

ECME Hard Thresholding Methods for Image Reconstruction from Compressive Samples

Kun Qiu and Aleksandar Dogandžić

ECpE Department, Iowa State University, 3119 Coover Hall, Ames, IA

ABSTRACT

We propose two hard thresholding schemes for image reconstruction from compressive samples. 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 derived an expectation-conditional maximization either (ECME) iteration that converges to a local maximum of the likelihood function of the unknown parameters for a given image sparsity level. Here, we present and analyze a *double overrelaxation (DORE)* algorithm that applies two successive overrelaxation steps after one ECME iteration step, with the goal to accelerate the ECME iteration. To analyze the reconstruction accuracy, we introduce *minimum sparse subspace quotient (minimum SSQ)*, a more flexible measure of the sampling operator than the well-established restricted isometry property (RIP). We prove that, if the minimum SSQ is sufficiently large, the DORE algorithm achieves perfect or near-optimal recovery of the true image, provided that its transform coefficients are sparse or nearly sparse, respectively. We then describe a multiple-initialization DORE algorithm (DOREMI) that can significantly improve DORE's reconstruction performance. We present numerical examples where we compare our methods with existing compressive sampling image reconstruction approaches.

Keywords: Compressive sampling, expectation-conditional maximization either (ECME) algorithm, sparse signal reconstruction, sparse subspace quotient, successive overrelaxation, iterative hard thresholding, double overrelaxation (DORE) thresholding.

1. INTRODUCTION

Sparsity is an important concept in modern image processing. Most natural images can be accurately characterized by a few significant coefficients in some (e.g. discrete wavelet- or cosine-) transform domain, where the number of significant coefficients is much smaller than the image size. Therefore, for an $m \times 1$ vector \mathbf{x} representing the image (stacked in a long column vector for notational convenience) and an appropriate $m \times m$ orthogonal sparsifying transform matrix Ψ , we have

$$\mathbf{x} = \Psi \mathbf{s} \quad (1.1)$$

where \mathbf{s} is an $m \times 1$ transform-coefficient vector with most elements having negligible magnitudes. The idea behind compressive sampling or compressed sensing¹⁻⁶ is to directly *sense* the non-negligible components of \mathbf{s} using a small number of linear measurements:

$$\mathbf{y} = \Phi \mathbf{x} = \Phi \Psi \mathbf{s} \triangleq H \mathbf{s} \quad (1.2)$$

where \mathbf{y} is an $N \times 1$ measurement vector, Φ is a known $N \times m$ *sampling matrix* with $N \leq m$, and $H = \Phi \Psi$ is the $N \times m$ *sensing matrix*. We note that only Φ is employed in the data collection (sampling) process, whereas Ψ and H are needed only for the image reconstruction. The compressive sampling theory asserts that it is possible to accurately recover the sparse or compressible coefficient vector \mathbf{s} from the measurements \mathbf{y} *provided that* the sampling matrix Φ is *incoherent* with the sparsifying transform matrix Ψ .⁷ For example, random Gaussian

This work was supported by the National Science Foundation under Grant CCF-0545571.
E-mail: {kqiu,ald}@iastate.edu.

and Bernoulli sampling matrices are incoherent with any transform basis,⁵ whereas the random noiselet⁷ and structurally random sampling matrices⁸ are incoherent with the discrete wavelet transform (DWT) bases.

For noiseless measurements and strictly sparse transform coefficient vector, the major reconstruction task is to find the sparsest solution of an underdetermined linear system $\mathbf{y} = H \mathbf{s}$ (see e.g. [6, eq. (2)]):

$$(P_0) : \quad \min_{\mathbf{s}} \|\mathbf{s}\|_{\ell_0} \quad \text{subject to } \mathbf{y} = H \mathbf{s} \quad (1.3)$$

where $\|\mathbf{s}\|_{\ell_0}$ counts the number of nonzero elements in the signal coefficient vector \mathbf{s} . The (P_0) problem requires combinatorial search and is known to be NP-hard.⁹ Practical sparse reconstruction approaches include convex relaxation, greed pursuit, and probabilistic methods, see e.g. [10, Sec. I] for a brief survey. Among the existing sparse reconstruction methods, iterative hard thresholding (IHT)^{11,12} and normalized iterative hard thresholding (NIHT) algorithms¹³ are particularly appealing for image reconstruction due to their low memory and computational requirements per iteration, as well as theoretical and empirical evidence of good reconstruction performance. However, the IHT and NIHT methods

- (1) are sensitive to scaling of the sensing matrix (IHT) or require elaborate adjustments in each iteration to compensate for the scaling problem (NIHT) and
- (2) typically converge slowly, demanding a fairly large number of iterations.

We proposed a linear measurement model where the transform coefficient vector consists of a sparse deterministic component and a random Gaussian component and developed an *expectation-conditional maximization either (ECME) iteration* that converges to a local maximum of the likelihood function of the unknown parameters.^{14,15} In contrast with IHT and NIHT, the ECME iteration is universal (monitoring- and adjustment-free) and invariant to the scaling of H , thus successfully addressing (1) above.

In this paper, we present and analyze an *double overrelaxation (DORE)* thresholding scheme that interleaves two successive overrelaxation steps with ECME steps, see also [15, Sec. III]. DORE inherits invariance to the scaling of the sensing matrix from ECME. It also significantly accelerates the convergence of the ECME algorithm, thus providing a solution to (2) above. The line searches in the overrelaxation steps have *closed-form* solutions, making these steps computationally efficient. We prove that this DORE iteration monotonically converges to a local maximum of the marginal likelihood function under our probabilistic model. The conditions that we use in this convergence analysis are *invariant* to invertible linear transforms of the rows or the columns of the sensing matrix H , which indicates that the convergence of our DORE iteration is robust to the corresponding linear transforms of H . In contrast, the IHT algorithm converges to a local minimum of the squared residual error for a specified sparsity level *only* if H is appropriately scaled. Indeed, [11, Theorem 4] demands that the spectral norm of the sensing matrix H is less than unity. If the spectral norm condition is violated, the IHT iteration may become unstable and diverge.¹³

We derive guarantees for DORE image recovery that are not based on the common assumption that H has a sufficiently small restricted isometry constant, known as the *restricted isometry property (RIP)*;⁴ rather, we introduce a new measure of H useful for sparse signal reconstruction analysis: the minimum sparse subspace quotient (minimum SSQ). The minimum SSQ measures how close an arbitrary nonzero sparse vector of a certain sparsity level can get to the null space of H . Unlike the restricted isometry constant, the minimum SSQ is *invariant* to invertible linear transforms of the rows of H . We prove that, if the appropriate minimum SSQ of the sensing matrix is sufficiently large, our DORE algorithm provides perfect or near-optimal reconstruction of the underlying image, provided that its transform coefficients are sparse or nearly sparse, respectively. Finally, we propose a DORE scheme with multiple initial values (termed DOREMI), which finds a higher maximum point of the likelihood function and thereby improves the reconstruction performance compared with the standard DORE algorithm; the improvement is particularly significant for images with purely sparse transform coefficients, see Section 5.1.

