boldsymbol{\theta})$, which is the Bayesian minimum mean-square error (MMSE) estimate of $\boldsymbol{z}$ for *known* $\boldsymbol{\theta}$ [14, Sec. 11.4]. If the rows of the sensing matrix H are orthonormal:

$$H\,H^T = I_N \tag{2.6}$$

the update of $\boldsymbol{s}^{(p+1)}$ in (2.5b) is equivalent to one *iterative hard-thresholding (IHT) step* in [6, eq. (10)].

Note that the parameter space Θ_r in (2.3) is non-convex with empty interior and that the complete-data and conditional unobserved data given the observed data distributions $p_{\boldsymbol{z},\boldsymbol{y}\,|\,\boldsymbol{\theta}}(\boldsymbol{z},\boldsymbol{y}\,|\,\boldsymbol{\theta})$ and $p_{\boldsymbol{z}\,|\,\boldsymbol{y},\boldsymbol{\theta}}(\boldsymbol{z}\,|\,\boldsymbol{y},\boldsymbol{\theta})$ are degenerate. Therefore, our probabilistic model *does not* satisfy the general regularity conditions assumed in standard convergence analysis of the EM-type algorithms in [12] and [13, Theorem 2]. Nevertheless, we have shown that, under certain mild conditions, our ECME iteration (2.5) ensures monotonically non-decreasing marginal likelihood (2.4a) between consecutive iterations and converges to a local maximum of (2.4a). However, the ECME iteration (2.5) converges slowly. In the following section, we propose our DORE method for accelerating the ECME iteration.

III. The DORE Algorithm for Known r

Assume that two consecutive estimates of the unknown parameters $\boldsymbol{\theta}^{(p-1)} = (\boldsymbol{s}^{(p-1)}, (\sigma^2)^{(p-1)})$ and $\boldsymbol{\theta}^{(p)} = (\boldsymbol{s}^{(p)}, (\sigma^2)^{(p)})$ are available from the $(p-1)$-th and p-th iterations, respectively. *Iteration* $p+1$ proceeds as follows:

1. ECME step. Compute

$$\widehat{\boldsymbol{s}} = \mathcal{T}_r\big(\boldsymbol{s}^{(p)} + H^T(HH^T)^{-1}(\boldsymbol{y} - H\boldsymbol{s}^{(p)})\big) \tag{3.1a}$$

$$\widehat{\sigma}^2 = (\boldsymbol{y} - H\,\widehat{\boldsymbol{s}})^T\,(H\,H^T)^{-1}\,(\boldsymbol{y} - H\,\widehat{\boldsymbol{s}})/N \tag{3.1b}$$

and define $\widehat{\boldsymbol{\theta}} = (\widehat{\boldsymbol{s}}, \widehat{\sigma}^2)$.

2. First overrelaxation. Compute the linear combination of $\widehat{\boldsymbol{s}}$ and $\boldsymbol{s}^{(p)}$:

$$\bar{\boldsymbol{z}} = \widehat{\boldsymbol{s}} + \alpha_1\,(\widehat{\boldsymbol{s}} - \boldsymbol{s}^{(p)}) \tag{3.2a}$$

where the weight

$$\alpha_1 = \frac{(H\,\widehat{\boldsymbol{s}} - H\,\boldsymbol{s}^{(p)})^T\,(HH^T)^{-1}(\boldsymbol{y} - H\,\widehat{\boldsymbol{s}})}{(H\,\widehat{\boldsymbol{s}} - H\,\boldsymbol{s}^{(p)})^T\,(HH^T)^{-1}(H\,\widehat{\boldsymbol{s}} - H\,\boldsymbol{s}^{(p)})} \tag{3.2b}$$

is the closed-form solution of the line search:

$$\alpha_1 = \arg\max_{\alpha} p_{\boldsymbol{y}\,|\,\boldsymbol{\theta}}\big(\boldsymbol{y}\,|\,(\widehat{\boldsymbol{s}} + \alpha\,(\widehat{\boldsymbol{s}} - \boldsymbol{s}^{(p)}), \sigma^2)\big) \tag{3.2c}$$

with the parameter space of $\boldsymbol{\theta}$ extended to Θ_{r_1}, where $r_1 = \dim(\mathrm{supp}(\widehat{\boldsymbol{s}}) \cup \mathrm{supp}(\boldsymbol{s}^{(p)}))$ is the sparsity level of $\widehat{\boldsymbol{s}} + \alpha\,(\widehat{\boldsymbol{s}} - \boldsymbol{s}^{(p)})$ and σ^2 is an arbitrary positive number, see also (2.4a).

