Greedy Algorithms for Joint Sparse Recovery

Jeffrey D. Blanchard, Michael Cermak, David Hanle, and Yirong Jing, *Grinnell College*

***Abstract*—Five known greedy algorithms designed for the single measurement vector setting in compressed sensing and sparse approximation are extended to the multiple measurement vector scenario: Iterative Hard Thresholding (IHT), Normalized IHT (NIHT), Hard Thresholding Pursuit (HTP), Normalized HTP (NHTP), and Compressive Sampling Matching Pursuit (CoSaMP). Using the asymmetric restricted isometry property (ARIP), sufficient conditions for all five algorithms establish bounds on the discrepancy between the algorithms' output and the optimal row-sparse representation. When the initial multiple measurement vectors are jointly sparse, ARIP-based guarantees for exact recovery are also established. The algorithms are then compared via the recovery phase transition framework. The strong phase transitions describing the family of Gaussian matrices which satisfy the sufficient conditions are obtained via known bounds on the ARIP constants. The algorithms' empirical weak phase transitions are compared for various numbers of multiple measurement vectors. Finally, the performance of the algorithms is compared against a known rank aware greedy algorithm, Rank Aware Simultaneous Orthogonal Matching Pursuit + MUSIC. Simultaneous recovery variants of NIHT, NHTP, and CoSaMP all outperform the rank-aware algorithm.**

***Index Terms*—Compressed sensing, greedy algorithms, multiple measurement vectors, joint sparsity, row sparse matrices, performance comparison**

I. Introduction

A. Joint Sparse Recovery of Multiple Measurement Vectors

The single measurement vector (SMV) formulation is now standard in sparse approximation and compressed sensing literature. For $m < n$, $x \in \mathbb{R}^n$, $A \in \mathbb{R}^{m\times n}$, and $y = Ax \in \mathbb{R}^m$, one seeks to recover the signal or vector x from the measurements y when the linear measurement process defined by A is known. While this problem is NP-hard in general [1], if A is chosen wisely and x is sparse, several reconstruction algorithms are known to guarantee exact recovery of x. When x is not exactly sparse, but instead has a good sparse approximation, or when the measurements y are corrupted by noise, bounds on the recovery error are also known.

A natural extension of this problem is the multiple measurement vector (MMV) problem where a single matrix A is utilized to obtain measurements of multiple signals: $y_1 = Ax_1, y_2 = Ax_2, \dots, y_l = Ax_l$. Rather than recovering the l signals separately, one attempts to simultaneously recover all l signals from the matrix formulation $Y = AX$ where $X = [x_1|x_2|\cdots|x_l]$ and thus $Y = [y_1|y_2|\cdots|y_l]$. When the target signals, $\{x_i\}_{i=1}^l$, are all predominantly supported on a common support set, this approach can lead to a computational advantage [2], [3]. If the cost per iteration of one run of the simultaneous recovery algorithm is no worse than l runs of an equivalent SMV algorithm, the common support set provides more information to the simultaneous recovery algorithm than running l independent instances of an SMV algorithm.

B. Prior Art and Contributions

Beginning with Leviatan, Lutoborski, and Temlyakov [4], [5], [6], a substantial body of work has been developed for the MMV problem including [7], [8], [9], [10], [11], [12], [13], [14]. The majority of the literature focuses on relaxations, mixed matrix norm techniques, and variants of orthogonal matching pursuit. Tropp et al. [3], [15] introduced simultaneous recovery algorithms based on Orthogonal Matching Pursuit (OMP) and convex relaxation. For the greedy algorithm, Simultaneous OMP (SOMP), Tropp et al. stated that the analysis of the MMV recovery algorithm permitted a straightforward extension of the analysis from the SMV setting. Foucart applied these "capitalization" techniques to Hard Thresholding Pursuit (HTP) to extend that algorithm to the MMV setting [2], [16].

In this article, we provide a comprehensive investigation of the extension to the MMV problem of five known greedy algorithms designed for the SMV setting: Iterative Hard Thresholding (IHT) [17], Normalized IHT (NIHT) [18], Hard Thresholding Pursuit (HTP) [16], Normalized HTP (NHTP) [16], and Compressive Sampling Matching Pursuit (CoSaMP) [19]. The article includes:

- a description of the simultaneous joint sparse recovery algorithms (Section II-B);
- sufficient conditions based on the asymmetric restricted isometry property which guarantee joint sparse recovery and bound recovery error for joint sparse approximation (Section II-C);
- a quantitative comparison of the theoretical sufficient conditions through the strong recovery phase transition framework (Section III-A);
- an empirical, average case performance comparison through the weak recovery phase transition framework (Section III-B);
- an empirical, average case performance comparison against a known rank-aware algorithm RA-SOMP+MUSIC (Section III-C).

The MMV algorithms Simultaneous IHT (SIHT), Simultaneous NIHT (SNIHT), Simultaneous HTP (SHTP)[1], Simultaneous NHTP (SNHTP), and Simultaneous CoSaMP (SCoSaMP) are natural extensions of the well-known SMV

This work was supported by grant NSF DMS 11126152 and the Grinnell College MAP program. The authors are with the Department of Mathematics and Statistics, Grinnell College, Grinnell, IA 50112.

Manuscript submitted July 2013; accepted January 2014.

[1]This algorithm and its associated convergence guarantee were originally presented by Foucart [2]

versions of the algorithms and reduce to the SMV versions when applied to the measurements of a single sparse vector. While the analysis closely follows the MMV extension techniques of Tropp et al. [3], [15] and the proofs closely follow the analysis of Foucart for the SMV versions of the algorithms [20], the convergence analysis provides three generalizations. The results are written in terms of the asymmetric restricted isometry constants [21] thereby providing weaker sufficient conditions than those derived with the standard, symmetric restricted isometry constants. Since empirical testing [22] suggests tuning the step size in SIHT and SHTP according to family from which A is drawn, the analysis permits an arbitrary fixed step size between 0 and 1. Finally, the results for the normalized algorithms NIHT and NHTP are stated explicitly.

These sufficient conditions are quantitatively compared by employing the techniques for the strong recovery phase transition framework of [21], [23]. The strong phase transitions associated with the sufficient conditions identify two important facts. First, simpler algorithms often admit a simpler analysis which yield more relaxed sufficient conditions even though the algorithms may have inferior observed performance. Second, the sufficient conditions obtained via the restricted isometry property are exceedingly pessimistic and apply to a regime of problems unlikely to be realized in practice. While critical to understanding the theoretical behavior of the algorithms, the pessimistic, worst-case sufficient conditions fail to inform practitioners about typical algorithm behavior. From this point of view, the empirical average-case performance comparisons provide the most important information for selecting an algorithm for application.

C. Organization

The algorithms are detailed in Section II-B with the joint sparse recovery guarantees provided in Section II-C. In Section III-A, the theoretical sufficient conditions for each of the algorithms are compared via the strong phase transition framework [21], [23]. In Section III-B, the average case performance of the algorithms is then compared via empirical weak recovery phase transitions similar to other empirical studies [22], [24]. In Section III-C the typical performance of these "rank blind" algorithms is then juxtaposed with the performance of the "rank aware" greedy algorithm Rank Aware SOMP + MUSIC [9], [25], [12].

As the convergence analysis leading to Theorem 1 closely follows the techniques of Foucart [20], a representative proof for SIHT and SNHTP is provided in Appendix A. For completeness, all omitted proofs are available in the supplementary material [26]. The supplementary material also includes the analysis required to employ the strong phase transition techniques of [23] and additional empirical performance comparisons with measurements obtained from randomly subsampled discrete cosine transforms.

II. Recovery Guarantees

A. Notation

Let $M(r,c)$ denote the set of matrices with r rows and c columns with entries drawn from $\mathbb{R}$ or $\mathbb{C}$. If X is a collection of l vectors in $\mathbb{R}^n$ or $\mathbb{C}^n$, then $X \in M(n,l)$ and we let $X_{(i)}$ denote the ith row of X while X_i represents the ith column. Let $S \subset \{1,\dots,n\}$ be an index set and define $X_{(S)}$ as the matrix X restricted to the rows indexed by this set; in other words, the entries in the rows indexed by S remain unchanged while all other rows of $X_{(S)}$ have all entries set to 0. The linear measurement process is defined by a matrix $A \in M(m,n)$ and the restriction A_S represents the sub-matrix of A obtained by selecting the columns of A indexed by S. A^* denotes the conjugate transpose of A.

Throughout the manuscript, the *row support*, or simply *support*, of a matrix $Z \in M(n,l)$ is the index set of rows which contain nonzero entries. Thus, when X is a collection of l column vectors, $X = [X_1|X_2|\dots|X_l]$, we have

$$\text{supp}(X) = \bigcup_{i=1}^{l} \text{supp}(X_i).$$

The matrix X is k*-row sparse* (or the set $\{X_i : i = 1,\dots,l\}$ is *jointly* k*-sparse*) if $|\text{supp}(X)| \le k$. In particular, if $|S|$ is the cardinality of the index set S, then $X_{(S)}$ is $|S|$-row sparse. Let $\chi_{n,l}(k) \subset M(n,l)$ be the subset of k-row sparse $n\times l$ matrices; the set of k-sparse column vectors will be abbreviated $\chi_n(k)$.

The MMV sparse approximation problem is equivalent to constructing a row sparse approximation of a matrix X from the measurements $Y = AX$. Consider first the ideal case of measuring a k-row sparse matrix $X \in \chi_{n,l}(k)$ where $T = \text{supp}(X)$. Given the measurements $Y = AX \in M(m,l)$, the task is to exactly recover the k-row sparse matrix $X = X_{(T)}$. This is equivalent to simultaneously recovering l jointly k-sparse vectors. This ideal setting of attempting to recover a perfectly row sparse matrix from clean measurements is unlikely to present itself in applications. Instead, the task will be to find an accurate row sparse approximation to a matrix $X \in M(n,l)$. Suppose T is the index set of rows of $X \in M(n,l)$ which have the k largest row-ℓ_2-norms, and the measurement process is corrupted by additive noise, namely $Y = AX + E$ for some noise matrix $E \in M(m,l)$. The row sparse approximation problem seeks an approximation to $X_{(T)}$. The recovery guarantees are presented in terms of the Frobenius norm of the discrepancy between the algorithms' output $\hat{X}$ and the optimal k-row sparse approximation $X_{(T)}$. The Frobenius norm of a matrix $X \in M(n,l)$ is defined by

$$\|X\|_F^2 = \sum_{j=1}^{l} \|X_j\|_2^2 = \sum_{j=1}^{l}\sum_{i=1}^{n} |X_{i,j}|^2 .$$

B. Greedy MMV Algorithms

To solve the MMV or row sparse approximation problem, we propose the extension of five popular greedy algorithms designed for the SMV problem: IHT, NIHT, HTP, NHTP, and CoSaMP. Each of these algorithms is a support identification algorithm. The simultaneous recovery algorithms, prefixed with the letter S, are defined in Algorithms 1–3. Each algorithm follows the same initialization procedure. The initial approximation matrix is the zero matrix $X^0 = 0$ and thus the initial residual is the matrix of input measurements $R^0 = Y$. When an initial proxy for the support set is needed, $T^0 =$

`DetectSupport`(A^*Y, k) where `DetectSupport`(Z, s) is a subroutine identifying the index set of the rows of Z with the s largest row-ℓ_2-norms. In Algorithms 1 and 3, the thresholding operator `Threshold`(Z, S) restricts the matrix Z to the row index set S, i.e. $Z_{(S)} = $ `Threshold`(Z, S). The choice of stopping criteria plays an important role for the algorithms, and the stopping criteria employed for the empirical testing are outlined in Section III-B.

Algorithm 1 SIHT / SNIHT

for iteration j **until stopping criteria do**
  if (SIHT) **then**
    $\omega^j = \omega$
  else if (SNIHT) **then**
    $\omega^j = \frac{\|(A^*R^{j-1})_{(T^{j-1})}\|_F^2}{\|A_{T^{j-1}}(A^*R^{j-1})_{(T^{j-1})}\|_F^2}$
  end if
  $X^j = X^{j-1} + \omega^j \left(A^*R^{j-1}\right)$
  $T^j = \texttt{DetectSupport}(X^j, k)$
  $X^j = \texttt{Threshold}(X^j, T^j)$
  $R^j = Y - AX^j$
end for
return $\hat{X} = X^{j^\star}$ when stopping at iteration $j^\star$.