In Section 2, we describe our probabilistic measurement model. The DORE algorithm is introduced in Section 3 and its convergence and recovery performance are analyzed in Sections 3.1 and 3.2, respectively. The DOREMI algorithm is proposed in Section 4. In Section 5, we compare the performances of the proposed and existing large-scale sparse signal reconstruction methods via numerical experiments. Concluding remarks are given in Section 6.

1.1 Notation and Terminology

We introduce the notation used in this paper: $\mathcal{N}(\mathbf{y}; \boldsymbol{\mu}, \Sigma)$ denotes the multivariate probability density function (pdf) of a real-valued Gaussian random vector \mathbf{y} with mean vector $\boldsymbol{\mu}$ and covariance matrix Σ ; $|\cdot|$, $\|\cdot\|_{\ell_p}$, $\det(\cdot)$, “ T ” denote the absolute value, ℓ_p norm, determinant, and transpose, respectively; the smallest integer larger than or equal to a real number x is $\lceil x \rceil$; I_n , $\mathbf{0}_{n \times 1}$, and $\mathbf{0}_{n \times m}$ are the identity matrix of size n , the $n \times 1$ vector of zeros, and the $n \times m$ matrix of zeros, respectively; $\dim(A)$ denotes the size of a set A ; $\text{supp}(\mathbf{x})$ returns the support set of a vector \mathbf{x} , i.e. the index set corresponding to the nonzero elements of \mathbf{x} , e.g. $\text{supp}([0, 1, -5, 0, 3, 0]^T) = \{2, 3, 5\}$; the thresholding operator $\mathcal{T}_r(\mathbf{x})$ keeps the r largest-magnitude elements of a vector \mathbf{x} intact and sets the rest to zero, e.g. $\mathcal{T}_2([0, 1, -5, 0, 3, 0]^T) = [0, 0, -5, 0, 3, 0]^T$.

We refer to an $N \times m$ sensing matrix H as *proper* if it has full rank and

$$N \leq m \quad (1.4)$$

which implies that the rank of H is equal to N . Throughout this paper, we assume that sensing matrices H are proper, which is satisfied in almost all practical sparse signal reconstruction scenarios.

2. MEASUREMENT MODEL

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

$$\mathbf{y} = H \mathbf{z} \quad (2.1a)$$

where H is an $N \times m$ real-valued proper sensing matrix, \mathbf{z} is an $m \times 1$ multivariate Gaussian vector with pdf

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

$\mathbf{s} = [s_1, s_2, \dots, s_m]^T$ is an *unknown* $m \times 1$ real-valued sparse signal vector containing *at most* r nonzero elements ($r \leq m$), and σ^2 is an unknown *variance-component parameter*; we refer to r as the *sparsity level* of the signal and to the signal \mathbf{s} as being *r-sparse*. Note that $\|\mathbf{s}\|_{\ell_0} = \dim(\text{supp}(\mathbf{s}))$ counts the number of nonzero elements in \mathbf{s} ; we refer to $\|\mathbf{s}\|_{\ell_0}$ as the *support size* of \mathbf{s} . Therefore, the support size $\|\mathbf{s}\|_{\ell_0}$ of a r -sparse vector \mathbf{s} is less than or equal to r .

The set of unknown parameters is

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

with the parameter space

$$\Theta_r = \mathcal{S}_r \times [0, +\infty) \quad (2.3a)$$

where

$$\mathcal{S}_r = \{\mathbf{s} \in \mathcal{R}^m : \|\mathbf{s}\|_{\ell_0} \leq r\} \quad (2.3b)$$

is the sparse signal parameter space. The marginal likelihood function of $\boldsymbol{\theta}$ is obtained by *integrating \mathbf{z} out* [see (2.1)]:

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

where the fact that H is a proper sensing matrix ensures that $\mathbf{H}\mathbf{H}^T$ is invertible and, consequently, that the pdf (2.4a) exists. For a given sparsity level r , the maximum likelihood (ML) estimate of $\boldsymbol{\theta}$ is

$$\hat{\boldsymbol{\theta}}_{\text{ML}} = (\hat{\mathbf{s}}_{\text{ML}}, \hat{\sigma}_{\text{ML}}^2) = \arg \max_{\boldsymbol{\theta} \in \Theta_r} p_{\mathbf{y}|\boldsymbol{\theta}}(\mathbf{y}|\boldsymbol{\theta}). \quad (2.4b)$$

Throughout this paper, we assume that the sparsity level r is *known*. Therefore, we simplify the notation and omit the dependence of the estimates of $\boldsymbol{\theta}$ on r .

For any fixed \mathbf{s} , the marginal likelihood (2.4a) is maximized by

$$\hat{\sigma}^2(\mathbf{s}) = (\mathbf{y} - H \mathbf{s})^T (H H^T)^{-1} (\mathbf{y} - H \mathbf{s}) / N. \quad (2.5)$$

Hence, maximizing (2.4a) with respect to $\boldsymbol{\theta}$ is equivalent to first maximizing the *concentrated likelihood function*

$$p_{\mathbf{y}|\boldsymbol{\theta}}(\mathbf{y} | \mathbf{s}, \hat{\sigma}^2(\mathbf{s})) = \frac{1}{\sqrt{\det(2\pi H H^T)}} [\hat{\sigma}^2(\mathbf{s})]^{-0.5N} \exp(-0.5N) \quad (2.6)$$

with respect to $\mathbf{s} \in \mathcal{S}_r$, yielding $\hat{\mathbf{s}}_{\text{ML}}$, and then determining the ML estimate of σ^2 by substituting $\hat{\mathbf{s}}_{\text{ML}}$ into (2.5). Obtaining the exact ML estimate $\boldsymbol{\theta}_{\text{ML}}$ in (2.4b) requires a combinatorial search and is therefore infeasible in practice. In the following section, we present a computationally feasible iterative approach that aims at maximizing (2.4a) with respect to $\boldsymbol{\theta} \in \Theta_r$ and circumvents the combinatorial search.

3. DORE ALGORITHM FOR KNOWN SPARSITY LEVEL

We now present our double overrelaxation (DORE) thresholding method that approximately finds the ML estimate in (2.4b), assuming a fixed sparsity level r . By applying two successive overrelaxation steps, DORE *accelerates* our expectation-conditional maximization either (ECME) thresholding algorithm,^{14,15} which treats \mathbf{z} as the *missing (unobserved) data* and 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.¹⁶

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

i. ECME step. Compute

$$\hat{\mathbf{s}} = \mathcal{T}_r(\mathbf{s}^{(p)} + H^T (H H^T)^{-1} (\mathbf{y} - H \mathbf{s}^{(p)})) \quad (3.1a)$$

$$\hat{\sigma}^2 = (\mathbf{y} - H \hat{\mathbf{s}})^T (H H^T)^{-1} (\mathbf{y} - H \hat{\mathbf{s}}) / N \quad (3.1b)$$