3. Second overrelaxation. Compute the linear combination of $\bar{\boldsymbol{z}}$ and $\boldsymbol{s}^{(p-1)}$:

$$\widetilde{\boldsymbol{z}} = \bar{\boldsymbol{z}} + \alpha_2\,(\bar{\boldsymbol{z}} - \boldsymbol{s}^{(p-1)}) \tag{3.3a}$$

where the weight

$$\alpha_2 = \frac{(H\bar{\boldsymbol{z}} - H\boldsymbol{s}^{(p-1)})^T(HH^T)^{-1}(\boldsymbol{y} - H\bar{\boldsymbol{z}})}{(H\bar{\boldsymbol{z}} - H\boldsymbol{s}^{(p-1)})^T(HH^T)^{-1}(H\bar{\boldsymbol{z}} - H\boldsymbol{s}^{(p-1)})} \tag{3.3b}$$

is the closed-form solution of the line search:

$$\alpha_2 = \arg\max_{\alpha} p_{\boldsymbol{y}\,|\,\boldsymbol{\theta}}\big(\boldsymbol{y}\,|\,(\bar{\boldsymbol{z}} + \alpha\,(\bar{\boldsymbol{z}} - \boldsymbol{s}^{(p-1)}), \sigma^2)\big) \tag{3.3c}$$

with the parameter space of $\boldsymbol{\theta}$ extended to Θ_{r_2}, where $r_2 = \dim(\mathrm{supp}(\bar{\boldsymbol{z}}) \cup \mathrm{supp}(\boldsymbol{s}^{(p-1)}))$ is the sparsity level of $\bar{\boldsymbol{z}} + \alpha\,(\bar{\boldsymbol{z}} - \boldsymbol{s}^{(p-1)})$ and σ^2 is an arbitrary positive number.

4. Thresholding. Threshold $\widetilde{\boldsymbol{z}}$ to the sparsity level r:

$$\widetilde{\boldsymbol{s}} = \mathcal{T}_r(\widetilde{\boldsymbol{z}}) \tag{3.4a}$$

compute the corresponding variance component estimate:

$$\widetilde{\sigma}^2 = (\boldsymbol{y} - H\,\widetilde{\boldsymbol{s}})^T(HH^T)^{-1}(\boldsymbol{y} - H\,\widetilde{\boldsymbol{s}})/N \tag{3.4b}$$

and define our final overrelaxation parameter estimate $\widetilde{\boldsymbol{\theta}} = (\widetilde{\boldsymbol{s}}, \widetilde{\sigma}^2)$.

5. Decision (between ECME and thresholded overrelaxation parameter estimates). If $p_{\boldsymbol{y}\,|\,\boldsymbol{\theta}(r)}(\boldsymbol{y}\,|\,\widetilde{\boldsymbol{\theta}}) \geq p_{\boldsymbol{y}\,|\,\boldsymbol{\theta}(r)}(\boldsymbol{y}\,|\,\widehat{\boldsymbol{\theta}})$ or, equivalently, if

$$\widetilde{\sigma}^2 < \widehat{\sigma}^2 \tag{3.5}$$

assign $\boldsymbol{\theta}^{(p+1)} = \widetilde{\boldsymbol{\theta}}$; otherwise, assign $\boldsymbol{\theta}^{(p+1)} = \widehat{\boldsymbol{\theta}}$ and complete *Iteration* $p+1$.

Iterate until two consecutive sparse-signal estimates $\boldsymbol{s}^{(p)}$ and $\boldsymbol{s}^{(p+1)}$ do not differ significantly. Since $(H\,H^T)^{-1}$ can be precomputed before the iteration starts, our DORE iteration *does not* require matrix inversion, see Section III-A for details on computational complexity.

If the rows of the sensing matrix H are orthonormal [i.e. (2.6) holds], Step **1** of the DORE scheme reduces to one IHT step. After Step **1**, we apply two overrelaxations (Steps **2** and **3**), that utilize the sparse signal estimates $\boldsymbol{s}^{(p)}$ and $\boldsymbol{s}^{(p-1)}$ from the two most recent completed iterations. The goal of the overrelaxation steps is to boost the marginal likelihood (2.4a) and accelerate the convergence of the ECME iteration. Using a single overrelaxation step based on the most recent parameter estimate is a common approach for accelerating fixed-point iterations, see [10]. Here, we adopt the idea in [10, Sect. 5.1] and apply the *second overrelaxation*, which mitigates the 'zigzagging' effect caused by the first overrelaxation and thereby converges more rapidly. Our algorithm differs from that in [10, Sect. 5.1], which focuses on continuous parameter spaces with marginal likelihood that is differentiable with respect to the parameters. Unlike [10, Sect. 5.1], here we

- apply overrelaxation steps on parameter spaces with variable dimensions (Steps **2** and **3**),
- threshold the second overrelaxation estimate to ensure that the resulting signal estimate is r-sparse (Step **4**), and
- test the thresholded estimate from Step **4** versus the ECME estimate from Step **1** and adopt the better of the two (Step **5**).