Algorithm 2 SHTP / SNHTP

for iteration j **until stopping criteria do**
  if (SHTP) **then**
    $\omega^j = \omega$
  else if (SNHTP) **then**
    $\omega^j = \frac{\|(A^*R^{j-1})_{(T^{j-1})}\|_F^2}{\|A_{T^{j-1}}(A^*R^{j-1})_{(T^{j-1})}\|_F^2}$
  end if
  $X^j = X^{j-1} + \omega^j \left(A^*R^{j-1}\right)$
  $T^j = \texttt{DetectSupport}(X^j, k)$
  $X^j = \arg\min\{\|Y - AZ\|_F : \text{supp}(Z) \subseteq T^j\}$
  $R^j = Y - AX^j$
end for
return $\hat{X} = X^{j^\star}$ when stopping at iteration $j^\star$.

In iteration j, SIHT and SHTP update the previous approximation X^{j-1} by taking a step of predefined, fixed length ω in the steepest descent direction A^*R^{j-1}. A new proxy for the support set, T^j, is then obtained by selecting the rows of X^j with greatest row-ℓ_2-norms. The two algorithms differ in how the support proxy T^j is utilized: SIHT employs a hard thresholding operator which restricts the approximation X^j to the rows indexed by T^j while SHTP projects the measurements Y onto the support set T^j.

The normalized variants of these two algorithms, SNIHT and SNHTP, proceed in a nearly identical fashion although the potentially inaccurate fixed step size is replaced by a near-optimal step size ω^j. If $T^j = T^{j-1}$ and T^j contains the support set T of the measured row-sparse matrix $X = X_{(T)}$, the normalized step-size

$$\omega^j = \frac{\|(A^*R^{j-1})_{(T^{j-1})}\|_F^2}{\|A_{T^{j-1}}(A^*R^{j-1})_{(T^{j-1})}\|_F^2}$$

is optimal in terms of minimizing the norm of the residual R^j. When elements of the support T of the measured matrix $X = X_{(T)}$ are missing from the current support proxy T^j, the step-size is nearly optimal in the sense that the unknown error in the step size is exclusively determined by the missing elements $T \backslash T^j$. In other words, when considering minimizing the norm of the residual

$$R^j = Y - AX_{(T^j)} = A(X_{(T)} - X_{(T^j)}),$$

the optimal step size is not computable without oracle information regarding the new support proxy T^j and the support T of the target matrix $X = X_{(T)}$.

Algorithm 3 SCoSaMP

for iteration j **until stopping criteria do**
  $S^j = \texttt{DetectSupport}(A^*R^{j-1}, 2k)$
  $Q^j = T^{j-1} \cup S^j$
  $U^j = \arg\min\{\|Y - AZ\|_F : \text{supp}(Z) \subseteq Q^j\}$
  $T^j = \texttt{DetectSupport}(U^j, k)$
  $X^j = \texttt{Threshold}(U^j, T^j)$
  $R^j = Y - AX^j$
end for
return $\hat{X} = X^{j^\star}$ when stopping at iteration $j^\star$.

SCoSaMP is also a support identification algorithm but takes a fundamentally different approach to constructing the approximation X^j. The support of the previous approximation T^{j-1} is combined with the set of $2k$ indices of the largest row-ℓ_2-norms of the residual A^*R^{j-1}. This larger set, Q^j, has at most $3k$ indices and the next approximation is determined by projecting the measurements Y onto this subspace. The best k-row-sparse approximation is then obtained by hard thresholding this projection to the rows with k largest row-ℓ_2-norms.

C. Sufficient Restricted Isometry Properties

The following recovery guarantees are based on the *restricted isometry property* (RIP) introduced by Candés and Tao [27]. The standard *RIP constant of order* k is the smallest value R_k such that

$$(1 - R_k)\|x\|_2^2 \leq \|Ax\|_2^2 \leq (1 + R_k)\|x\|_2^2$$

for all $x \in \chi_n(k)$. The RIP constants are clearly determined by the most extreme singular values of all $m \times k$ submatrices of A formed by selecting k columns. However, the smallest and largest singular values of the submatrices can deviate from 1 in a highly asymmetric fashion since the smallest singular values are nonnegative while the largest singular values can be much greater than 1. Therefore, it is beneficial to treat the sets of smallest and largest singular values independently. A natural relaxation of the standard RIP constants is to use an asymmetric version of Candés and Tao's RIP constants; the asymmetric RIP constants presented in [21] capture the most extreme smallest and largest singular values from the set of all $m \times k$ matrices formed by selecting k columns of A.

alg	ARIP Condition	$\mu^{alg}(k;A)$	$\xi^{alg}(k;A)$
SIHT	$2\phi_\omega(3k) < 1$	$2\phi_\omega(3k)$	$2\omega\sqrt{1+U_{2k}}$
SNIHT	$2U_{3k} + 2L_{3k} + L_k < 1$	$2\psi(3k)$	$2\left(\frac{\sqrt{1+U_{2k}}}{1-L_k}\right)$
SHTP	$\sqrt{3}\phi_\omega(3k) < 1$	$\sqrt{\frac{2[\phi_\omega(3k)]^2}{1-[\phi_\omega(2k)]^2}}$	$\sqrt{\frac{2(1+U_{2k})}{1-[\phi_\omega(2k)]^2}} + \frac{\sqrt{1+U_k}}{(1-L_k)(1-\phi_\omega(2k))}$
SNHTP	$\sqrt{3}U_{3k} + \sqrt{3}L_{3k} + L_k < 1$	$\sqrt{\frac{2[\psi(3k)]^2}{1-[\psi(2k)]^2}}$	$\sqrt{\frac{2(1+U_{2k})}{1-[\psi(2k)]^2}} + \frac{\sqrt{1+U_k}}{(1-L_k)(1-\psi(2k))}$
SCoSaMP	$\sqrt{\frac{5+\sqrt{73}}{2}}\max\{U_{4k}, L_{4k}\} < 1$	$\sqrt{\frac{4[R_{4k}]^2(1+3[R_{4k}]^2)}{1-[R_{4k}]^2}}$	$\sqrt{3(1+U_{3k})} + \sqrt{1+3[R_{4k}]^2}\left(\sqrt{\frac{2(1+U_{4k})}{1-[R_{4k}]^2}} + \sqrt{\frac{1+U_{3k}}{1-R_{4k}}}\right)$

TABLE I
SUFFICIENT ARIP CONDITONS WITH CONVERGENCE FACTORS $\mu^{alg}(k;A)$ AND STABILITY FACTORS $\xi^{alg}(k;A)$ FOR ALGS. 1–3. LET $\phi_\omega(ck) = \max\{|1-\omega(1+U_{ck})|, |1-\omega(1-L_{ck})|\}$ AND $\psi(ck) = \frac{U_{ck}+L_{ck}}{1-L_k}$ FOR THE ARIP CONSTANTS L_k, L_{ck}, AND U_{ck} OF THE $m \times n$ MATRIX A.

Definition 1 (RIP Constants)**.** *For $A \in M(m,n)$, the* lower and upper asymmetric restricted isometry property (ARIP) constants of order k *are denoted L_k and U_k, respectively, and are defined as:*

$$L_k := \min_{c\geq 0} c \quad \text{subject to} \quad \begin{cases} (1-c)\|x\|_2^2 \leq \|Ax\|_2^2 \\ \text{for all } x \in \chi_n(k) \end{cases} \quad (1)$$

$$U_k := \min_{c\geq 0} c \quad \text{subject to} \quad \begin{cases} (1+c)\|x\|_2^2 \geq \|Ax\|_2^2 \\ \text{for all } x \in \chi_n(k) \end{cases} \quad (2)$$

The standard (symmetric) restricted isometry property (RIP) constant of order k *is denoted R_k and can be defined in terms of the ARIP constants:*

$$R_k := \max\{L_k, U_k\}. \quad (3)$$

The main result for each of the algorithms takes on the same formulation. Therefore, we consolidate the results into a single theorem where the sufficient ARIP conditions are stated in Table I along with the appropriate convergence and stability factors. Theorem 1 provides a bound on the discrepancy of the row sparse approximation obtained by the greedy algorithms and the optimal row sparse approximation.

Theorem 1 (Simultaneous Sparse Approximation)**.** *Suppose $A \in M(m,n)$, $X \in M(n,l)$, T is the index set of rows of X with the k largest row-ℓ_2-norms, $Y = AX + E = AX_{(T)} + \tilde{E}$ for some error matrix E and $\tilde{E} = AX_{(T^c)} + E$. Assume the initial approximation is the zero matrix $X^0 = 0$. If A satisfies the sufficient ARIP conditions stated in Table I, then each algorithm, alg from {SIHT, SNIHT, SHTP, SNHTP, SCoSaMP}, is guaranteed after j iterations to return an approximation X^j satisfying*

$$\|X^j - X_{(T)}\|_F \leq \left(\mu^{alg}\right)^j \|X_{(T)}\|_F + \frac{\xi^{alg}}{1-\mu^{alg}}\|\tilde{E}\|_F; \quad (4)$$

where $\mu^{alg} \equiv \mu^{alg}(k;A)$ and $\xi^{alg} \equiv \xi^{alg}(k;A)$ are defined in Table I.

In the ideal, exact row sparse setting, a more specific result applies. Under the same sufficient ARIP conditions the greedy algorithms are all guaranteed to converge to the targeted row sparse matrix and the support set is identified in a finite number of iterations.

Corollary 1 (Simultaneous Exact Recovery)**.** *Suppose $A \in M(m,n)$, $X \in \chi_{n,l}(k)$, $Y = AX$, and the initial approximation is the zero matrix $X^0 = 0$. If A satisfies the sufficient ARIP conditions stated in Table I, then each algorithm, alg from {SIHT, SNIHT, SHTP, SNHTP, SCoSaMP}, is guaranteed after j iterations to return an approximation X^j satisfying*

$$\|X^j - X\|_F \leq \left(\mu^{alg}\right)^j \|X\|_F, \quad (5)$$

where $\mu^{alg} \equiv \mu^{alg}(k;A)$ is defined in Table I.

Moreover, define

$$j^{alg}_{max} = \left\lceil \frac{\log \nu_{min}(X)}{\log \mu^{alg}(k;A)} \right\rceil + 1 \quad (6)$$

$$\text{where} \quad \nu_{min}(X) = \frac{\min_{i\in supp(X)} \|X_{(i)}\|_2}{\|X\|_F}.$$

Then, if $j \geq j^{alg}_{max}$, $supp(X^j) \subset supp(X)$.

The ARIP analyses of the MMV variants of the greedy algorithms are clearly independent of the number of vectors (columns) contained in X, and the sufficient conditions therefore apply to the SMV case. Hence the sufficient ARIP conditions in Table I capture the known conditions for the SMV case presented by Foucart [20] namely $R_{3k} < 1/2$ for IHT, $R_{3k} < 1/\sqrt{3}$ for HTP, and $R_{4k} < \sqrt{2/(5+\sqrt{73})}$ for CoSaMP. The standard RIP extensions to the normalized versions are therefore $R_{3k} < 1/5$ for NIHT and $R_{3k} < 1/(2\sqrt{3}+1)$ for NHTP. The proofs of Theorem 1 and Corollary 1 appear in Appendix A and [26].