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

ii. First overrelaxation. Compute the linear combination of $\hat{\mathbf{s}}$ and $\mathbf{s}^{(p)}$:

$$\bar{\mathbf{z}} = \hat{\mathbf{s}} + \alpha_1 (\hat{\mathbf{s}} - \mathbf{s}^{(p)}) \quad (3.2a)$$

where the weight

$$\alpha_1 = \frac{(H \hat{\mathbf{s}} - H \mathbf{s}^{(p)})^T (H H^T)^{-1} (\mathbf{y} - H \hat{\mathbf{s}})}{(H \hat{\mathbf{s}} - H \mathbf{s}^{(p)})^T (H H^T)^{-1} (H \hat{\mathbf{s}} - H \mathbf{s}^{(p)})} \quad (3.2b)$$

is the *closed-form* solution of the line search:

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

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

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

$$\tilde{\mathbf{z}} = \bar{\mathbf{z}} + \alpha_2 (\bar{\mathbf{z}} - \mathbf{s}^{(p-1)}) \quad (3.3a)$$

where the weight

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

is the closed-form solution of the line search:

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

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

iv. Thresholding. Threshold $\tilde{\mathbf{z}}$ to the sparsity level r :

$$\tilde{\mathbf{s}} = \mathcal{T}_r(\tilde{\mathbf{z}}) \quad (3.4a)$$

compute the corresponding variance component estimate:

$$\tilde{\sigma}^2 = (\mathbf{y} - H\tilde{\mathbf{s}})^T (H H^T)^{-1} (\mathbf{y} - H\tilde{\mathbf{s}}) / N \quad (3.4b)$$

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

v. Decision (between ECME and thresholded overrelaxation parameter estimates). If $p_{\mathbf{y}|\boldsymbol{\theta}}(\mathbf{y} | \tilde{\boldsymbol{\theta}}) \geq p_{\mathbf{y}|\boldsymbol{\theta}}(\mathbf{y} | \hat{\boldsymbol{\theta}})$ or, equivalently, if

$$\tilde{\sigma}^2 < \hat{\sigma}^2 \quad (3.5)$$

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

Iterate until two consecutive sparse-signal estimates $\mathbf{s}^{(p)}$ and $\mathbf{s}^{(p+1)}$ do not differ significantly. Since $(H H^T)^{-1}$ can be pre-computed, our DORE iteration *does not* require matrix inversion. After Step **i**, we apply two overrelaxations (Steps **ii** and **iii**) that utilize the sparse signal estimates $\mathbf{s}^{(p)}$ and $\mathbf{s}^{(p-1)}$ from the two most recent completed DORE iterations, respectively. The goal of the overrelaxation steps is to boost the marginal likelihood (2.4a) and accelerate the convergence of the ECME iterations.¹⁵ The line searches in the two overrelaxation steps have closed-form solutions and are therefore computationally efficient.

If the rows of the sensing matrix H are orthonormal:

$$H H^T = I_N \quad (3.6)$$

Step **i** of the DORE scheme reduces to one *iterative hard-thresholding (IHT) step* in [12, eq. (10)].

Note that one ECME step (Step **i**) ensures monotonically nondecreasing marginal likelihood (2.4a).¹⁵ Step **v** ensures that the resulting new DORE parameter estimate $\boldsymbol{\theta}^{(p+1)}$ yields the marginal likelihood function that is higher than or equal to that of the standard ECME step (Step **i**). Therefore, the DORE iteration (3.1)–(3.5) ensures monotonically nondecreasing marginal likelihood between consecutive iteration steps:

$$p_{\mathbf{y}|\boldsymbol{\theta}}(\mathbf{y} | \boldsymbol{\theta}^{(p+1)}) \geq p_{\mathbf{y}|\boldsymbol{\theta}}(\mathbf{y} | \boldsymbol{\theta}^{(p)}). \quad (3.7)$$

DORE Initialization. The parameter estimates $\boldsymbol{\theta}^{(1)}$ and $\boldsymbol{\theta}^{(2)}$ are obtained by applying two consecutive ECME steps (3.1) to an initial sparse signal estimate $\mathbf{s}^{(0)}$.

The empirical Bayesian signal estimate. We construct the following empirical Bayesian estimate of the missing data vector \mathbf{z} :

$$\mathbb{E}_{\mathbf{z}|\mathbf{y}, \boldsymbol{\theta}}[\mathbf{z} | \mathbf{y}, \boldsymbol{\theta}^{(+\infty)}] = \mathbf{s}^{(+\infty)} + H^T (H H^T)^{-1} (\mathbf{y} - H \mathbf{s}^{(+\infty)}). \quad (3.8)$$

where $\boldsymbol{\theta}^{(+\infty)} = (\mathbf{s}^{(+\infty)}, (\sigma^2)^{(+\infty)})$ denotes the estimate of the unknown parameter set upon convergence of the DORE iteration. Unlike $\mathbf{s}^{(+\infty)}$, the empirical Bayesian estimate (3.8) is not r -sparse in general, and is therefore preferable for reconstructing images that have nearly sparse transform coefficients. Observe that

$$\mathbf{y} = H \mathbb{E}_{\mathbf{z}|\mathbf{y}, \boldsymbol{\theta}}[\mathbf{z} | \mathbf{y}, \boldsymbol{\theta}^{(+\infty)}] \quad (3.9)$$

implying that the empirical Bayesian estimate (3.8) always achieves zero squared residual error (unlike $\mathbf{s}^{(+\infty)}$).

3.1 Convergence Analysis

We first define the local maximum of a function over the parameter space \mathcal{S}_r in (2.3b).

Definition 1. (r -local maximum) For a function $f(\mathbf{s}) : \mathcal{R}^m \rightarrow \mathcal{R}$, a vector $\mathbf{s}^* \in \mathcal{S}_r$ is an r -local maximum point of $f(\mathbf{s})$ if there exists a $\delta > 0$, such that, for all $\mathbf{s} \in \mathcal{S}_r$ satisfying $\|\mathbf{s} - \mathbf{s}^*\|_{\ell_2} < \delta$, we have

$$f(\mathbf{s}^*) \geq f(\mathbf{s}).$$

Then, $f(\mathbf{s}^*)$ is the corresponding r -local maximum of $f(\mathbf{s})$.

Definition 1 states that an r -sparse vector is an r -local maximum point of a function $f(\mathbf{s})$ if, in some small neighborhood, this vector attains the largest function value among all sparse vectors within that small neighborhood. The following theorem states that our DORE algorithm for sparsity level r converges to an r -local maximum point of the concentrated marginal likelihood function (2.6).

Theorem 1. Assume that the sparsity level r satisfies

$$r \leq \frac{1}{2}(m - N) \quad (3.10)$$

and that the sensing matrix H satisfies the unique representation property (URP) stating that all $N \times N$ submatrices of H are invertible.¹⁷ Then, the DORE signal iterate $\mathbf{s}^{(p)}$ for sparsity level r converges monotonically to its fixed point, which corresponds to an r -local maximum point of the concentrated marginal likelihood function (2.6).

Note that (3.10) is a mild condition. Since $N \ll m$ in practice, (3.10) specifies a large range of sparsity levels r for which the DORE iteration converges. The conditions of Theorem 1 are invariant to the scaling of H and hold even when H is pre- or post-multiplied by full-rank square matrices of appropriate size.