Step **5** ensures that the resulting new parameter estimate $\boldsymbol{\theta}^{(p+1)}$ yields the marginal likelihood function (2.4a) that is higher than or equal to that of the standard ECME step (Step **1**). Therefore, the DORE iteration (3.1)–(3.5) ensures monotonically non-decreasing marginal likelihood between consecutive iteration steps

$$p_{\boldsymbol{y}\,|\,\boldsymbol{\theta}}(\boldsymbol{y}\,|\,\boldsymbol{\theta}^{(p+1)}) \geq p_{\boldsymbol{y}\,|\,\boldsymbol{\theta}}(\boldsymbol{y}\,|\,\boldsymbol{\theta}^{(p)}) \tag{3.6}$$

and converges to a local maximum of the marginal likelihood (2.4a), see also the discussion in Section II-A.

DORE initialization. The first two parameter estimates $\boldsymbol{\theta}^{(1)}$ and $\boldsymbol{\theta}^{(2)}$ are obtained by applying two consecutive ECME steps (2.5) initialized by the zero sparse signal estimate

$$\boldsymbol{s}^{(0)} = \boldsymbol{0}_{m\times 1}. \tag{3.7}$$

A. Computational Complexity and Memory Requirements

The major computation complexity of the ECME algorithm lies in the sorting operation of the $m \times 1$ vector and the matrix-vector operations. Assuming that the common bubble sorting is employed, sorting $\boldsymbol{z}^{(p+1)}$ in (2.5b) requires $\mathcal{O}(m^2)$ operations. There are three matrix-vector multiplications in one ECME iteration, namely $H\boldsymbol{s}^{(p)}$, $(HH^T)^{-1}[H\boldsymbol{s}^{(p)}]$ and $H^T[(HH^T)^{-1}H\boldsymbol{s}^{(p)}]$, which requires $\mathcal{O}(Nm)$, $\mathcal{O}(N^2)$ and $\mathcal{O}(Nm)$ operations, respectively. The intermediate computation results of $(\sigma^2)^{(p+1)}$ in (2.5c) can be stored and used to compute (2.5b) of the next iteration; therefore, this step does not cause additional computation. In summary, the complexity of one ECME iteration is $\mathcal{O}(m^2 + 2Nm + N^2)$. If H has orthonormal rows satisfying (2.6), ECME reduces to the IHT iteration, and in this case $(HH^T)^{-1}[H\boldsymbol{s}^{(p)}]$ is simply $H\boldsymbol{s}^{(p)}$. The computation complexity of one IHT step is therefore $\mathcal{O}(m^2 + 2Nm)$.

For DORE, there are two sorting operations per iteration, one in step **1** and the other in step **4**, requiring $\mathcal{O}(2m^2)$ operations. In one DORE iteration, we need to compute $H^T[(HH^T)^{-1}H\boldsymbol{s}^{(p)}]$, $H\,\widehat{\boldsymbol{s}}$, $(H\,H^T)^{-1}\,[H\,\widehat{\boldsymbol{s}}]$, $H\,\widetilde{\boldsymbol{s}}$, and $(H\,H^T)^{-1}\,[H\,\widetilde{\boldsymbol{s}}]$, which require total of $\mathcal{O}(3Nm + 2N^2)$ operations. Note that $H\boldsymbol{s}^{(p-1)}$, $(HH^T)^{-1}[H\boldsymbol{s}^{(p-1)}]$, $H\boldsymbol{s}^{(p)}$, and $(HH^T)^{-1}[H\boldsymbol{s}^{(p)}]$ in (3.1a), (3.2b) and (3.3b) can be adopted from the previous two iterations and do not need to be computed again in the current iteration; in addition, the quantities $H\bar{\boldsymbol{z}}$ and $(HH^T)^{-1}[H\bar{\boldsymbol{z}}]$ in (3.3b) are simple linear combinations of computed terms and do not require

additional matrix-vector computations. To summarize, one DORE iteration requires $\mathcal{O}(2m^2 + 3Nm + 2N^2)$, which is slightly less than twice the complexity of one ECME step. When H has orthonormal rows, we do not need to compute $(H\,H^T)^{-1}\,[H\,\widehat{\boldsymbol{s}}]$ and $(H\,H^T)^{-1}\,[H\,\widetilde{\boldsymbol{s}}]$, which brings the complexity down to $\mathcal{O}(2m^2 + 3Nm)$, slightly less than twice the complexity of one IHT step.