III. ALGORITHM COMPARISON

A. Strong Phase Transitions

The comparison of sufficient conditions based on restricted isometry properties can be challenging when the conditions do not take on the same formulation or use different support sizes for the RIP constants. Blanchard, Cartis, and Tanner [21] developed bounds on the ARIP constants for Gaussian matrices which permit a quantitative comparison of sufficient ARIP conditions via the phase transition framework. The unit square defines a phase space for the ARIP conditions under a proportional growth asymptotic, namely $(m/n, k/m) \to (\delta, \rho)$ as $m \to \infty$ for $(\delta, \rho) \in [0,1]^2$. Utilizing the bounds on the ARIP constants it is possible to identify lower bounds on

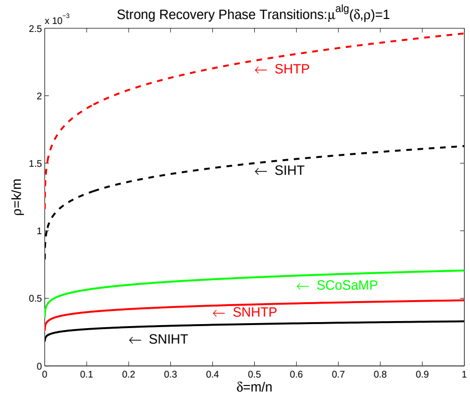


Fig. 1. Lower bounds on the strong recovery phase transition curves for SIHT, SNIHT, SHTP, SNHTP, and SCoSaMP. Beneath the line, the associated sufficient condition from Table I is satisfied with overwhelming probability on the draw of A from the Gaussian matrix ensemble; therefore $\mu^{alg}(k;A) < 1$.

strong phase transition curves $\rho_S^{alg}(\delta)$ which delineate a region $k/m = \rho < \rho_S^{alg}(\delta)$ where the sufficient ARIP condition is satisfied with overwhelming probability on the draw of A from the Gaussian ensemble, i.e. the entries of A are drawn i.i.d. from the normal distribution $\mathcal{N}(0, m^{-1})$. For a more general description of the phase transition framework in the context of compressed sensing, see [28].

For each algorithm, the strong phase transition curve $\rho_S^{alg}(\delta)$ is the solution to the equation $\mu^{alg}(\delta,\rho) \equiv 1$ where $\mu^{alg}(\delta,\rho)$ is obtained by replacing the ARIP constants in $\mu^{alg}(k;A)$ by their respective probabilistic bounds from [21]. A higher strong phase transition curve indicates a sufficient condition which is satisfied by a larger family of Gaussian matrices since the region below the curves $\rho_S^{alg}(\delta)$ demonstrate that $\mu^{alg}(k,A) < 1$ with overwhelming probability. Figure 1 shows that SHTP (with $\omega = 1$) has the best sufficient ARIP condition among these five algorithms while the efficacy of the conditions for the remaining algorithms from largest region of the phase space to smallest is SIHT, SCoSaMP, SNHTP, SNIHT.

When $l = 1$, the simultaneous recovery algorithms are identical to their SMV variants. Moreover, the sufficient conditions in Table I are independent of the number of multiple measurement vectors and therefore apply directly to the SMV algorithms. A similar analysis for SMV greedy algorithms was performed in [23]. Figure 1 shows the lower bound on the strong phase transition for the five algorithms. The improved analysis leading to the sufficient conditions in Table I yields phase transition curves for IHT and CoSaMP that capture a larger region of the phase space then the phase transition curves reported in [23]. The strong phase transition curves for the sufficient ARIP conditions for NIHT, HTP, and NHTP are reported for the first time.

The lower bounds on the strong phase transition curves point out the pessimism in the worst case analysis. Notice that Figure 1 implies the sufficient conditions from Theorem 1 require $\rho = k/m < .0008$ for SCoSaMP, SNIHT, and SHTP. The bounds on the ARIP constants are surprisingly tight and improved bounds by Bah and Tanner [29] show that the curves defined by the functions $\rho_S^{alg}(\delta)$ closely identify the regions of the phase space in which one can expect to satisfy the sufficient conditions. As shown in [23], empirical identifications of RIP constants show upper bounds on these phase transition curves are no more than twice as high as those depicted in Figure 1. The analysis required to employ the techniques outlined in [23] is contained in the supplementary material for this paper [26] along with the phase transition representation of the stability factors $\frac{\xi^{alg}}{1-\mu^{alg}}$ from Theorem 1.

B. Weak Phase Transitions

It is often more useful to understand the average case performance of the algorithms rather than the worst case guarantees provided by the sufficient conditions and delineated by the strong phase transition curves of Section III-A. In this section, we provide empirical average case performance comparisons via a weak recovery phase transition framework. Although empirical testing has its limitations, the results presented here provide insight into the expected relative performance of the greedy simultaneous sparse recovery algorithms SNIHT, SNHTP, and SCoSaMP.

The empirical testing was performed using an MMV extension of the Matlab version of the software *GAGA for Compressed Sensing* [30], [31]. The setup and procedures are similar to those outlined in [24]. A random problem instance consists of generating a random matrix $A \in M(m,n)$ and a random MMV matrix $X \in \chi_{n,l}(k)$, forming the measurements $Y = AX$, and passing to each algorithm the information (Y, A, k). To form the MMV matrix X, a row support set T with $|T| = k$ is chosen randomly and the entries of the multiple measurement vectors are selected from $\{-1, 1\}$ with probability $1/2$, thereby forming the matrix $X = X_{(T)} \in \chi_{n,l}(k)$.

For the results presented here, $n = 1024$ with tests conducted for 15 values of m where $m = \lceil \delta \cdot n \rceil$ for

$$\delta \in \{0.01, 0.02, 0.04, 0.06, 0.08, 0.1, \dots, 0.99\}$$

with 8 additional, linearly spaced values of δ from 0.1 to 0.99. For each (m,n) pair, a binary search determines an interval $[k_{min}, k_{max}]$ where the algorithm is observed to have successfully recovered 8 of 10 trials at k_{min} and 2 of 10 trials at k_{max}. The interval $[k_{min}, k_{max}]$ is then sampled with 50 independent, linearly spaced values of k from k_{min} to k_{max}, or every value of $k \in [k_{min}, k_{max}]$ if $k_{max} - k_{min} \leq 50$. Ten tests are conducted for each of the sampled values of $k \in [k_{min}, k_{max}]$.

The matrix X is determined to be successfully recovered if the output of the algorithm, $\hat{X}$, satisfies

$$\|\hat{X} - X_{(T)}\|_F \leq 0.001.$$

The empirical weak phase transitions are defined by a logistic regression of the data which determines a curve $\rho_W^{alg}(\delta)$ in the phase space identifying the location of 50% success. For a detailed explanation of the logistic regression, see [24].

For computational efficiency, the algorithms have been altered slightly in the testing regime. The projection steps in Algorithms 2 and 3 have been replaced with a subspace restricted conjugate gradient projection (see [30]). Empirically, SCoSaMP has improved performance when the index set S^j in Step 2 has k entries rather than $2k$ entries; this change was implemented in the testing.

Critically important to the empirical testing is establishing suitable stopping criteria for the greedy algorithms. Following the extensive work presented in [24], [30], the algorithms continue to iterate until one of the following stopping criteria is met:

- the residual is small: $\|R^j\|_F < 0.001 \cdot \frac{m}{n}$;
- a maximum number of iterations has been met: 5000 for Algorithm 1 and 300 for Algorithms 2 and 3;
- the algorithm is diverging: $\|R^j\|_F > 100 \cdot \|Y\|_F$;
- the residual has failed to change significantly in 16 iterations:

$$\max_{i=1,\dots,16} \left| \|R^{j-i+1}\|_F - \|R^{j-i}\|_F \right| < 10^{-6};$$

- after many iterations, the convergence rate is close to one: let c=700 for Algorithm 1 and c=125 for Algorithms 2 and 3,

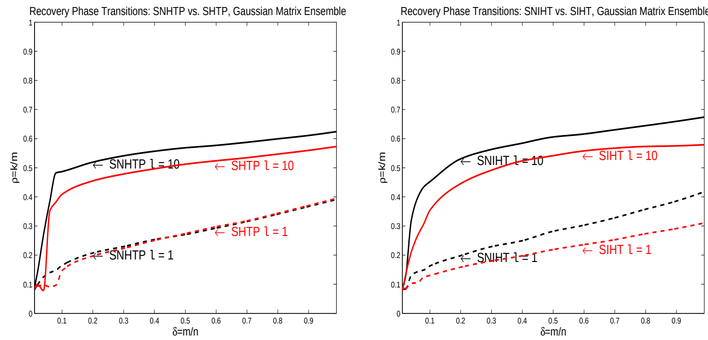
$$\text{if } j > c \text{ and } \left(\frac{\|R^{j-15}\|_F^2}{\|R^j\|_F^2} \right)^{\frac{1}{15}} > 0.999.$$

When any one of the stopping criteria is met at iteration j, the algorithm terminates and returns the k-row sparse matrix $\hat{X} = X^j$.

As in Section III-A, a higher empirical weak recovery phase transition curve indicates that the algorithm successfully recovers a larger set of MMV matrices X. All results presented in this section have the nonzero entries of the MMV matrix X selected with equal probability from $\{-1, 1\}$; alternative MMV matrix ensembles, for example selecting the nonzeros from $\mathcal{N}(0, 1)$, result in higher weak phase transition curves. These findings are consistent with other empirical studies [22], [24]. Also, throughout this section, the matrix A is selected from the Gaussian ensemble with entries drawn i.i.d. from $\mathcal{N}(0, m^{-1})$ for consistency with the strong recovery phase transition curves from Section III-A. The weak phase transition curves are higher when the matrix A is constructed by randomly selecting m rows of the discrete cosine transform; the empirical results for the DCT matrix ensemble are included in the supplementary material [26].

To demonstrate the improved performance of the algorithms from the SMV setting ($l = 1$) to the MMV setting, the weak phase transition curves are presented for $l = 1, 2, 5, 10$. In Section III-B1, the optimal step size selection in SNIHT and SNHTP is shown to provide a noticeable advantage over the fixed step size variants SIHT and SHTP, particularly as the number of multiple measurement vectors increases. The performance gain in the exact sparsity MMV setting is detailed in Section III-B2.

1) Optimal Step Size Selection: The fixed step size in SIHT and SHTP permits simplified analyses leading to weaker sufficient conditions than for the optimal step size variants SNIHT and SNHTP. This is clear from the strong phase transitions presented in Figure 1. Intuitively, selecting the step size to minimize the residual in the subspace restricted steepest descent direction should lead to improved performance. For



(a) (b)

Fig. 2. Empirical weak recovery phase transitions: fixed versus optimal step size with A from the Gaussian matrix ensemble. SHTP ($\omega = 1$) versus SNHTP (a), SIHT ($\omega = 1$) versus SNIHT (b).

the SMV setting, the introduction of the optimal step size in NIHT provides a significant improvement in average case performance [18], [24] even when compared to the tuned step size $\omega = .65$ identified for Gaussian matrices A in [22]. Interestingly, the improvement in the recovery phase transition for $l = 1$ is not nearly as dramatic for NHTP compared to HTP with $\omega = 1$.

In Figure 2, we see that for both SNIHT and SNHTP (Algorithms 1 and 2), the inclusion of the optimal step size improves performance in the MMV setting, and the advantage increases as the number of multiple measurement vectors increases. Although the analysis is simplified with a fixed step size, the improved empirical performance suggests that implementations should utilize the optimal step size, especially in the MMV setting. When A is a subsampled DCT matrix, SNIHT and SNHTP are more efficient than the fixed step size variants, especially in the most interesting regime for compressing sensing with $m/n \to 0$. The comparisons of the associated weak phase transitions for the DCT matrix ensembles are displayed in [26, Figure 6].

2) Exact Recovery: Figure 3 shows the empirical weak recovery phase transition curves $\rho_W^{alg}(\delta)$ for $X \in \chi_{n,l}(k)$ with $n = 1024$ and $l = 1, 2, 5, 10$ for SNIHT, SHTP, and SCoSaMP. A theoretical average case analysis for the greedy algorithms considered here is currently unavailable in the literature as the lack of Lipschitz continuity for the thresholding operation imposes a significant impediment. For the SMV setting, Donoho and Tanner utilized stochastic geometry to identify the weak phase transition for recovering a sparse vector via ℓ_1-minimization when A is Gaussian [32], [33]. For reference, the theoretical weak phase transition for ℓ_1-minimization with $l = 1$ is included as the blue, dashed curve in Figures 3(a)–(c).