3.2 Minimum SSQ and Near-optimal DORE Recovery

Now, we study the reconstruction performance of DORE via a new property of the sensing matrix H : minimum r -sparse subspace quotient, a more flexible measure than the well established restricted isometry property (RIP).⁴

Definition 2. (minimum r -SSQ) The minimum r -sparse subspace quotient (minimum r -SSQ) of a nonzero r -sparse vector \mathbf{s} of size $m \times 1$ (i.e. $\mathbf{s} \in \mathcal{S}_r \setminus \mathbf{0}_{m \times 1}$) and a proper $N \times m$ sensing matrix H is defined as

$$\rho_{r,\min}(H) \triangleq \min_{\mathbf{s} \in \mathcal{S}_r \setminus \mathbf{0}_{m \times 1}} \frac{\|H^T (H H^T)^{-1} H \mathbf{s}\|_{\ell_2}^2}{\|\mathbf{s}\|_{\ell_2}^2} = \min_{\mathbf{s} \in \mathcal{S}_r \setminus \mathbf{0}_{m \times 1}} \frac{\mathbf{s}^T H^T (H H^T)^{-1} H \mathbf{s}}{\mathbf{s}^T \mathbf{s}}. \quad (3.11)$$

The minimum r -SSQ in (3.11) is the smallest normalized squared magnitude of the projection of a nonzero r -sparse vector onto the row space of the sensing matrix H . Here, it is the *row space* of H that matters, rather than H itself. The minimum r -SSQ is invariant to invertible linear transforms of the rows of H , i.e.

$$\rho_{r,\min}(H) = \rho_{r,\min}(G H) \quad (3.12)$$

for any full-rank $N \times N$ matrix G . In contrast, the restricted isometry property (RIP), which is used extensively to analyze various sparse reconstruction methods,^{4,5,12,13} is vulnerable to the above transform. Specifically, the RIP condition requires that any subset of columns of H of certain size is nearly orthonormal: all columns of H should have approximately unit magnitudes and the correlation between distinct columns within the subset should be small. In comparison, due to the row transform invariance property, our minimum r -SSQ measure allows arbitrary nonzero column magnitudes and highly correlated columns.

We now give a geometric interpretation of $\rho_{r,\min}(H)$. Note that the *subspace quotient*

$$\frac{\|H^T (H H^T)^{-1} H \mathbf{s}\|_{\ell_2}^2}{\|\mathbf{s}\|_{\ell_2}^2} \quad (3.13)$$

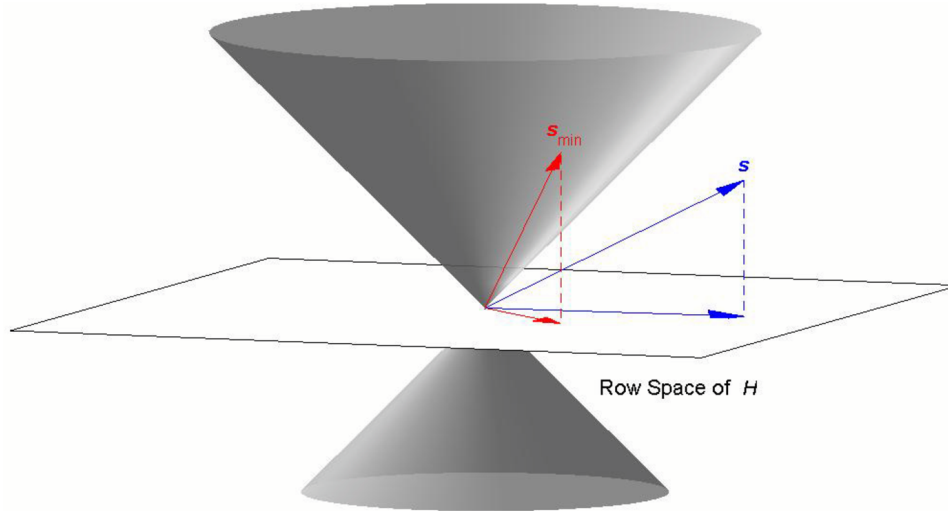


Figure 1. A geometric interpretation of minimum SSQ.

is simply the squared cosine of the angle between \mathbf{s} and the row space of H , see Fig. 1. The geometric interpretation of the subspace quotient is similar to that of the matched-subspace detector.¹⁸ Hence, the minimum r -SSQ is the smallest squared cosine between a nonzero r -sparse vector and the row space of H .

We have derived reconstruction performance guarantees for our DORE algorithm that are based on the minimum r -SSQ measure. These results are summarized in the following two theorems.

Theorem 2. (Exact Sparse Signal Reconstruction From Noiseless Samples) Suppose that we have collected a measurement vector

$$\mathbf{y} = H \mathbf{s}^\diamond \quad (3.14a)$$

where $\mathbf{s}^\diamond \in \mathcal{S}_r$ is an r -sparse signal vector, i.e. $\|\mathbf{s}^\diamond\|_{\ell_0} \leq r$. If the minimum $2r$ -SSQ of the sensing matrix H satisfies*

$$\rho_{2r, \min}(H) > 0.5 \quad (3.14b)$$

then the DORE iteration for the sparsity level r converges to the ML estimate of $\boldsymbol{\theta}$:

$$\hat{\boldsymbol{\theta}}_{\text{ML}} = (\mathbf{s}^\diamond, 0) \quad (3.14c)$$

and therefore recovers the true sparse signal \mathbf{s}^\diamond perfectly. The empirical Bayesian estimate (3.8) is equal to \mathbf{s}^\diamond , thus also achieving perfect reconstruction.

Theorem 2 shows that, upon convergence and if the minimum $2r$ -SSQ of the sensing matrix is sufficiently large, the DORE algorithm recovers the true sparse signal \mathbf{s}^\diamond perfectly from the noiseless measurements. In this case, the DORE iteration converges to the *global maximum* of the marginal likelihood (2.4a), which is infinitely large since the ML estimate of σ^2 is zero. This global convergence is guaranteed regardless of the initial estimate of $\boldsymbol{\theta}$ used to start the DORE iteration.

Next, we consider a more practical scenario where the signal \mathbf{s}^\diamond is not strictly sparse and the measurements \mathbf{y} are corrupted by noise.