Regarding the memory storage, the largest quantity that ECME (and its special case IHT) and DORE need to store is the sensing matrix H requiring memory storage of order $\mathcal{O}(Nm)$. In large-scale applications, H is typically not explicitly stored but instead appears in the function-handle form [for example, random DFT sensing matrix can be implemented via the fast Fourier transform (FFT)]. In this case, the storage requirement of ECME, IHT and DORE is just $\mathcal{O}(m)$.

Although a single DORE step is about twice more complex than the ECME and IHT steps, it converges in much fewer iterations than the ECME and IHT iterations in the numerical examples in Section V, see Figs. 2 (b).

IV. Unconstrained Sparsity Selection Criterion and the ADORE Algorithm

We propose the following *unconstrained sparsity selection (USS)* objective function for selecting the proper sparsity level r that strikes a balance between the efficiency and accuracy of signal representation:

$$\begin{aligned}\mathrm{USS}(r) \;&=\; -\tfrac{1}{2}\,r\,\ln\big(\frac{N}{m}\big)\\ &\quad -\tfrac{1}{2}\,(N-r-2)\,\ln\Big(\frac{\widehat{\sigma}^2_{\mathrm{ML}}(r)}{\boldsymbol{y}^T\,(H\,H^T)^{-1}\,\boldsymbol{y}/N}\Big)\end{aligned}\tag{4.1}$$

where $\widehat{\sigma}^2_{\mathrm{ML}}(r)$ is the ML estimate of the variance component σ^2 in the parameter space Θ_r, see (2.4b). $\mathrm{USS}(r)$ in (4.1) is developed from the approximate generalized maximum likelihood (GML) objective function in [9, e.q. (13)]; in particular, when $\boldsymbol{y}^T\,(H\,H^T)^{-1}\,\boldsymbol{y}/N = 1$, the two functions are equal up to an additive constant. However, unlike the GML objective function, (4.1) is scale-invariant, i.e. scaling the measurements $\boldsymbol{y}$ by a nonzero constant does not change $\mathrm{USS}(r)$, which is a desirable property.

The following theorem states that, under certain conditions, the optimal sparsity level selected by maximizing $\mathrm{USS}(r)$ with respect to r *coincides* with the support size of the solution to the well-known (P_0) problem [5, eq. (2)]:

$$(\mathrm{P}_0): \qquad \min_{\boldsymbol{s}} \|\boldsymbol{s}\|_{\ell_0} \quad \text{subject to } \boldsymbol{y} = H\,\boldsymbol{s} \tag{4.2}$$

which seeks the sparsest solution of an underdetermined linear system $\boldsymbol{y} = H\,\boldsymbol{s}$. Denote by $\boldsymbol{s}^\diamond$ the solutions to (4.2) and by $r^\diamond = \|\boldsymbol{s}^\diamond\|_{\ell_0}$ the support size of $\boldsymbol{s}^\diamond$.

Theorem 1: Assume that the $N \times m$ sensing matrix H has full rank and $N < m$. If

(1) the sensing matrix H satisfies the unique representation property (URP) [1], stating that all $N \times N$ submatrices of H are invertible, and

(2) the number of measurements N satisfies

$$N > r^\diamond + 2 \tag{4.3}$$

then $\mathrm{USS}(r)$ in (4.1) is *globally and uniquely maximized* at $r = r^\diamond$ and the (P_0)-optimal solutions $\boldsymbol{s}^\diamond$ coincide with the ML signal estimates $\widehat{\boldsymbol{s}}_{\mathrm{ML}}(r^\diamond)$.