Clearly, each of the algorithms takes advantage of additional information about the support set provided by the jointly sparse multiple measurement vectors. As l increases, the weak phase transitions increase for all three algorithms in Figure 3. For direct performance comparison, Figure 3(d) displays the empirical weak recovery phase transition curves for all three algorithms with $l = 2, 10$. SNIHT and SNHTP have very similar weak phase transition curves in the MMV setting, extending the similar observation for the SMV case detailed in [24]. When A is Gaussian, SCoSaMP recovers row

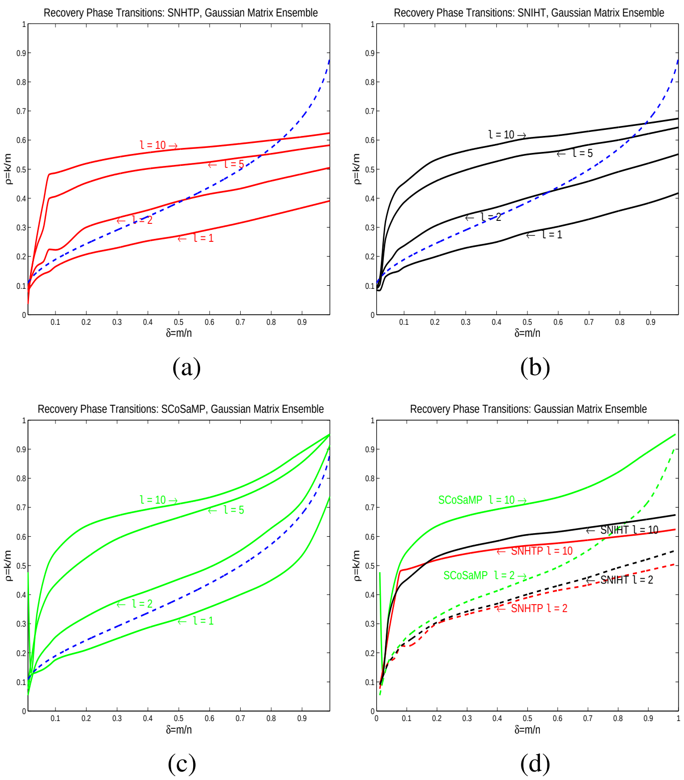


Fig. 3. Empirical weak recovery phase transitions for joint sparsity levels $l = 1, 2, 5, 10$ with A from the Gaussian matrix ensemble: SNHTP (a); SNIHT (b); SCoSaMP (c); All algorithms (d). Theoretical weak phase transition for ℓ_1-minimization is the blue, dashed curve in (a)–(c).

alg	Matrix Ensemble	$l=2$	$l=5$	$l=10$
SNIHT	Gaussian	1.42	1.89	2.09
	DCT	1.33	1.74	1.88
SNHTP	Gaussian	1.38	1.83	2.03
	DCT	1.31	1.71	1.87
SCoSaMP	Gaussian	1.40	1.92	2.10
	DCT	1.38	1.82	1.98

TABLE II
THE RATIO OF THE AREA OF THE RECOVERY REGION FOR $l = 2, 5, 10$ COMPARED TO THE AREA OF THE SINGLE MEASUREMENT VECTOR RECOVERY REGION.

sparse matrices X for noticeably larger values of $\rho = k/m$ throughout the phase space, especially as $m/n \to 1$, and for all four values $l \in \{1, 2, 5, 10\}$ ($l = 1, 5$ are omitted from the plot for clarity). However, when A is a subsampled DCT matrix, the advantage shown by SCoSaMP is removed. Similar plots for the subsampled DCT are given in [26].

Referring to the area below the empirical weak phase transition curves as the *recovery region*, Table II provides the ratio of the areas of the recovery regions for $l = 2, 5, 10$ compared to the area of the recovery region for the SMV setting ($l = 1$). When A is drawn from the Gaussian matrix ensemble, the area of the recovery regions for all three algorithms more than doubles when 10 jointly sparse vectors are simultaneously recovered. For $l = 10$, the area of the recovery region for SCoSaMP is approximately 1.3 times larger than the area of the recovery region for SNHTP; see Figure 3(d).

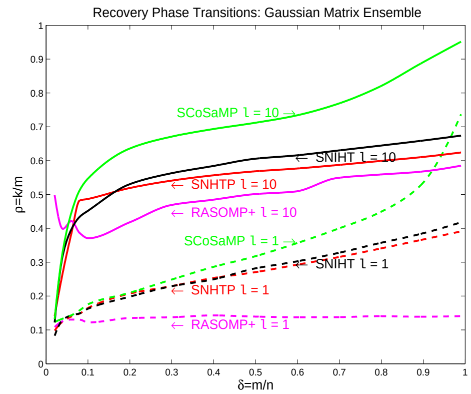


Fig. 4. Weak Recovery Phase Transitions for Algorithms 1–3 and RA-SOMP+MUSIC with joint sparsity levels $l = 1, 10$ with A from the Gaussian matrix ensemble.

C. Comparison to Rank Aware Algorithms

Most greedy simultaneous recovery algorithms, including Algorithms 1–3, fail to incorporate the rank of X in the algorithms' definition and analysis. From this point of view, the algorithms are "rank blind". In fact, the ARIP analysis presented here and elsewhere requires the number of observations (rows of A) to satisfy $m \gtrsim Ck\log(n)$ for a constant C. Davies and Eldar analyzed "rank aware" (RA) greedy algorithms for the MMV problem [9] which incorporate an orthogonalization of the column space of the residual in each iteration. Blanchard and Davies [25] and Lee, Bressler, and Junge [12] considered rank aware greedy algorithms followed by an application of MUSIC [34] for incorporating rank awareness in the MMV setting. For A Gaussian, Blanchard and Davies established that the logarithmic term in the requisite number of measurements is reduced by the rank so that $m \gtrsim Ck\left(\frac{1}{r}\log(n) + 1\right)$ [25].

Interestingly, Figure 4 shows the seemingly rank blind greedy algorithms presented here have superior weak phase transitions than the rank aware algorithm RA-SOMP+MUSIC. This empirical observation suggests that SNIHT, SNHTP, and SCoSaMP are somehow rank aware[2] and calls for further exploration. One possible explanation is that when selecting support sets based on the largest row-ℓ_2-norms of the residual or the current approximation, the `DetectSupport` step in Algorithms 1–3 is inherently rank aware providing the performance gain with the increase in the number of multiple measurement vectors.

IV. Conclusion

Five greedy algorithms designed for the SMV sparse approximation problem have been extended to the MMV problem with ARIP guarantees on the approximation errors and convergence for the ideal exact row sparse situations. The sufficient ARIP conditions for the algorithms have been compared via the strong phase transition framework for Gaussian matrices providing the best available strong recovery phase transitions curves. The importance of the optimal step size selection in the normalized variants of the algorithms was

[2] An alternative interpretation is that rank awareness in OMP based algorithms is insufficient to close the performance gap on these more sophisticated greedy algorithms.

shown through empirical testing to provide a more significant advantage in the MMV setting than in the SMV setting. Also, through empirical testing, an average case performance comparison of the algorithms was presented through the weak phase transition framework. These greedy algorithms appear to outperform an explicitly rank aware algorithm.

In this work, we have identified the location of the weak phase transition curves. Future empirical investigations on additional performance characteristics for more realistic sized problems and noisy signals, similar to [24], will better inform algorithm selection in regions below the weak phase transition curves for multiple algorithms.

Appendix A
Proofs of Recovery Guarantees

All inner products in this manuscript are Frobenius matrix inner products. For $Z, W \in M(r,c)$, the Frobenius matrix inner product is defined by

$$\langle Z, W\rangle = \text{trace}(W^T Z).$$

The Frobenius norm defined in Section I can be equivalently defined via the Frobenius matrix inner product: for $Z \in M(r,c)$,

$$\|Z\|_F^2 = \langle Z, Z\rangle .$$

A. Technical Lemmas

The straightforward proofs of Lemmas 1–2 are available in the supplementary material [26] for completeness.

Lemma 1. *Let $Z \in M(n,l)$ and let $S, T \subset \{1, 2, \ldots, n\}$ be row index sets with $|S| = |T| = k$. If T is the index set of the rows of Z with the k largest row-ℓ_2-norms, then*

$$\|Z - Z_{(T)}\|_F \leq \|Z - Z_{(S)}\|_F. \tag{7}$$

Lemma 2. *Suppose $Y = AX + E$, $T = supp(X)$, and $\tilde{E} = AX_{(T^c)} + E$. Let alg be any algorithm from Algorithms 1–3 and let X^j denote the approximation in iteration j from alg. If there exist nonnegative constants μ^{alg} and ξ^{alg} such that $\mu^{alg} < 1$ and for any iteration $j > 1$,*

$$\|X^j - X_{(T)}\|_F \leq \mu^{alg}\|X^{j-1} - X_{(T)}\|_F + \xi^{alg}\|\tilde{E}\|_F, \tag{8}$$

then

$$\|X^j - X_{(T)}\|_F \leq \left(\mu^{alg}\right)^j \|X^0 - X_{(T)}\|_F + \frac{\xi^{alg}}{1-\mu^{alg}}\|\tilde{E}\|_F. \tag{9}$$

Lemma 3. *If $A \in M(m,n)$ has ARIP constants L_k, U_k and $Z \in \chi_{n,l}(k)$, then*

$$(1 - L_k)\|Z\|_F^2 \leq \|AZ\|_F^2 \leq (1 + U_k)\|Z\|_F^2. \tag{10}$$

Proof: For each column of Z, Definition 1 states $(1 - L_k)\|Z_i\|_2^2 \leq \|AZ_i\|_2^2 \leq (1 + U_k)\|Z_i\|_2^2$. Therefore, summing over all columns in Z provides the ARIP statement in terms of the Frobenius norm. ∎

Lemma 4. *Let $A \in M(m,n)$ have ARIP constants L_k and U_k, and let T^j be the index set from the `DetectSupport` step in iteration j of SNIHT or SNHTP (Algorithms 1 or 2). Then, w^{j+1}, the optimal steepest descent step size in iteration $j+1$ of SNIHT or SNHTP, satisfies*

$$\frac{1}{1+U_k} \leq w^{j+1} \leq \frac{1}{1-L_k}. \tag{11}$$

Proof: Let $Z^j = \left(A^*(Y - AX^j)\right)_{(T^j)}$ where $Y \in M(m,l)$ is the input measurements and X^j is the approximation after iteration j for SNIHT or SNHTP. Then

$$w^{j+1} = \frac{\|Z^j\|_F^2}{\|A_{T^j}Z^j\|_F^2}.$$

By Lemma 3,

$$(1 - L_k) \leq \frac{\|A_{T^j}Z^j\|_F^2}{\|Z^j\|_F^2} \leq (1 + U_k),$$

and (11) follows. ∎

Lemma 5. *Let $A \in M(m,n)$ with ARIP constants L_k, L_{ck}, U_k, and U_{ck} where $k, ck \in \mathbb{N}$. Let S be any column index set with $|S| = ck$ and let $\{w^j\}_{j=1}^{\infty}$ be a sequence of positive scalars. Then,*

(i) if $\omega^j = \omega$ is constant for all j, then

$$\|I - \omega A_S^* A_S\|_2 \leq \max\left\{|1 - \omega(1+U_{ck})|, |1-\omega(1-L_{ck})|\right\};$$

(ii) if $\frac{1}{1+U_k} \leq \omega^j \leq \frac{1}{1-L_k}$ for all j, then

$$\|I - \omega^j A_S^* A_S\|_2 \leq \frac{U_{ck} + L_{ck}}{1 - L_k}.$$

Proof: From Definition 1 and as described in [21], the ARIP constants are equal to extreme eigenvalues on the set of all Gram matrices comprised of ck columns of A:

$$1 - L_{ck} = \min_{Q:|Q|=ck} \lambda\left(A_Q^* A_Q\right);$$
$$1 + U_{ck} = \max_{Q:|Q|=ck} \lambda\left(A_Q^* A_Q\right).$$

Hence, for any set S with $|S| = ck$, $1 - L_{ck} \leq \lambda(A_S^* A_S) \leq 1 + U_{ck}$. Therefore, the eigenvalues of the matrix $I - \omega^j A_S^* A_S$ are bounded by

$$1 - \omega^j(1 + U_{ck}) \leq \lambda\left(I - \omega^j A_S^* A_S\right) \leq 1 - \omega^j(1 - L_{ck}).$$