Theorem 3. (Near-Optimal Non-sparse Signal Reconstruction From Noisy Samples) Suppose that we have collected a measurement vector

$$\mathbf{y} = H \mathbf{s}^\diamond + \mathbf{n} \quad (3.15a)$$

*The condition (3.14b) requires that all nonzero $2r$ -sparse vectors are within $\pm\pi/4$ from the row space of H , i.e. *outside* the double cone with aperture $\pi/2$, depicted in Fig. 1.

where the signal vector \mathbf{s}^\diamond is not necessarily sparse and $\mathbf{n} \in \mathcal{R}^N$ is a noise vector. Denote by \mathbf{s}_r^\diamond the best r -term ℓ_2 -norm approximation to \mathbf{s}^\diamond , i.e.

$$\mathbf{s}_r^\diamond = \arg \min_{\mathbf{s} \in \mathcal{S}_r} \|\mathbf{s} - \mathbf{s}^\diamond\|_{\ell_2} = \mathcal{T}_r(\mathbf{s}^\diamond) \quad (3.15b)$$

and by $\mathbf{s}^{(+\infty)}$ the r -sparse signal estimate obtained upon convergence of the DORE iteration for the sparsity level r . If the minimum $2r$ -SSQ of the sensing matrix H satisfies

$$\rho_{2r, \min}(H) > 0.5 \quad (3.15c)$$

which is the same as the condition (3.14b) in Theorem 2, then

$$\|\mathbf{s}^{(+\infty)} - \mathbf{s}_r^\diamond\|_{\ell_2} \leq \frac{\|\mathbf{s}^\diamond - \mathbf{s}_r^\diamond\|_{\ell_2} + \|H^T (HH^T)^{-1} \mathbf{n}\|_{\ell_2}}{\sqrt{\rho_{2r, \min}(H)} - \sqrt{1 - \rho_{2r, \min}(H)}}. \quad (3.15d)$$

Theorem 3 shows that, for a general (not necessarily sparse) signal vector \mathbf{s}^\diamond , noisy measurements satisfying (3.15a), and sensing matrix satisfying (3.15c), the DORE estimate is *close* to the best r -term ℓ_2 -norm approximation of \mathbf{s}^\diamond . (Note that \mathbf{s}_r^\diamond can be viewed as the r -term *transform-coding* estimate of \mathbf{s}^\diamond .) This result holds regardless of the initial estimate of $\boldsymbol{\theta}$ employed by the DORE iteration. Our analysis of the DORE method in Theorems 2 and 3 applies to the cases where the columns of H are not approximately orthonormal and can be heavily correlated, thus widening the class of sensing matrices for which it is possible to derive reconstruction performance guarantees.

4. DORE WITH MULTIPLE INITIAL VALUES (DOREMI)

The reconstruction performance of the DORE method can be improved by using multiple initial values. We now propose a DOREMI scheme that performs DORE multiple times using different initial values. The idea is to randomly perturb the best available signal estimate in the previous DORE iterations and use the obtained perturbed signal vector as an initial estimate for the new DORE iteration. We refer to each DORE iteration performed as a ‘run’, where i denotes the run index. We now describe the DOREMI scheme.

Initial stage: $i = 0$. Run the 0th DORE iteration initialized by $\mathbf{s}_0^{(0)} = \mathbf{0}_{m \times 1}$, which yields

$$\boldsymbol{\theta}_0^{(+\infty)} = \boldsymbol{\theta}^* = (\mathbf{s}^*, (\sigma^2)^*) \quad (4.1)$$

and $\mathbf{z}^* = E_{\mathbf{z} | \mathbf{y}, \boldsymbol{\theta}}[\mathbf{z} | \mathbf{y}, \boldsymbol{\theta}^*]$, see also (3.8).

Calibration stage: $i = 1, 2, \dots, L_P$. Perform L_P DORE iterations with the goal to select a good variance parameter for our random perturbations; we suggest setting

$$L_P = 7. \quad (4.2)$$

Run the i th DORE iteration initialized by

$$\mathbf{s}_i^{(0)} = \mathcal{T}_r(\mathbf{z}^* + \mathbf{e}_i) \quad (4.3)$$

which yields $\boldsymbol{\theta}_i^{(+\infty)}$, where the random perturbation terms \mathbf{e}_i are zero-mean Gaussian random vectors with covariance matrices

$$\text{cov}(\mathbf{e}_i) = 2^i (\sigma^2)^* I_m, \quad i = 1, 2, \dots, L_P. \quad (4.4)$$

Note that \mathbf{z}^* in (4.3) has been computed in the initial stage.

Upon completion of the above L_P DORE iterations, update $\boldsymbol{\theta}^* = (\mathbf{s}^*, (\sigma^2)^*)$ by selecting the best estimate of $\boldsymbol{\theta}$ found in all past runs [that achieves the largest marginal likelihood function (2.4a)] and compute $\mathbf{z}^* = E_{\mathbf{z} | \mathbf{y}, \boldsymbol{\theta}}[\mathbf{z} | \mathbf{y}, \boldsymbol{\theta}^*]$. Denote by $i_{\text{opt}} \in \{0, 1, \dots, L_P\}$ the run index that yields the best parameter estimate $\boldsymbol{\theta}^*$ and observe that

$$\boldsymbol{\theta}^* = \boldsymbol{\theta}_{i_{\text{opt}}}^{(+\infty)}.$$

The perturbation covariance matrix that we adopt in the main stage is based on i_{opt} , as described in the following.

Main stage: $i = L_P + 1, L_P + 2, \dots, L_M$. Perform $L_M - L_P$ DORE iterations with the goal to find a higher local maximum of the marginal likelihood function (2.4a) than the standard DORE algorithm; we suggest setting

$$L_M = 100. \quad (4.5)$$

The i th DORE iteration in the main stage is initialized by

$$\mathbf{s}_i^{(0)} = \mathcal{T}_r(\mathbf{z}^* + \mathbf{e}_i) \quad (4.6)$$

where the random perturbation terms \mathbf{e}_i are zero-mean Gaussian random vectors with covariance matrix

$$\text{cov}(\mathbf{e}_i) = 2^{i_{\text{opt}}} (\sigma^2)^* I_m. \quad (4.7)$$

Note that the multiplier $2^{i_{\text{opt}}}$ is *constant* (not a function of the run index i) within the main stage. At the end of each run within the main stage, we select $\boldsymbol{\theta}^* = (\mathbf{s}^*, (\sigma^2)^*)$ as the best estimate of $\boldsymbol{\theta}$ that achieves the largest marginal likelihood function (2.4a) and, if needed, update the corresponding empirical Bayesian missing-data estimate $\mathbf{z}^* = \mathbb{E}_{\mathbf{z}|\mathbf{y}, \boldsymbol{\theta}}[\mathbf{z} | \mathbf{y}, \boldsymbol{\theta}^*]$. Hence, in the main stage, \mathbf{z}^* is allowed to vary between consecutive DORE runs.

To have more than the $L_M + 1$ runs described above, reset $i = 1$ and go back to the calibration stage, using the best $\boldsymbol{\theta}^*$ and corresponding \mathbf{z}^* found so far, with $\boldsymbol{\theta}_0^{(+\infty)} = \boldsymbol{\theta}^*$. Here, we wish to ensure that the calibration stage is repeated every L_M runs. The process ceases when a specified total number of runs is reached.