Proof: We outline the proof here. For $r = r^\diamond$, the ML estimates of $\boldsymbol{\theta}$ are $\widehat{\boldsymbol{\theta}}_{\mathrm{ML}}(r^\diamond) = \big(\widehat{\boldsymbol{s}}_{\mathrm{ML}}(r^\diamond), \widehat{\sigma}^2_{\mathrm{ML}}(r^\diamond)\big) = (\boldsymbol{s}^\diamond, 0)$, since they lead to infinite likelihood function (2.4a) and no other $\boldsymbol{\theta}$ yields infinite likelihood. Furthermore, since (4.3) holds, $\mathrm{USS}(r)$ is infinite as well. Note that

$$\mathrm{USS}(r) = \widetilde{\mathrm{USS}}\big(r, \widehat{\sigma}^2_{\mathrm{ML}}(r)\big) \tag{4.4}$$

where

$$\begin{aligned}\widetilde{\mathrm{USS}}(r,\sigma^2) \;&=\; -\tfrac{1}{2}\,r\,\ln\big(\frac{N}{m}\big)\\ &\quad -\tfrac{1}{2}\,(N-r-2)\,\ln\Big(\frac{\sigma^2}{\boldsymbol{y}^T\,(H\,H^T)^{-1}\,\boldsymbol{y}/N}\Big).\end{aligned}\tag{4.5}$$

Now,

$$\lim_{\sigma^2 \searrow 0} \frac{\widetilde{\mathrm{USS}}(r^\diamond, \sigma^2)}{\ln(1/\sigma^2)} = \tfrac{1}{2}\,(N - r^\diamond - 2) > 0 \tag{4.6}$$

specifying the rate of growth to infinity of $\widetilde{\mathrm{USS}}(r^\diamond, \sigma^2)$ as σ^2 approaches the ML estimate $\widehat{\sigma}^2_{\mathrm{ML}}(r^\diamond) = 0$.

For $r < r^\diamond$, $\boldsymbol{y} \neq H\,\boldsymbol{s}$ for any r-sparse vector $\boldsymbol{s}$; consequently, $\sigma^2_{\mathrm{ML}}(r) > 0$ and $\mathrm{USS}(r)$ is finite.

For $r > r^\diamond$, the ML estimate of σ^2 must be $\widehat{\sigma}^2_{\mathrm{ML}}(r) = 0$, which leads to infinite likelihood. However, in this case,

$$\lim_{\sigma^2 \searrow 0} \frac{\widetilde{\mathrm{USS}}(r, \sigma^2)}{\ln(1/\sigma^2)} = \tfrac{1}{2}(N - r - 2) < \tfrac{1}{2}(N - r^\diamond - 2). \tag{4.7}$$

Therefore, if $r \geq N - 2$, $\mathrm{USS}(r)$ is either finite or goes to negative infinity. For $r^\diamond < r < N - 2$, $\mathrm{USS}(r)$ is infinitely large, but the rate at which $\widetilde{\mathrm{USS}}(r, \sigma^2)$ grows to infinity as σ^2 approaches the ML estimate $\widehat{\sigma}^2_{\mathrm{ML}}(r) = 0$ is smaller than that specified by (4.6).

The claim follows by combining the above conclusions. □

Theorem 1 demonstrates that the USS objective function *transforms* the constrained optimization problem (P_0) in (4.2) into an equivalent unconstrained problem (4.1) and that it is capable of optimally selecting the sparsity level r that allows accurate signal representation and keeps as few nonzero signal elements as possible.

Note that Theorem 1 *does not* guarantee the uniqueness of the (P_0)-optimal signal estimate $\boldsymbol{s}^\diamond$. For H satisfying the URP condition (1), [5, Theorem 2 in Sect. 2.1.1] states that this uniqueness is guaranteed by

$$N \geq 2\,r^\diamond \tag{4.8}$$

which is more stringent than (4.3) in practical scenarios where $r^\diamond > 2$.

Maximizing (4.1) with respect to r by an exhaustive search may be computationally expensive because we need to apply a full DORE iteration for each sparsity level r in the set of integers between 0 and $N/2$. Note that, based on (4.8), $N/2$ is the largest sparsity level for which reasonable reconstruction is possible from N measurements.

A. The ADORE Algorithm for Unknown Sparsity Level r

Instead of the exhaustive search, our proposed ADORE algorithm applies the *golden-section search* [15, Sec. 4.5.2.1] to maximize (4.1) with respect to r, with the initial search boundaries set to 0 and $\lceil N/2 \rceil$. Note that $\mathrm{USS}(0) = 0$ assuming that $\boldsymbol{y} \neq \mathbf{0}_{N\times 1}$, which is of practical interest. For each candidate $0 < r \leq \lceil N/2 \rceil$, we estimate $\widehat{\sigma}^2_{\mathrm{ML}}(r)$ using the DORE iteration. After running one golden sectioning step, the length of the new search interval is approximately 0.618 of the previous interval (rounded to the closest integer). The search process ceases when the desired resolution L is reached, i.e. when the searching interval becomes shorter than the prescribed resolution level L. Therefore, ADORE requires roughly $1.4\,(\log_2 \frac{N}{L} - 1)$ full DORE iterations. For the golden-section search to find the exact maximum of (4.1), $\mathrm{USS}(r)$ must be *unimodal* in r, which is not true in general. Hence, ADORE maximizes (4.1) only approximately, yielding r_{ADORE}; then, our ADORE sparse-signal estimate is equal to the corresponding DORE estimate at $r = r_{\mathrm{ADORE}}$.