In case (i),

$$\begin{aligned}\|I - \omega A_S^* A_S\|_2 &= \max\left|\lambda\left(I - \omega^j A_S^* A_S\right)\right| \\ &\leq \max\left\{|1 - \omega(1+U_{ck})|, |1-\omega(1-L_{ck})|\right\}.\end{aligned}$$

In case (ii), for each j Lemma 4 ensures

$$1 - \frac{1}{1-L_k}(1 + U_{ck}) \leq 1 - \omega^j(1 + U_{ck}),$$
$$1 - \frac{1}{1+U_k}(1 - L_{ck}) \geq 1 - \omega^j(1 - L_{ck}).$$

Hence,

$$\frac{-U_{ck} - L_k}{1 - L_k} \leq \lambda\left(I - \omega^j A_S^* A_S\right) \leq \frac{U_k + L_{ck}}{1+U_k}.$$

Thus

$$\begin{aligned}\|I-\omega A_S^*A_S\|_2 &\le \max\left|\lambda\left(I-\omega^j A_S^*A_S\right)\right| \\ &\le \max\left\{\left|\frac{-U_{ck}-L_k}{1-L_k}\right|,\left|\frac{U_k+L_{ck}}{1+U_k}\right|\right\} \\ &= \max\left\{\frac{U_{ck}+L_k}{1-L_k},\frac{U_k+L_{ck}}{1+U_k}\right\}.\end{aligned}$$

Since $U_k \le U_{ck}, L_k \le L_{ck}$, and $1-L_k \le 1+U_k$, (ii) follows from the bound

$$\max\left\{\frac{U_{ck}+L_k}{1-L_k},\frac{U_k+L_{ck}}{1+U_k}\right\} \le \frac{U_{ck}+L_{ck}}{1-L_k}.$$

■

B. Algorithm Specific Theorems

The following theorem and its proof are representative of the analysis for all five greedy algorithms. This proof includes the use of ARIP constants and simultaneously treats both fixed and normalized step-sizes. The proof is based on the IHT proof of Foucart [20].

Theorem 2 (SIHT and SNIHT). *Let $X \in M(n,l)$ and let T be the row index set for the k rows of X with largest ℓ_2 norm. Let $A \in M(m,n)$ with ARIP constants L_k, L_{3k}, U_{2k}, and U_{3k}, and let $\varphi(3k)$ be a function of ARIP constants such that for all ω^j in Algorithm 1, $\|I-\omega^j A_Q^*A_Q\|_2 \le \varphi(3k) < 1$ for all index sets Q with $|Q| = 3k$. Define $Y = AX+E = AX_{(T)}+\tilde{E}$ for some error matrix $E \in M(m,l)$ and $\tilde{E} = AX_{(T^c)}+E$. Define the ARIP functions*

$$\mu_1(k) = 2\varphi(3k) \tag{12}$$

$$\xi_1(k) = 2(\max_j \omega^j)\sqrt{1+U_{2k}}. \tag{13}$$

If $\{X^j\}$ is the sequence of approximations from SIHT or SNIHT (Algorithm 1), then for all j

$$\|X^{j+1}-X_{(T)}\|_F \le \mu_1(k)\|X^j-X_{(T)}\|_F+\xi_1(k)\|\tilde{E}\|_F. \tag{14}$$

Proof: Let $V^j = X^j + w^j A^*(Y-AX^j)$ be the update step from Algorithm 1. By substituting $Y = AX_{(T)}+AX_{(T^c)}+E = AX_{(T)}+\tilde{E}$, we have

$$V^j = X^j+\omega^j A^*A(X_{(T)}-X^j)+\omega^j A^*\tilde{E}. \tag{15}$$

By the `DetectSupport` and `Threshold` steps in Algorithm 1, Lemma 1 implies

$$\|V^j-X^{j+1}\|_F^2 \le \|V^j-X_{(T)}\|_F^2. \tag{16}$$

Writing $V^j = V^j - X_{(T)} + X_{(T)}$, the left hand side of (16) can be expanded via the Frobenius inner product to reveal

$$\begin{aligned}\|V^j-X^{j+1}\|_F^2 &= \|V^j-X_{(T)}\|_F^2+\|X_{(T)}-X^{j+1}\|_F^2 \\ &\quad -2\left[\textsc{Real}\left(\left\langle V^j-X_{(T)},X_{(T)}-X^{j+1}\right\rangle_F\right)\right].\end{aligned} \tag{17}$$

Combining (16) and (17) and bounding the real part of the inner product by its magnitude,

$$\|X_{(T)}-X^{j+1}\|_F^2 \le 2\left|\left\langle V^j-X_{(T)},X_{(T)}-X^{j+1}\right\rangle\right|. \tag{18}$$

From (15),

$$V^j-X_{(T)} = (I-\omega^j A^*A)(X^j-X_{(T)})+\omega^j A^*\tilde{E},$$

so applying the triangle inequality to (18), we have

$$\begin{aligned}&\|X^{j+1}-X_{(T)}\|_F^2 \\ &\quad\le 2\left|\left\langle (I-\omega^j A^*A)(X^j-X_{(T)}),(X^{j+1}-X_{(T)})\right\rangle\right| \\ &\qquad + 2\omega^j\left|\left\langle \tilde{E},A(X^{j+1}-X_{(T)})\right\rangle\right| \\ &\quad= 2\left|\left\langle (I-\omega^j A_Q^*A_Q)(X^j-X_{(T)}),(X^{j+1}-X_{(T)})\right\rangle\right| \\ &\qquad + 2\omega^j\left|\left\langle \tilde{E},A(X^{j+1}-X_{(T)})\right\rangle\right|\end{aligned} \tag{19}$$

where $Q = T\cup T^j\cup T^{j+1}$.

Now, let $\varphi(3k)$ be a function of ARIP constants such that for any set Q with $|Q| = 3k$, we have $\|I-\omega^j A_Q^*A_Q\|_2 \le \varphi(3k)$. Then

$$\begin{aligned}&\left|\left\langle (I-\omega^j A_Q^*A_Q)(X^j-X_{(T)}),(X^{j+1}-X_{(T)})\right\rangle\right| \\ &\qquad\le \varphi(3k)\|X^j-X_{(T)}\|_F\|X^{j+1}-X_{(T)}\|_F.\end{aligned} \tag{20}$$

By Definition 1 and the Cauchy-Schwartz inequality,

$$\left|\left\langle \tilde{E},A(X^{j+1}-X_{(T)})\right\rangle\right| \le \sqrt{1+U_{2k}}\|\tilde{E}\|_F\|X^{j+1}-X_{(T)}\|_F. \tag{21}$$

With (20) and (21), (19) simplifies to

$$\|X^{j+1}-X_{(T)}\|_F \le 2\varphi(3k)\|X^j-X_{(T)}\|_F+2\omega^j\sqrt{1+U_{2k}}\|\tilde{E}\|_F, \tag{22}$$

establishing (14). ■

The proofs of the following theorems are presented in [26] for completeness. The proofs closely follow the analysis for the SMV variants presented by Foucart [20] while incorporating ARIP constants and simultaneously treating fixed and normalised step-sizes.

Theorem 3 (SHTP and SNHTP). *Let $X \in M(n,l)$ and let T be the row index set for the k rows of X with largest ℓ_2 norm. Let $A \in M(m,n)$ with ARIP constants L_{ck} and U_{ck} for $c = 1,2,3$, and let $\varphi(ck)$ be a function of ARIP constants such that for all ω^j in Algorithm 2, $\|I-\omega^j A_Q^*A_Q\|_2 \le \varphi(ck) < 1$ for all index sets Q with $|Q| = ck$. Define $Y = AX+E = AX_{(T)}+\tilde{E}$ for some error matrix $E \in M(m,l)$ and $\tilde{E} = AX_{(T^c)}+E$. Define the ARIP functions*

$$\mu_2(k) = \sqrt{\frac{2\left[\varphi(3k)\right]^2}{1-\left[\varphi(2k)\right]^2}} \tag{23}$$

$$\xi_2(k) = \sqrt{\frac{2(1+U_{2k})}{1-\left[\varphi(2k)\right]^2}}+(\max_j \omega^j)\frac{\sqrt{1+U_k}}{1-\varphi(2k)}. \tag{24}$$

If $\{X^j\}$ is the sequence of approximations from SHTP or SNHTP (Algorithm 2), then for all j

$$\|X^{j+1}-X_{(T)}\|_F \le \mu_2(k)\|X^j-X_{(T)}\|_F+\xi_2(k)\|\tilde{E}\|_F. \tag{25}$$

Theorem 4 (SCoSaMP). *Let $X \in M(n,l)$ and let T be the row index set for the k rows of X with largest ℓ_2 norm. Let $A \in M(m,n)$ with ARIP constants L_{ck} and U_{ck} for $c = 2,3,4$, and let $\varphi(ck)$ be a function of ARIP constants such that $\|I-A_Q^*A_Q\|_2 \le \varphi(ck) < 1$ for all index sets Q with $|Q| = ck$. Define $Y = AX+E = AX_{(T)}+\tilde{E}$ for some error* *matrix $E \in M(m,l)$ and $\tilde{E} = AX_{(T^c)} + E$. Define the ARIP functions*

$$\mu_3(k) = \sqrt{\frac{(\varphi(2k)+\varphi(4k))^2\,(1+3[\varphi(4k)]^2)}{1-[\varphi(4k)]^2}} \quad (26)$$

$$\xi_3(k) = \sqrt{3(1+U_{3k})} + \sqrt{1+3[\varphi(4k)]^2}\left(\sqrt{\frac{2(1+U_{4k})}{1-[\varphi(4k)]^2}} + \sqrt{\frac{1+U_{3k}}{1-\varphi(4k)}}\right). \quad (27)$$

If $\{X^j\}$ is the sequence of approximations from SCoSaMP (Algorithm 3), then for all j

$$\|X^{j+1} - X_{(T)}\|_F \leq \mu_3(k)\|X^j - X_{(T)}\|_F + \xi_3(k)\|\tilde{E}\|_F. \quad (28)$$

C. Proof of Main Results

Proof of Theorem 1: For the fixed step size variants SIHT or SHTP, Lemma 5 ensures that the ARIP function $\varphi(ck)$ from Theorems 2 and 3 can be chosen to be the function $\phi_\omega(ck) = \max\{|1-\omega(1+U_{ck})|, |1-\omega(1-L_{ck})|\}$. Likewise, for the normalised variants SNIHT and SNHTP, Lemma 5 ensures that the optimal subspace restricted steepest descent steps permit the substitution of ARIP function $\psi(ck) = \frac{U_{ck}+L_{ck}}{1-L_k}$ for the ARIP function $\varphi(ck)$ in Theorems 2 and 3. Finally, for SCoSaMP, it is clear that we can select $\varphi(4k) = \max\{U_{4k}, L_{4k}\} = R_{4k}$.

All three choices of ARIP functions are nondecreasing. In the following, each ARIP function $\mu^{alg}(k;A)$ is defined in Table I. Therefore, it is clear that with $\varphi(ck) = \phi_\omega(ck)$,

$$\mu_1(k) \leq \mu^{siht}(k;A), \quad (29)$$

$$\mu_2(k) \leq \mu^{shtp}(k;A). \quad (30)$$

For $\varphi(ck) = \psi(ck)$,

$$\mu_1(k) \leq \mu^{sniht}(k;A), \quad (31)$$

$$\mu_2(k) \leq \mu^{snhtp}(k;A). \quad (32)$$

Finally, with $\varphi(4k) = R_{4k}$,

$$\mu_3(k) \leq \mu^{scosamp}(k;A). \quad (33)$$

The sufficient ARIP conditions in Table I guarantee that the associated ARIP functions $\mu^{alg}(k;A) < 1$. Therefore, combining Lemma 2 with Theorems 2–4 proves Theorem 1. ■

Proof of Corollary 1: In the ideal setting where $T = \text{supp}(X)$ and $E = 0$, with $X^0 = 0$ (5) follows directly from Theorem 1. The number of iterations follows from a minor generalization of analogous results in [2], [23] since it is clear that

$$\|X^{j^{alg}_{max}} - X\|_F < \min_{i\in T}\|X_{(i)}\|_2$$

and therefore $\text{supp}(X^{j^{alg}_{max}}) \subset \text{supp}(X)$. ■

References

[1] B. K. Natarajan, "Sparse approximate solutions to linear systems," *SIAM J. Computing*, vol. 24, no. 2, pp. 227–234, 1995.