5. NUMERICAL EXAMPLES

We compare our proposed methods with existing large-scale sparse reconstruction techniques using two image recovery experiments, with purely and nearly sparse transform coefficients, respectively. In particular, we compare

- the ECME^{14,15} scheme, initialized by the zero sparse signal estimate:

$$\mathbf{s}^{(0)} = \mathbf{0}_{m \times 1}; \quad (5.1)$$

- the DORE scheme in Section 3, initialized by the zero sparse signal estimate (5.1), with MATLAB implementation available at <http://home.eng.iastate.edu/~ald/DORE.htm>;
- the DOREMI scheme in Section 4, where we choose $L_P = 7$, $L_M = 100$ [as suggested in (4.2) and (4.5)], and the total of 200 runs;
- the automatic double overrelaxation (ADORE) scheme in [15, Section IV] that selects the signal sparsity level from the data, initialized by the zero sparse signal estimate and with search resolution set to $L = 500$;
- the NIHT¹³ scheme, initialized by the zero sparse signal estimate (5.1);
- the debiased gradient-projection for sparse reconstruction method in [19, Sec. III.B] with the convergence threshold $\text{tolP} = 10^{-5}$ and regularization parameter set to
 - (i) $\tau = 0.1 \|H^T \mathbf{y}\|_{\ell_\infty}$, suggested in [19, e.q. (22)] (labeled GPSR₀) and
 - (ii) $\tau = 0.001 \|H^T \mathbf{y}\|_{\ell_\infty}$, obtained by *manual tuning* for good performance in the following two numerical examples (labeled GPSR);
- the minimum-norm signal estimate (labeled MN):

$$\hat{\mathbf{s}}_{\text{MN}} = H^T (H H^T)^{-1} \mathbf{y} \quad (5.2)$$

which achieves zero squared residual error by ignoring signal sparsity.

For the ECME, DORE, DOREMI, ADORE, and NIHT iterations, we use the following convergence criterion:

$$\|\mathbf{s}^{(p+1)} - \mathbf{s}^{(p)}\|_{\ell_2}^2 / m < 10^{-14}. \quad (5.3)$$

To implement the NIHT scheme, we incorporated the convergence criterion (5.3) into the corresponding MATLAB codes from the `sparsify` toolbox at <http://www.see.ed.ac.uk/~tblumens/sparsify/sparsify.html>.

The sensing matrix H has the following structure (see e.g. [1, eq. (2) and Fig. 1]):

$$H = \Phi \Psi \quad (5.4)$$

where Φ is an $N \times m$ *sampling matrix* and Ψ is an appropriate $m \times m$ orthogonal *sparsifying transform matrix*. In our examples presented here, Φ are *structurally random sampling matrices*⁸ and Ψ are inverse discrete wavelet transform (DWT) matrices. Under these choices of Φ and Ψ , the rows of H are orthonormal, i.e. (3.6) holds.

For an underlying image $\Psi \mathbf{s}$, the signal vector \mathbf{s} is the wavelet coefficient vector of the image. Our performance metric is the peak signal-to-noise ratio (PSNR) of a reconstructed image $\Psi \hat{\mathbf{s}}$, where $\hat{\mathbf{s}}$ is the estimated wavelet coefficients vector:

$$\text{PSNR (dB)} = 10 \log_{10} \left\{ \frac{[(\Psi \mathbf{s})_{\text{MAX}} - (\Psi \mathbf{s})_{\text{MIN}}]^2}{\|\Psi \hat{\mathbf{s}} - \Psi \mathbf{s}\|_{\ell_2}^2 / m} \right\} = 10 \log_{10} \left\{ \frac{[(\Psi \mathbf{s})_{\text{MAX}} - (\Psi \mathbf{s})_{\text{MIN}}]^2}{\|\hat{\mathbf{s}} - \mathbf{s}\|_{\ell_2}^2 / m} \right\} \quad (5.5)$$

where $(\Psi \mathbf{s})_{\text{MIN}}$ and $(\Psi \mathbf{s})_{\text{MAX}}$ denote the smallest and largest elements of $\Psi \mathbf{s}$.

5.1 Shepp-Logan Phantom Reconstruction

Consider the reconstruction of the Shepp-Logan phantom of size $m = 256^2$ in Fig. 2 (a) from structurally random compressive samples. In this example, we select the inverse Haar (Daubechies-2) DWT matrix to be the orthogonal sparsifying transform matrix Ψ . The true signal vector \mathbf{s} consists of the Haar wavelet transform coefficients of the phantom and is sparse, with $\|\mathbf{s}\|_{\ell_0} = 3769 \approx 0.06 m$.

For all hard thresholding methods (ECME, DORE, DOREMI, ADORE, and NIHT), we estimate the transform coefficient vector \mathbf{s} using the sparse signal estimates obtained upon their convergence.

ECME, DORE, DOREMI, and NIHT require knowledge of the signal sparsity level r ; in this example, we set r to the true signal support size:

$$r = 3769. \quad (5.6)$$

In contrast, ADORE is *automatic* and *estimates* r from the measurements.

Figs. 2 (b)–(d) show the PSNRs, numbers of iterations, and CPU times of various methods as functions of the normalized number of measurements (subsampling factor) N/m . The reconstructions having PSNR higher than 100 dB achieve almost perfect recovery. In this example, all hard-thresholding methods have significantly sharper phase transitions than the tuned GPSR, and outperform the tuned GPSR after the phase transitions. DOREMI exhibits the phase transition around $N/m = 0.17$; the phase transitions of the other hard thresholding methods occur around $N/m = 0.2$. Furthermore, after the phase transition, DOREMI settles at a higher PSNR than the other methods. Note that ADORE, which estimates the sparsity level from the data, performs as well as the ECME, DORE, and NIHT methods that require prior knowledge of the signal sparsity level. 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. For the subsampling factor $N/m = 0.18$, DORE needs 8.6 times less iterations than ECME and 3.9 times less iterations than NIHT; in terms of the CPU time, DORE is about 5.5 times faster than both ECME and NIHT. In this example, DORE is also faster than the tuned GPSR method.

The importance of tuning the GPSR's regularization parameter τ is evident from Fig. 2 (b): GPSR_0 reconstructs the underlying image poorly, with reconstruction performance inferior even to that of the MN method.