V. Numerical Examples

Consider the reconstruction of the Shepp-Logan phantom of size $m = 256^2$ in Fig. 1 (a) from tomographic projections. The elements of $\boldsymbol{y}$ are 2-D discrete Fourier transform (DFT) coefficients of the image in Fig. 1 (a) sampled over a star-shaped domain, as illustrated in Fig. 1 (b); see also [2], [7], and [9]. As in [7] and [9], the sensing matrix is $H = \Phi\,\Psi$ with $N \times m$ *sampling matrix* Φ and $m \times m$ orthonormal sparsifying matrix Ψ constructed using selected rows of the DFT matrix (yielding the corresponding DFT coefficients of the phantom image that are within the star-shaped domain) and inverse Haar wavelet transform, respectively. Here, the rows of H are orthonormal, satisfying (2.6). The Haar wavelet transform coefficients of the phantom image in Fig. 1 (a) are sparse, with $\|\boldsymbol{s}\|_{\ell_0} = 3769 \approx 0.06\,m$, where the true signal vector $\boldsymbol{s}$ consists of the Haar wavelet transform coefficients of the phantom in Fig. 1 (a).

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

$$\mathrm{PSNR\ (dB)} = 10\,\log_{10}\left\{\frac{[(\Psi\,\boldsymbol{s})_{\mathrm{MAX}} - (\Psi\,\boldsymbol{s})_{\mathrm{MIN}}]^2}{\|\widehat{\boldsymbol{s}} - \boldsymbol{s}\|^2_{\ell_2}/m}\right\} \quad (5.1)$$

where $(\Psi\,\boldsymbol{s})_{\mathrm{MIN}}$ and $(\Psi\,\boldsymbol{s})_{\mathrm{MAX}}$ denote the smallest and largest elements of the phantom image $\Psi\,\boldsymbol{s}$. We compare the PSNRs (5.1) of

- the DORE and ADORE schemes, ADORE search resolution set to $L = 500$, with MATLAB implementations available at http://home.eng.iastate.edu/~ald/DORE.htm,
- the IHT and NIHT schemes in [6] and [7],
- the automatic hard thresholding (AHT) method in [9] using the moving-average window length 100, initialized with $\boldsymbol{z}_{\mathrm{init}} = \mathbf{0}_{m\times 1}$ and $r_{\mathrm{init}} = 1$,
- the debiased gradient-projection for sparse reconstruction method in [11, Sec. III.B] with the convergence threshold $\mathtt{tolP} = 10^{-5}$ and regularization parameter set to (i) $\tau = 0.1\,\|H^T\,\boldsymbol{y}\|_{\ell_\infty}$, suggested in [11, e.q. (22)] (labeled $\mathrm{GPSR_0}$) and (ii) $\tau = 0.001\,\|H^T\,\boldsymbol{y}\|_{\ell_\infty}$ obtained by manual tuning for good performance (labeled GPSR),
- the standard filtered back-projection that corresponds to setting the unobserved DFT coefficients to zero and taking the inverse DFT, see [2].