[2] S. Foucart, "Recovering jointly sparse vectors via hard thresholding pursuit," in *Proc. of SAMPTA*, 2011, Online.

[3] J. A. Tropp, A. C. Gilbert, and M. J. Strauss, "Algorithms for simultaneous sparse approximation. Part I: Greedy pursuit," *Signal Processing*, vol. 86, pp. 572–588, 2006.

[4] D. Leviatan and V. N. Temlyakov, "Simultaneous approximation by greedy algorithms," *Adv. Comput. Math.*, vol. 25, no. 1-3, pp. 73–90, 2006.

[5] A. Lutoborski and V. N. Temlyakov, "Vector greedy algorithms," *J. Complexity*, vol. 19, no. 4, pp. 458–473, 2003.

[6] V. N. Temlyakov, "A remark on simultaneous greedy approximation," *East J. Approx.*, vol. 10, no. 1-2, pp. 17–25, 2004.

[7] J. Chen and X. Huo, "Theoretical results on sparse representations of multiple-measurement vectors," *IEEE Trans. Sig. Proc.*, vol. 54, no. 12, pp. 4634–4643, 2006.

[8] S. Cotter, B. Rao, K. Engan, and K. Kreutz-Delgado, "Sparse solutions to linear inverse problems with multiple measurement vectors," *IEEE Trans. Sig. Proc.*, vol. 53, no. 7, pp. 2477–2488, 2005.

[9] M. E. Davies and Y. C. Eldar, "Rank awareness in joint sparse recovery," *IEEE Trans. Inform. Theory*, vol. 58, no. 2, pp. 1135–1146, 2012.

[10] Y. Eldar and M. Mishali, "Robust recovery of signals from a structured union of subspaces," *IEEE Trans. Inform. Theory*, vol. 55, no. 11, pp. 5302–5316, 2009.

[11] Y. Eldar and H. Rauhut, "Average case analysis of multichannel sparse recovery using convex relaxation," *IEEE Trans. Inform. Theory*, vol. 56, no. 1, pp. 505–519, 2010.

[12] K. Lee, Y. Bresler, and M. Junge, "Subspace methods for joint sparse recovery," *IEEE Trans. Inform. Theory*, vol. 58, no. 6, pp. 3613–3641, 2012.

[13] M.-J. Lai and Y. Liu, "The null space property for sparse recovery from multiple measurement vectors," *Appl. Comp. Harmon. Anal.*, vol. 30, no. 3, pp. 402–406, 2011.

[14] E. van den Berg and M. Friedlander, "Theoretical and empirical results for recovery from multiple measurements," *IEEE Trans. Inform. Theory*, vol. 56, no. 5, pp. 2516–2527, 2010.

[15] J. A. Tropp, "Algorithms for simultaneous sparse approximation. Part II: Convex relaxation," *Signal Processing*, vol. 86, pp. 589–602, 2006.

[16] S. Foucart, "Hard thresholding pursuit: an algorithm for compressive sensing," *SIAM J. Numerical Analysis*, vol. 49, no. 6, pp. 2543–2563, 2011.

[17] T. Blumensath and M. E. Davies, "Iterative hard thresholding for compressed sensing," *Appl. Comput. Harmon. Anal.*, vol. 27, no. 3, pp. 265–274, 2009.

[18] ——, "Normalised iterative hard thresholding; guaranteed stability and performance," *IEEE J. Selected Topics in Signal Processing*, vol. 4, no. 2, pp. 298–309, 2010.

[19] D. Needell and J. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," *Appl. Comput. Harmon. Anal.*, vol. 26, no. 3, pp. 301–321, 2009.

[20] S. Foucart, "Sparse recovery algorithms: Sufficient conditions in terms of restricted isometry constants," in *Approximation Theory XIII: San Antonio 2010*, ser. Springer Proceedings in Mathematics, M. Neamtu and L. Schumaker, Eds. Springer New York, 2012, vol. 13, pp. 65–77.

[21] J. D. Blanchard, C. Cartis, and J. Tanner, "Compressed Sensing: How sharp is the restricted isometry property?" *SIAM Review*, vol. 53, no. 1, pp. 105–125, 2011.

[22] A. Maleki and D. Donoho, "Optimally tuned iterative reconstruction algorithms for compressed sensing," *IEEE J. Selected Topics in Signal Processing*, vol. 4, no. 2, pp. 330 –341, april 2010.

[23] J. D. Blanchard, C. Cartis, J. Tanner, and A. Thompson, "Phase transitions for greedy sparse approximation algorithms," *Appl. Comput. Harmon. Anal.*, vol. 30, no. 2, pp. 188–203, 2011.

[24] J. D. Blanchard and J. Tanner, "Performance comparisons of greedy algorithms for compressed sensing," 2013, Submitted.

[25] J. Blanchard and M. Davies, "Recovery guarantees for rank aware pursuits," *IEEE Sig. Proc. Letters*, vol. 19, no. 7, pp. 427–430, 2012.

[26] J. Blanchard, M. Cermak, D. Hanle, and Y. Jing, "Greedy algorithms for joint sparse recovery: supplementary material," 2013, www.math.grinnell.edu/∼blanchaj/GAJSsupp.pdf.

[27] E. J. Candes and T. Tao, "Decoding by linear programming," *IEEE Trans. Inform. Theory*, vol. 51, no. 12, pp. 4203–4215, 2005.

[28] D. L. Donoho and J. Tanner, "Precise undersampling theorems," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 913–924, 2010.

[29] B. Bah and J. Tanner, "Improved bounds on restricted isometry constants for gaussian matrices," *SIAM Journal on Matrix Analysis*, vol. 31, no. 5, pp. 2882–2898, 2010.

[30] J. D. Blanchard and J. Tanner, "GPU accelerated greedy algorithms for compressed sensing," *Mathematical Programming Computation*, vol. 5, no. 3, pp. 267–304, 2013.

[31] ——, "GAGA: GPU Accelerated Greedy Algorithms," 2013, version 1.0.0. [Online]. Available: www.gaga4cs.org

[32] D. L. Donoho and J. Tanner, "Counting faces of randomly projected polytopes when the projection radically lowers dimension," *J. AMS*, vol. 22, no. 1, pp. 1–53, 2009.

[33] ——, "Neighborliness of randomly projected simplices in high dimensions," *Proc. Natl. Acad. Sci. USA*, vol. 102, no. 27, pp. 9452–9457 (electronic), 2005.

[34] R. O. Schmidt, "Multiple emitter location and signal parameter estimation," *Proceedings of RADC Spectral Estimation Workshop*, pp. 243–258, 1979.

Jeffrey D. Blanchard recieved the B.A. (Hons.) degree in mathematics from Benedictine College, Atchison, KS, USA, in 1998. After serving as an officer in the US Army, he recieved the A.M. and Ph.D. degrees in mathematics from Washington University in St. Louis, St. Louis, MO, USA, in 2004 and 2007, respectively, where he held a Department of Homeland Security Fellowship. From 2007 to 2009, he was a VIGRE Research Assistant Professor in the Department of Mathematics at the University of Utah, Salt Lake City, UT, USA. Since August 2009, he has been an Assistant Professor in the Department of Mathematics and Statistics at Grinnell College, Grinnell, IA, USA. From January to December 2010, he was a National Science Foundation International Research Fellow at the School of Mathematics and the School of Electronics and Engineering at the University of Edinburgh, Endiburgh, UK. He was a 2008-2009 Mathematical Association of America Project NExT Fellow and a 2013-2014 Grinnell College Harris Faculty Fellow. His current research interests include composite dilation wavelets, compressed sensing, matrix completion, scientific computing with graphics processing units, and directing undergraduate research.

Michael Cermak is a fourth-year student from Chotebor, Czech Republic, double majoring in Mathematics & Statistics and Economics at Grinnell College, Grinnell, IA, USA. In summer 2012, he conducted research in compressed sensing with the group led by Professor Blanchard. During the 2012-2013 academic year, he completed a study abroad program at the London School of Economics in London, UK and an internship the following summer with JP Morgan in London. He intends to pursue graduate work and a career in applied mathematics with a focus on industry.

David Hanle is a fourth-year student from Madison, WI, USA, double majoring in Mathematics & Statistics and Computer Science at Grinnell College, IA, USA. In summer 2012, he conducted research in compressed sensing with the group led by Professor Blanchard. In summer 2013, he was an intern with Administrative Information Management Services at the University of Wisconsin - Madison. He hopes to pursue a career in software development. Outside of the classroom, David enjoys playing sports and piano.

Yirong Jing is a fourth-year student from Taiyuan, Shanxi, China, double majoring in Mathematics & Statistics and Economics at Grinnell College, IA, USA. In summer 2012, she conducted research in compressed sensing with the group led by Professor Blanchard. In fall 2012, she was a research assistant in the Woodrow Wilson International Center for Scholars, Washington, DC. In summer 2013, she participated in a machine learning research project led by Professor Jerod Weinman at Grinnell College.

APPENDIX B
GREEDY ALGORITHMS FOR JOINT SPARSE RECOVERY: SUPPLEMENTARY MATERIAL

This document includes supplementary material for the paper *Greedy Algorithms for Joint Sparse Recovery* and the references to definitions, theorems, lemmas and equations refer to that document. The numbering in this document is a continuation of that in the main document. First, for completeness the omitted proofs are included in Section B-A. The analysis verifying the use of the asymptotic bounds on the ARIP constants to determine the strong phase transition curves in Section III-A is included here in Section B-B. Also, various level curves for convergence and stability factors are provided in Figure 5. The additional empirical weak phase transitions for A drawn from the DCT matrix ensemble appear in Section B-C.

A. Omitted Proofs

In the following, if S, T are index sets, let $T\backslash S := \{t \in T : t \notin S\}$ and define the symmetric difference of the two sets $T\Delta S := (T \cup S)\backslash(T \cap S)$. We first prove an additional technical lemma utilized in the proofs of Theorems 3 and 4.

Lemma 6. *Let $Z \in M(n,l)$ and let $S, T \subset \{1,2,\ldots,n\}$ be row index sets. Then*

$$\|Z_{T\backslash S}\|_F + \|Z_{S\backslash T}\|_F \leq \sqrt{2}\|Z_{T\Delta S}\|_F. \tag{34}$$

Proof: For any real numbers a, b, $2ab \leq a^2 + b^2$ so that

$$\begin{aligned}\left(\|Z_{T\backslash S}\|_F + \|Z_{S\backslash T}\|_F\right)^2 &= \|Z_{T\backslash S}\|_F^2 + \|Z_{S\backslash T}\|_F^2 \\ &\quad + 2\|Z_{T\backslash S}\|_F\|Z_{S\backslash T}\|_F \\ &\leq 2\left(\|Z_{T\backslash S}\|_F^2 + \|Z_{S\backslash T}\|_F^2\right) \\ &= 2\|Z_{T\Delta S}\|_F^2\end{aligned}$$

and (34) is equivalent. ■

The following four proofs were omitted from the main manuscript and are included here for completeness.