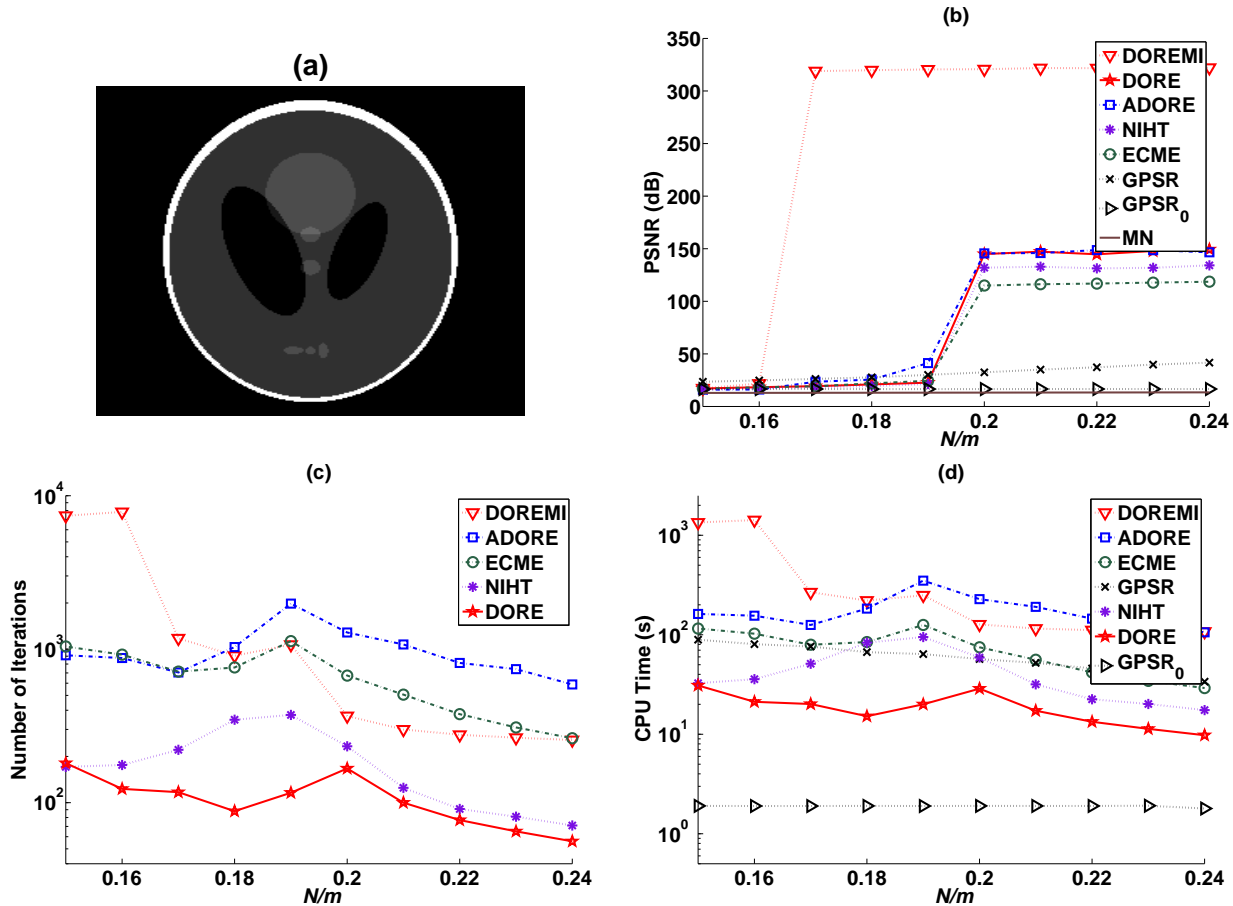


Figure 2. (a) The 256×256 Shepp-Logan phantom and (b) PSNR, (c) number of iterations, and (d) CPU time for its reconstruction as functions of the normalized number of measurements N/m .

5.2 Lena Reconstruction

We now reconstruct the standard Lena image of size $m = 512^2$ in Fig. 3 (a) from structurally random compressive samples. In this example, we select the inverse Daubechies-6 DWT matrix to be the orthogonal sparsifying transform matrix Ψ . The wavelet coefficients of the Lena image are only nearly sparse.

For all hard thresholding methods, we estimate the transform coefficient vector \mathbf{s} using the empirical Bayesian estimate (3.8), with $\theta^{(+\infty)}$ equal to the parameter estimates obtained upon their convergence.

To implement the ECME, DORE, DOREMI, and NIHT iterations, we chose the sparsity level as follows:

$$r = 50000 \frac{N}{m} \approx 0.19 N. \quad (5.7)$$

Figs. 3 (b)–(d) show the PSNRs, numbers of iterations, and CPU times of various methods as functions of the normalized number of measurements N/m . As a benchmark, we also show the PSNR of the transform-coding signal estimate $\mathcal{T}_r(\mathbf{s})$, which is the best r -term ℓ_2 -norm approximation of the true wavelet coefficient vector \mathbf{s} . We do not display the PSNRs of the MN estimates (5.2) because they are poor: they vary between 14.21 dB and 16.24 dB for the range of N/m in Figs. 3 (b)–(d).

The importance of tuning the GPSR's regularization parameter τ is evident from Fig. 3 (b): GPSR₀ reconstructs the underlying image poorly compared with the tuned GPSR method.

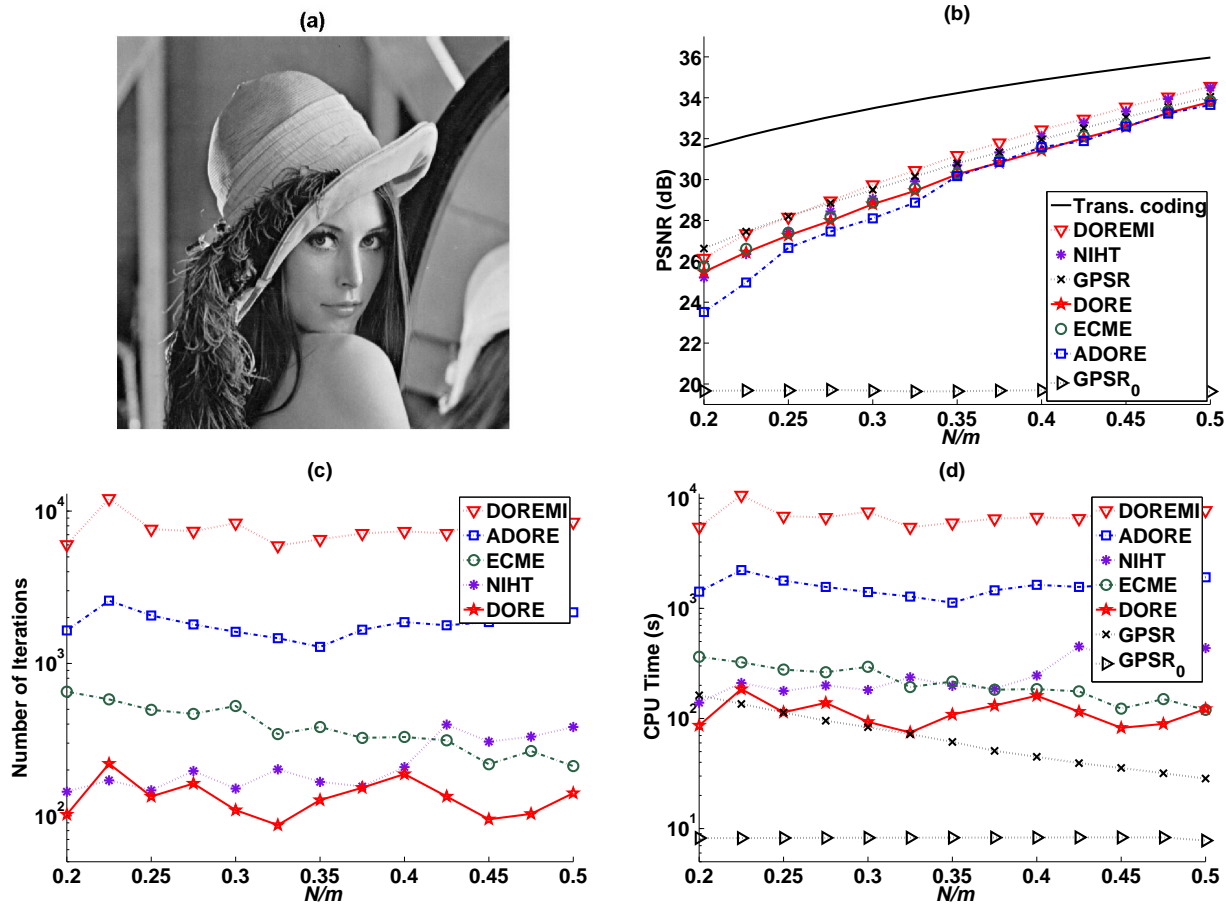


Figure 3. (a) The 512×512 Lena image and (b) PSNR, (c) number of iterations, and (d) CPU time for its reconstruction as functions of the normalized number of measurements N/m .