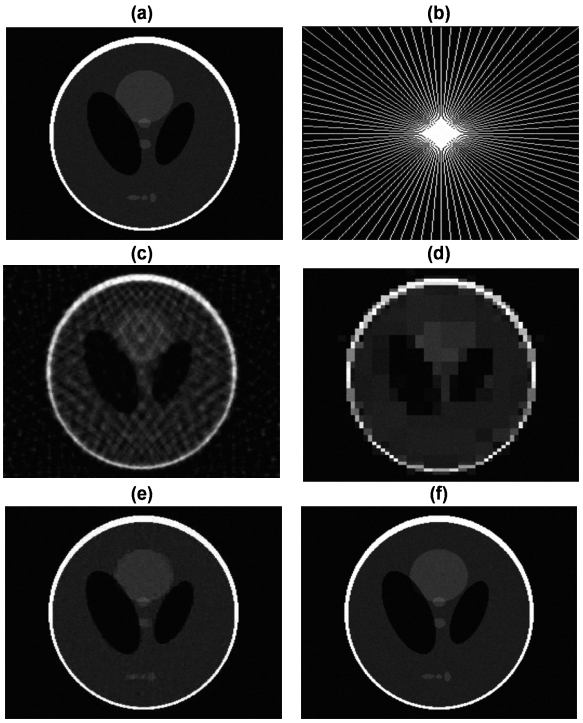

Fig. 1. (a) The size-256^2 Shepp-Logan phantom, (b) a typical star-shaped sampling domain in the frequency plane containing 44 radial lines, (c) the filtered back-projection reconstruction, (d) $\mathrm{GPSR_0}$ reconstruction, (e) GPSR reconstruction, and (f) the almost perfect reconstruction achieved by DORE, ADORE, AHT, IHT, and NIHT schemes, for the sampling pattern in (b).

Since the sensing matrix H is orthonormal satisfying (2.6), IHT is equivalent to the ECME iteration in Section II-A.

For the DORE, ADORE, IHT and NIHT algorithms, we use the following convergence criterion:[1]

$$\|\boldsymbol{s}^{(p+1)} - \boldsymbol{s}^{(p)}\|^2_{\ell_2}/m < 10^{-14}. \quad (5.2)$$

DORE, IHT, and NIHT require knowledge of the signal sparsity level r; to implement these methods, we set r to the true signal support size: $r = 3769$. In contrast, the ADORE and AHT methods are *automatic* and *estimate* r from the measurements.

Fig. 1 (c)–(f) present the reconstructed images from the 44 radial lines given in Fig. 1 (b) using the above methods. Here, all hard-thresholding methods (DORE, IHT, NIHT, ADORE, and AHT) recovered the original image almost perfectly, in contrast with the filtered back-projection and GPSR methods that achieve inferior reconstructions.

Fig. 2 shows (a) the PSNRs, (b) numbers of iterations, and (c) CPU times of the above methods (implemented in MATLAB) as we change N/m by varying the number of

[1] To implement the IHT and NIHT schemes, we incorporated the convergence criterion (5.2) into the corresponding MATLAB codes from the SPARSIFY toolbox at http://www.see.ed.ac.uk/~tblumens/sparsify/sparsify.html.

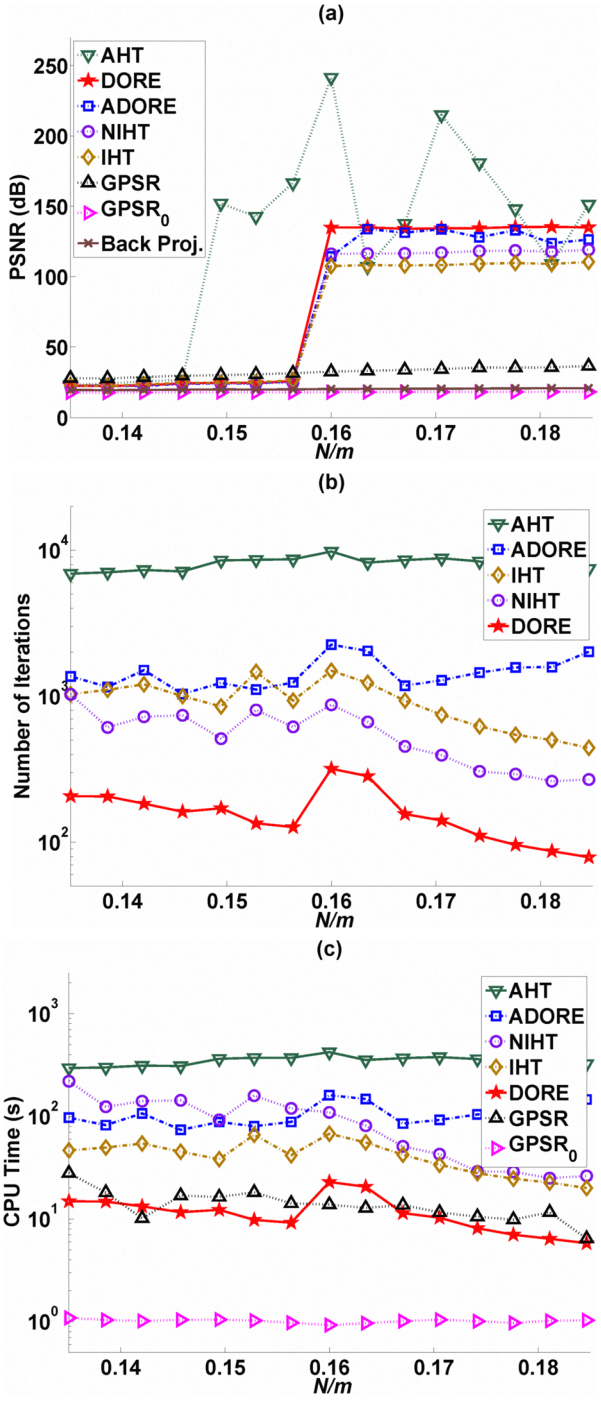

Fig. 2. (a) PSNR, (b) number of iterations, and (c) CPU time as functions of the normalized number of measurements N/m.