Proof of Lemma 1: By the choice of T, $\|Z_{(T)}\|_F^2 = \sum_{t\in T}\|Z_{(t)}\|_2^2 \geq \sum_{s\in S}\|Z_{(s)}\|_2^2 = \|Z_{(S)}\|_F^2$. Thus,

$$\begin{aligned}\|Z - Z_{(T)}\|_F^2 &= \|Z\|_F^2 - \|Z_{(T)}\|_F^2 \\ &\leq \|Z\|_F^2 - \|Z_{(S)}\|_F^2 = \|Z - Z_{(S)}\|_F^2\end{aligned}$$

and (7) is equivalent. ■

Proof of Lemma 2: This is a straightforward induction argument. For $X^0 = 0$, the base case is trivial. Assuming the inductive hypotheses (9) for iteration $j-1$, then (8) implies that at iteration j,

$$\begin{aligned}\|X^j - X_{(T)}\|_F \leq \mu^{alg}&\left(\left(\mu^{alg}\right)^{j-1}\|X_{(T)}\|_F + \frac{\xi^{alg}}{1-\mu^{alg}}\|\tilde{E}\|_F\right) \\ &+ \xi^{alg}\|\tilde{E}\|_F,\end{aligned}$$

which is equivalent to (9) for iteration j. ■

Proof of Theorem 3: From the projection step in Algorithm 2, $Y - AX^{j+1}$ is Frobenius-orthogonal to the subspace $\{AZ : \text{supp}(Z) \subset T^{j+1}\}$. Letting $Y = AX + E = AX_{(T)} + AX_{(T^c)} + E = AX_{(T)} + \tilde{E}$, we have $Y - AX^{j+1} = A(X_{(T)} - X^{j+1}) + \tilde{E}$. Therefore, for all vectors Z with $\text{supp}(Z) \subset T^{j+1}$,

$$\begin{aligned}0 &= \left\langle Y - AX^{j+1}, AZ\right\rangle \\ &= \left\langle A(X_{(T)} - X^{j+1}), AZ\right\rangle + \left\langle \tilde{E}, AZ\right\rangle \\ &= \left\langle X^{j+1} - X_{(T)}, -A^*AZ\right\rangle + \left\langle \tilde{E}, AZ\right\rangle.\end{aligned} \tag{35}$$

Select $Z = (X^{j+1} - X_{(T)})_{(T^{j+1})}$ so that

$$\begin{aligned}\|Z\|_F^2 &= \|(X^{j+1} - X_{(T)})_{(T^{j+1})}\|_F^2 \\ &= \left\langle X^{j+1} - X_{(T)}, (X^{j+1} - X_{(T)})_{(T^{j+1})}\right\rangle \\ &= \left\langle X^{j+1} - X_{(T)}, Z\right\rangle.\end{aligned}$$

Scaling (35) by ω^j and adding 0 to $\|Z\|_F^2$ yields

$$\|Z\|_F^2 = \left\langle X^{j+1} - X_{(T)}, (I - w^jA^*A)Z\right\rangle + \left\langle w^j\tilde{E}, AZ\right\rangle. \tag{36}$$

Now, let $\varphi(ck)$ be a function of ARIP constants such that for any set Q with $|Q| = ck$, we have $\|I - \omega^jA_Q^*A_Q\|_2 \leq \varphi(ck) < 1$. Then with $Q = T \cup T^{j+1}$, the first term in the right hand side of (36) is bounded above by

$$\begin{aligned}\left\langle X^{j+1} - X_{(T)},\right.&\left.(I - w^jA^*A)Z\right\rangle \\ &= \left\langle X^{j+1} - X_{(T)}, (I - w^jA_Q^*A_Q)Z\right\rangle \\ &\leq \varphi(2k)\|X^{j+1} - X_{(T)}\|_F\|Z\|_F.\end{aligned} \tag{37}$$

The second term of (36) is bounded above by

$$\left\langle w^j\tilde{E}, AZ\right\rangle \leq w^j\sqrt{1+U_k}\|\tilde{E}\|_F\|Z\|_F. \tag{38}$$

Applying the bounds (37) and (38) to (36),

$$\|Z\|_F \leq \varphi(2k)\|X^{j+1} - X_{(T)}\|_F + w^j\sqrt{1+U_k}\|\tilde{E}\|_F. \tag{39}$$

Let $W = (X^{j+1} - X_{(T)})_{((T^{j+1})^c)}$ so that $X^{j+1} - X_{(T)} = Z + W$. Then, by (39)

$$\begin{aligned}\|X^{j+1} &- X_{(T)}\|_F^2 - \|W\|_F^2 = \|Z\|_F^2 \\ &\leq \left(\varphi(2k)\|X^{j+1} - X_{(T)}\|_F + w^j\sqrt{1+U_k}\|\tilde{E}\|_F\right)^2 \\ &= [\varphi(2k)]^2\|X^{j+1} - X_{(T)}\|_F^2 + \left(w^j\sqrt{1+U_k}\right)^2\|\tilde{E}\|_F^2 \\ &\quad + 2\varphi(2k)\left(w^j\sqrt{1+U_k}\right)\|X^{j+1} - X_{(T)}\|_F\|\tilde{E}\|_F.\end{aligned} \tag{40}$$

Define the convex polynomial

$$\begin{aligned}p(t) = &\left(1 - [\varphi(2k)]^2\right)t^2 - \left(2\varphi(2k)w^j\sqrt{1+U_k}\|\tilde{E}\|_F\right)t \\ &- \left(\|W\|_F^2 + \left(w^j\sqrt{1+U_k}\right)^2\|\tilde{E}\|_F^2\right).\end{aligned}$$

The larger root $t^\star$ of $p(t)$ is therefore

$$\begin{aligned}t^\star = &\frac{\varphi(2k)}{1 - [\varphi(2k)]^2}w^j\sqrt{1+U_k}\|\tilde{E}\|_F \\ &+ \frac{\sqrt{\left(1 - [\varphi(2k)]^2\right)\|W\|_F^2 + \left(w^j\sqrt{1+U_k}\right)^2\|\tilde{E}\|_F^2}}{1 - [\varphi(2k)]^2}.\end{aligned}$$

By the sub-additivity of the square root,

$$t^\star \le \frac{1+\varphi(2k)}{1-[\varphi(2k)]^2} w^j \sqrt{1+U_k}\|\tilde{E}\|_F + \frac{1}{\sqrt{1-[\varphi(2k)]^2}}\|W\|_F. \quad (41)$$

By (40), $p(\|X^{j+1} - X_{(T)}\|_F) \le 0$ and therefore $\|X^{j+1} - X_{(T)}\|_F \le t^\star$. (41) implies

$$\|X^{j+1} - X_{(T)}\|_F \le \frac{\|W\|_F}{\sqrt{1-[\varphi(2k)]^2}} + \frac{w^j\sqrt{1+U_k}}{1-\varphi(2k)}\|\tilde{E}\|_F. \quad (42)$$

To complete the proof, we find an upper bound for $\|W\|_F$. Let $V^j = X^j + \omega^j A^*(Y - AX^j)$ be the update step for Algorithm 2. The `DetectSupport` step selects T^{j+1} so that $\|V^j_{(T)}\|_F \le \|V^j_{(T^{j+1})}\|_F$ and therefore

$$\|V^j_{(T\setminus T^{j+1})}\|_F \le \|V^j_{(T^{j+1}\setminus T)}\|_F. \quad (43)$$

Substituting $Y = AX_{(T)} + \tilde{E}$,

$$\begin{aligned} V^j &= X^j + \omega^j A^*A(X_{(T)} - X^j) + \omega^j A^*\tilde{E} \\ &= X_{(T)} + (I - \omega^j A^*A)(X^j - X_{(T)}) + \omega^j A^*\tilde{E}. \quad (44)\end{aligned}$$

With $\mathrm{supp}(X^{j+1}) = T^{j+1}$, $(X_{(T)})_{(T\setminus T^{j+1})} = (X_{(T)} - X^{j+1})_{(T\setminus T^{j+1})}$ and since $W = (X^{j+1} - X_{(T)})_{((T^{j+1})^c)} = (X^{j+1} - X_{(T)})_{(T\setminus T^{j+1})}$, $V^j_{(T\setminus T^{j+1})}$ can be written

$$\begin{aligned} V^j_{(T\setminus T^{j+1})} =& - W + \omega^j (A^*\tilde{E})_{(T\setminus T^{j+1})} \\ &+ (I - \omega^j A^*A)(X^j - X_{(T)})_{(T\setminus T^{j+1})}. \quad (45)\end{aligned}$$

Therefore, the left hand side of (43) can be bounded below by applying the reverse triangle inequality to (45);

$$\begin{aligned} \|V^j_{(T\setminus T^{j+1})}\|_F \ge& \|W\|_F - \omega^j \left\|(A^*\tilde{E})_{(T\setminus T^{j+1})}\right\|_F \\ &- \left\|(I - \omega^j A^*A)(X^j - X_{(T)})_{(T\setminus T^{j+1})}\right\|_F. \quad (46)\end{aligned}$$

Since $(X_{(T)})_{(T^{j+1}\setminus T)} = 0$, (44) permits the straightforward upper bound on the right hand side of (43),

$$\begin{aligned} \|V^j_{(T^{j+1}\setminus T)}\|_F \le& \left\|(I - \omega^j A^*A)(X^j - X_{(T)})_{(T^{j+1}\setminus T)}\right\|_F \\ &+ \omega^j \left\|(A^*\tilde{E})_{(T^{j+1}\setminus T)}\right\|_F. \quad (47)\end{aligned}$$

Applying (46), (47) and Lemma 6 to (43) establishes

$$\begin{aligned} \|W\|_F \le& \sqrt{2}\left\|(I - \omega^j A^*A)(X^j - X_{(T)})_{(T\Delta T^{j+1})}\right\|_F \\ &+ \sqrt{2}\omega^j \left\|(A^*\tilde{E})_{(T\Delta T^{j+1})}\right\|_F. \quad (48)\end{aligned}$$

With $Q = T \cup T^j \cup T^{j+1}$, the first norm on the right hand side of (48) satisfies

$$\begin{aligned} &\left\|(I - \omega^j A^*A)\ (X^j - X_{(T)})_{(T\Delta T^{j+1})}\right\|_F \\ &\quad\le \left\|(I - \omega^j A^*_Q A_Q)(X^j - X_{(T)})\right\|_F \\ &\quad\le \varphi(3k)\left\|(X^j - X_{(T)})\right\|_F, \quad (49)\end{aligned}$$

while the second norm of (48) satisfies

$$\left\|(A^*\tilde{E})_{(T\Delta T^{j+1})}\right\|_F \le \sqrt{1+U_{2k}}\|\tilde{E}\|_F. \quad (50)$$

Hence, (49) and (50) yield

$$\|W\|_F \le \sqrt{2}\varphi(3k)\left\|(X^j - X_{(T)})\right\|_F + \sqrt{2(1+U_{2k})}\|\tilde{E}\|_F. \quad (51)$$

Therefore, combining (42) and (51) establishes (25). ■

Proof of Theorem 4: From the projection step in Algorithm 3, $Y - AU^j$ is Frobenius-orthogonal to the subspace $\{AZ : \mathrm{supp}(Z) \subset Q^j = S^j \cup T^j\}$. By an argument almost identical to that at the beginning of the proof of Theorem 3, we establish the upper bound

$$\|(U^j - X_{(T)})_{(Q^j)}\|_F \le \varphi(4k)\|U^j - X_{(T)}\|_F + \sqrt{1+U_{3k}}\|\tilde{E}\|_F \quad (52)$$

where $\varphi(4k)$ is any function of ARIP constants such that $\|I - A^*_Q A_Q\|_2 \le \varphi(4k) < 1$ for any index set Q with $|Q| = 4k$. In this case $Q = Q^j \cup T$ ensures $|Q| \le 4k$.