Unlike the phantom reconstruction example in Section 5.1, here the underlying signal is *not* strictly sparse and, consequently,

- the achieved PSNRs of all methods are significantly smaller than in the phantom example, even though the subsampling factor N/m ranges over a fairly wide interval;
- the difference in reconstruction accuracy between the tuned convex GPSR method and hard thresholding methods is significantly smaller than in the phantom example: compare Fig. 2 (b) with Fig. 3 (b).

In this example, DOREMI achieves the highest PSNR among all hard thresholding methods and also outperforms GPSR when $N/m \geq 0.25$. Interestingly, for the sparsity level r in (5.7), the PSNR of the transform-coding signal estimate $\mathcal{T}_r(s)$ varies between 31.6 dB and 36 dB for the range of N/m in Figs. 3 (b)–(d). When $N/m > 0.4$, DOREMI achieves PSNRs that are within 2.5 dB from those of the transform-coding signal estimates. For $N/m > 0.35$, the ADORE method, which estimates the sparsity level from the data, performs as well as the ECME, DORE, and NIHT methods for the manually tuned sparsity levels in (5.7).

Figs. 3 (c) and (d) show that DORE requires the smallest number of iterations and CPU time among the hard thresholding methods. In terms of the CPU time, DORE is up to 4.2 times faster than ECME and up to 4.3 times faster than NIHT in this example. When $N/m < 0.35$, the CPU times of DORE and tuned GPSR are comparable.

6. CONCLUDING REMARKS

We presented and analyzed double overrelaxation ECME hard thresholding schemes for sparse signal reconstruction. The DORE algorithm applies two consecutive overrelaxation steps with the goal to accelerate the convergence of the ECME iteration. We showed that, under certain mild conditions, DORE converges monotonically to a local maximum of the concentrated marginal likelihood function. To develop the near-optimal recovery results for DORE, we introduced new measures of the sensing matrix: sparse subspace quotient (SSQ) and minimum SSQ. We proposed a DOREMI algorithm that runs DORE multiple times and found that such multiple initialization brings particularly significant improvement when reconstructing images that have purely sparse transform coefficients.

Further research will include developing hard thresholding sparse signal reconstruction methods for quantized measurements.

REFERENCES

- [1] Baraniuk, R. G., “Compressive sensing,” *IEEE Signal Processing Mag.* **24**, 118–121 (July 2007).
- [2] [*IEEE Signal Processing Mag. Special Issue on Sensing, Sampling, and Compression*] (Mar. 2008).
- [3] Candès, E. J., Romberg, J., and Tao, T., “Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information,” *IEEE Trans. Inform. Theory* **52**, 489–509 (Feb. 2006).
- [4] Candès, E. J. and Tao, T., “Decoding by linear programming,” *IEEE Trans. Inform. Theory* **51**, 4203–4215 (Dec. 2005).
- [5] Candès, E. J. and Tao, T., “Near-optimal signal recovery from random projections: Universal encoding strategies?” *IEEE Trans. Inform. Theory* **52**, 5406–5425 (Dec. 2006).
- [6] Bruckstein, A. M., Donoho, D. L., and Elad, M., “From sparse solutions of systems of equations to sparse modeling of signals and images,” *SIAM Review* **51**, 34–81 (Mar. 2009).
- [7] Candès, E. J. and Romberg, J., “Sparsity and incoherence in compressive sampling,” *Inverse Problems* **23**, 969–985 (Apr. 2007).
- [8] Do, T. T., Tran, T. D., and Gan, L., “Fast compressive sampling with structurally random matrices,” in [*Proc. Int. Conf. Acoust., Speech, Signal Process.*], 3369–3372 (Apr. 2008).
- [9] Natarajan, B. K., “Sparse approximate solutions to linear systems,” *SIAM J. Comput.* **24**, 227–234 (1995).
- [10] Qiu, K. and Dogandžić, A., “Variance-component based sparse signal reconstruction and model selection,” *IEEE Trans. Signal Processing* **58**, 2935–2952 (June 2010).
- [11] Blumensath, T. and Davies, M. E., “Iterative thresholding for sparse approximations,” *J. Fourier Anal. Appl.* **14**, 629–654 (Dec. 2008).
- [12] Blumensath, T. and Davies, M. E., “Iterative hard thresholding for compressed sensing,” *Appl. Comp. Harmonic Anal.* **27**, 265–274 (Nov. 2009).
- [13] Blumensath, T. and Davies, M. E., “Normalized iterative hard thresholding: Guaranteed stability and performance,” *IEEE J. Select. Areas Signal Processing* **4**, 298–309 (Apr. 2010).
- [14] Dogandžić, A. and Qiu, K., “Automatic hard thresholding for sparse signal reconstruction from NDE measurements,” in [*Rev. Progress Quantitative Nondestructive Evaluation*], Thompson, D. O. and Chimenti, D. E., eds., **29**, 806–813, Amer. Inst. Phys., Melville NY (2010).
- [15] Qiu, K. and Dogandžić, A., “Double overrelaxation thresholding methods for sparse signal reconstruction,” in [*Proc. 44rd Annu. Conf. Inform. Sci. Syst.*], (2010).
- [16] McLachlan, G. J. and Krishnan, T., [*The EM Algorithm and Extensions*], Wiley, New York, second ed. (2008).
- [17] Gorodnitsky, I. F. and Rao, B. D., “Sparse signal reconstruction from limited data using FOCUSS: A re-weighted minimum norm algorithm,” *IEEE Trans. Signal Processing* **45**, 600–616 (Mar. 1997).
- [18] Scharf, L. L., Kraut, S., and McCloud, M. L., “A review of matched and adaptive subspace detectors,” in [*Proc. Adapt. Sys. Sig. Proc. Comm. Control Symp.*], 82–86 (Oct. 2000).
- [19] Figueiredo, M. A. T., Nowak, R. D., and Wright, S. J., “Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems,” *IEEE J. Select. Areas Signal Processing* **1**, 586–597 (Dec. 2007).

Copyright 2010 Society of Photo-Optical Instrumentation Engineers. One print or electronic copy may be made for personal use only. Systematic reproduction and distribution, duplication of any material in this paper for a fee or for commercial purposes, or modification of the content of the paper are prohibited.