radial lines in our frequency-domain sampling pattern. In this example, all hard-thresholding methods have significantly sharper phase transitions than GPSR and outperform GPSR after the phase transitions. AHT exhibits the phase transition at $N/m \approx 0.15$, and the other hard thresholding methods at $N/m \approx 0.16$. Among all hard-thresholding methods, DORE needs the smallest number of iterations to converge and is also the fastest in terms of the CPU time. DORE needs 4.4 to 10.9 times less iterations than IHT and 2.3 to 6 times less iterations than NIHT; in terms of the CPU time, DORE is 2.7 to 6.7 times faster than IHT and 3.6 to 16.3 times faster than NIHT. Between the two automatic thresholding methods (AHT and ADORE), ADORE requires 3.7 to 7.7 times less iterations and is 2.2 to 4.6 times faster than AHT. The CPU times of DORE, IHT, and ADORE *per iteration* are approximately constant as N/m varies. One DORE step is about twice slower than one IHT step, validating the computational complexity analysis in Section III-A.

Since only a single choice (3.7) is used to initialize the DORE and ADORE methods, the PSNR curves for these methods in Fig. 2 (a) are only lower bounds on the PSNR performances achievable by these methods.

VI. Concluding Remarks

We developed double overrelaxation thresholding methods for reconstructing sparse signals from compressive samples and applied them to tomographic reconstruction from sparse projections. In large-scale problems, our DORE method is significantly faster than existing iterative hard thresholding methods and our ADORE scheme automatically estimates the signal sparsity level. Future research will include analyzing the speed of convergence of the DORE method.

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*, vol. 45, pp. 600–616, Mar. 1997.

[2] E.J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inform. Theory*, vol. 52, pp. 489–509, Feb. 2006.

[3] *IEEE Signal Processing Mag. Special Issue on Sensing, Sampling, and Compression*, Mar. 2008.

[4] A. Maleki and D.L. Donoho, "Optimally tuned iterative thresholding algorithms for compressed sensing," to appear in *IEEE J. Select. Areas Signal Processing*, vol. 4, 2010.

[5] A.M. Bruckstein, D.L. Donoho, and M. Elad, "From sparse solutions of systems of equations to sparse modeling of signals and images," *SIAM Review*, vol. 51, pp. 34–81, Mar. 2009.

[6] T. Blumensath and M.E. Davies, "Iterative hard thresholding for compressed sensing," *Appl. Comp. Harmonic Anal.*, vol. 27, pp. 265–274, Nov. 2009.

[7] T. Blumensath and M.E. Davies, "Normalised iterative hard thresholding; guaranteed stability and performance," to appear in *IEEE J. Select. Areas Signal Processing*, vol. 4, Apr. 2010.

[8] K.K. Herrity, A.C. Gilbert, and J.A. Tropp, "Sparse approximation via iterative thresholding," in *Proc. Int. Conf. Acoust., Speech, Signal Processing*, Toulouse, France, May 2006, pp. 624–627.

[9] A. Dogandžić and K. Qiu, "Automatic hard thresholding for sparse signal reconstruction from NDE measurements," in *Proc. Annu. Rev. Progress Quantitative Nondestructive Evaluation*, Kingston, RI, Jul. 2009. Available at http://home.eng.iastate.edu/~ald/publication.html.

[10] Y.X. He and C.H. Liu, "The dynamic ECME algorithm," Dept. Statistics, Purdue Univ., Tech. Report, 2009.

[11] 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*, vol. 1, pp. 586–597, Dec. 2007.

[12] G.J. McLachlan and T. Krishnan, *The EM Algorithm and Extensions*, 2nd. ed., New York: Wiley, 2008.

[13] C.H. Liu and D.B. Rubin, "The ECME algorithm: A simple extension of EM and ECM with fast monotone convergence," *Biometrika*, vol. 81, pp. 633–648, Dec. 1994.

[14] S.M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1993.

[15] R.A. Thisted, *Elements of Statistical Computing: Numerical Computation*, Chapman & Hall, 1988.