Let $W = (U^j - X_{(T)})_{((Q^j)^c)}$ so that $U^j - X_{(T)} = W + (U^j - X_{(T)})_{(Q^j)}$. Then (52) implies

$$\begin{aligned} &\|U^j - X_{(T)}\|_F^2 \le \|W\|_F^2 \\ &\quad + \left(\varphi(4k)\|U^j - X_{(T)}\|_F + \sqrt{1+U_{3k}}\|\tilde{E}\|_F\right)^2. \quad (53)\end{aligned}$$

Define the convex polynomial

$$\begin{aligned} p(t) =& \left(1 - [\varphi(4k)]^2\right)t^2 - \left(2\varphi(4k)\sqrt{1+U_{3k}}\|\tilde{E}\|_F\right)t \\ &- \left(\|W\|_F^2 + (1+U_{3k})\|\tilde{E}\|_F^2\right). \end{aligned}$$

Again, as in the proof of Theorem 3, since (53) ensures $p(\|U^j - X_{(T)}\|_F) \le 0$, bounding the larger root of $p(t)$ via the sub-additivity of the square root produces

$$\|U^j - X_{(T)}\|_F \le \frac{\|W\|_F}{\sqrt{1-[\varphi(4k)]^2}} + \frac{\sqrt{1+U_{3k}}}{1-\varphi(4k)}\|\tilde{E}\|_F. \quad (54)$$

Since $X^{j+1} - X_{(T)} = (U^j - X_{(T)}) - (U^j - X^{j+1})$, expanding the norm and bounding the real part of the inner product with its magnitude as in the proof of Theorem 2, we have

$$\begin{aligned} \|X^{j+1} - X_{(T)}\|_F^2 \le& \left\|U^j - X_{(T)}\right\|_F^2 + \left\|U^j - X^{j+1}\right\|_F^2 \\ &+ 2\left|\left\langle U^j - X_{(T)}, U^j - X^{j+1}\right\rangle\right|. \quad (55)\end{aligned}$$

Applying the triangle and Cauchy-Schwartz inequalities followed by an ARIP bound, we have

$$\begin{aligned} &\left|\left\langle U^j - X_{(T)},\ U^j - X^{j+1}\right\rangle\right| \\ &\quad\le \varphi(4k)\left\|U^j - X_{(T)}\right\|_F \left\|U^j - X^{j+1}\right\|_F \\ &\quad+ \sqrt{1+U_{3k}}\|\tilde{E}\|_F \left\|U^j - X^{j+1}\right\|_F. \quad (56)\end{aligned}$$

Note that $\mathrm{supp}(U^j - X^{j+1}) = Q^j$ and by the `DetectSupport` and `Threshold` steps in Algorithm 3, Lemma 1 ensures

$$\|U^j - X^{j+1}\|_F \le \|(U^j - X_{(T)})_{(Q^j)}\|_F. \quad (57)$$

Therefore, applying (52), (56), and (57) to (55), and rearranging yields

$$\begin{aligned} \|X^{j+1} - X_{(T)}\|_F^2 \le& (1 + 3[\varphi(4k)]^2)\|U^j - X_{(T)}\|_F^2 \\ &+ 6\varphi(4k)\sqrt{1+U_{3k}}\|U^j - X_{(T)}\|_F\|\tilde{E}\|_F \\ &+ 3(1+U_{3k})\|\tilde{E}\|_F^2. \quad (58)\end{aligned}$$

Since $36[\varphi(4k)]^2 \leq 12 + 36[\varphi(4k)]^2$, then $6\varphi(4k) \leq 2\sqrt{3(1+3[\varphi(4k)]^2)}$. Using this observation to bound (58) and simplifying produces the bound

$$\|X^{j+1} - X_{(T)}\|_F^2 \leq \sqrt{1+3[\varphi(4k)]^2}\|U^j - X_{(T)}\|_F + \sqrt{3(1+U_{3k})}\|\tilde{E}\|_F. \tag{59}$$

To complete the proof via (54), we establish an upper bound on $\|W\|_F$. Notice that $\text{supp}(X^j), \text{supp}(U^j) \subset Q^j$, and therefore $W = (U^j - X_{(T)})_{((Q^j)^c)} = (X^j - X_{(T)})_{((Q^j)^c)}$. Also, since $Q^j = S^j \cup T^j$, then $(Q^j)^c \subset (S^j)^c$ and thus

$$\|W\|_F \leq \|(X^j - X_{(T)})_{((S^j)^c)}\|_F = \|(X^j - X_{(T)})_{((T\cup T^j)\setminus S^j)}\|_F. \tag{60}$$

By the definition of S^j from Algorithm 3, Lemma 1 implies

$$\left\|\left(A^*(Y - AX^j)\right)_{((T\cup T^j)\setminus S^j)}\right\|_F \leq \left\|\left(A^*(Y - AX^j)\right)_{(S^j\setminus(T\cup T^j))}\right\|_F. \tag{61}$$

Writing $Y = AX_{(T)} + \tilde{E}$ and observing that $(X^j - X_{(T)})_{(S^j\setminus(T\cup T^j))} = 0$, the argument of the norm on the right side of (61) can be written

$$\begin{aligned}&\left(A^*(Y - AX^j)\right)_{(S^j\setminus(T\cup T^j))}\\ &= \left(A^*A(X_{(T)} - X^j)\right)_{(S^j\setminus(T\cup T^j))} + \left(A^*\tilde{E}\right)_{(S^j\setminus(T\cup T^j))}\\ &= \left((I - A^*A)(X^j - X_{(T)})\right)_{(S^j\setminus(T\cup T^j))} + \left(A^*\tilde{E}\right)_{(S^j\setminus(T\cup T^j))}.\end{aligned} \tag{62}$$

Letting $Q = T \cup Q^j = T \cup T^j \cup S^j$,

$$\begin{aligned}&\left\|\left(A^*(Y - AX^j)\right)_{(S^j\setminus(T\cup T^j))}\right\|_F\\ &\leq \left\|\left((I - A_Q^*A_Q)(X^j - X_{(T)})\right)\right\|_F + \left\|\left(A^*\tilde{E}\right)_{(S^j\setminus(T\cup T^j))}\right\|_F\\ &\leq \varphi(4k)\left\|X^j - X_{(T)}\right\|_F + \left\|\left(A^*\tilde{E}\right)_{(S^j\setminus(T\cup T^j))}\right\|_F.\end{aligned} \tag{63}$$

Similarly,

$$\begin{aligned}&\left(A^*(Y - AX^j)\right)_{((T\cup T^j)\setminus S^j)}\\ &= \left(A^*A(X_{(T)} - X^j) + A^*\tilde{E}\right)_{((T\cup T^j)\setminus S^j)}\\ &= (X^j - X_{(T)})_{((T\cup T^j)\setminus S^j)}\\ &\quad - \left((I - A^*A)(X^j - X_{(T)})\right)_{((T\cup T^j)\setminus S^j)}\\ &\quad + \left(A^*\tilde{E}\right)_{(S^j\setminus(T\cup T^j))}\end{aligned} \tag{64}$$

Therefore, (60) and (64) provide a lower bound for the left hand side of (61).

$$\begin{aligned}&\left\|\left(A^*(Y - AX^j)\right)_{((T\cup T^j)\setminus S^j)}\right\|_F\\ &\geq \|W\|_F - \left\|(I - A^*_{(T\cup T^j)}A_{(T\cup T^j)})(X^j - X_{(T)})\right\|_F\\ &\quad - \left\|\left(A^*\tilde{E}\right)_{(S^j\setminus(T\cup T^j))}\right\|_F\\ &\geq \|W\|_F - \varphi(2k)\left\|X^j - X_{(T)}\right\|_F - \left\|\left(A^*\tilde{E}\right)_{(S^j\setminus(T\cup T^j))}\right\|_F.\end{aligned} \tag{65}$$

Applying (63) and (65) to (61), solving for $\|W\|_F$, and applying Lemma 6 and the upper ARIP bound, we have

$$\|W\|_F \leq (\varphi(2k)+\varphi(4k))\left\|X^j - X_{(T)}\right\|_F + \sqrt{2(1+U_{4k})}\|\tilde{E}\|_F. \tag{66}$$

Combining (54), (59), and (66) establishes (28). ■

B. Strong Phase Transitions

Under the proportional growth asymptotic $(m/n, k/m) \to (\delta, \rho)$, computable bounds, $\mathcal{L}(\delta,\rho), \mathcal{U}(\delta,\rho)$, on the ARIP constants, L_k, U_k, were established for matrices drawn from the Gaussian ensemble [21]. The exact formulation of the bounds is available in [21].

Definition 2 (Proportional-Growth Asymptotic)**.** *A sequence of problem sizes (k,m,n) is said to* grow proportionally *if, for $(\delta,\rho) \in [0,1]^2$, $\frac{m}{n} \to \delta$ and $\frac{k}{m} \to \rho$ as $m \to \infty$.*

The following is an adaptation of [21, Thm. 1].

Theorem 5 (Blanchard, Cartis, Tanner [21])**.** *Fix $\epsilon > 0$. Under the proportional-growth asymptotic, Definition 2, sample each matrix $A \in M(m,n)$ from the Gaussian ensemble. Let $\mathcal{L}(\delta,\rho)$ and $\mathcal{U}(\delta,\rho)$ be defined as in [21, Thm. 1]. Define $\mathcal{R}(\delta,\rho) = \max\{\mathcal{L}(\delta,\rho), \mathcal{U}(\delta,\rho)\}$. Then for any $\epsilon > 0$, as $m \to \infty$,*

$$\textit{Prob}\left[L_k < \mathcal{L}(\delta,\rho) + \epsilon\right] \to 1, \tag{67}$$

$$\textit{Prob}\left[U_k < \mathcal{U}(\delta,\rho) + \epsilon\right] \to 1, \tag{68}$$

$$\textit{and} \quad \textit{Prob}\left[R_k < \mathcal{R}(\delta,\rho) + \epsilon\right] \to 1. \tag{69}$$

To employ the bounds on the ARIP constants in order to define the strong phase transition curves $\rho_S^{alg}(\delta)$, the stability factor $\mu^{alg}(k;A)$ and stability factor $\xi_S^{alg}(\delta)$ must satisfy the sufficient conditions of the following lemma:

Lemma 7 (Lemma 12, [23])**.** *For some $\tau < 1$, define the set $\Omega := (0,\tau)^p \times (o,\infty)^q$ and let $F : \Omega \to \mathbb{R}$ be continuously differentiable on Ω. Let A be a Gaussian matrix of size $m \times n$ with ARIP constants $L_k, \dots, L_{pk}, U_k, \dots, U_{qk}$. Let $\mathcal{L}(\delta,\rho), \mathcal{U}(\delta,\rho)$ be the ARIP bounds defined in Theorem 5. Define $\mathbf{1}$ to be the vector of all ones, and*

$$z(k) := \left[L_k, \dots, L_{pk}, U_k, \dots, U_{qk}\right], \tag{70}$$

$$z(\delta,\rho) := \left[\mathcal{L}(\delta,\rho), \dots, \mathcal{L}(\delta,p\rho), \mathcal{U}(\delta,\rho), \dots, \mathcal{U}(\delta,q\rho)\right]. \tag{71}$$

(i) Suppose, for all $t \in \Omega$, $(\nabla F[t])_i \geq 0$ for all $i = 1,\dots,p+q$ and for any $v \in \Omega$ we have $\nabla F[t]\cdot v > 0$. Then for any $c\epsilon > 0$, as $(k,m,n) \to \infty$ with $\frac{m}{n} \to \delta, \frac{k}{n} \to \rho$, there is overwhelming probability on the draw of the matrix A that

$$\textit{Prob}\left(F[z(k)] < F[z(\delta,\rho) + \mathbf{1}c\epsilon]\right) \to 1 \qquad \textit{as } m \to \infty. \tag{72}$$

(ii) Suppose, for all $t \in \Omega$, $(\nabla F[t])_i \geq 0$ for all $i = 1,\dots,p+q$ and there exists $j \in \{1,\dots,p\}$ such that $(\nabla F[t])_j > 0$. Then there exists $c \in (0,1)$ depending only on F, δ,and ρ such that for any $\epsilon \in (0,1)$

$$F[z(\delta,\rho) + \mathbf{1}c\epsilon] < F[z(\delta,(1+\epsilon)\rho)], \tag{73}$$

and so there is overwhelming probability on the draw of A that

$$\text{Prob}\,(F[z(k)] < F[z(\delta,(1+\epsilon)\rho)]) \to 1 \qquad \text{as } m \to \infty. \tag{74}$$

Also, $F(z(\delta,\rho))$ is strictly increasing in ρ.

Definition 3. *For $(\delta,\rho) \in (0,1)^2$, define the asymptotic bounds on the convergence factors as follows:*

$$\mu^{siht}(\delta,\rho) := 2\mathcal{R}(\delta,3\rho); \tag{75}$$

$$\mu^{sniht}(\delta,\rho) := 2\frac{\mathcal{U}(\delta,3\rho)+\mathcal{L}(\delta,3\rho)}{1-\mathcal{L}(\delta,\rho)}; \tag{76}$$

$$\mu^{shtp}(\delta,\rho) := \sqrt{\frac{2[\mathcal{R}(\delta,3\rho)]^2}{1-[\mathcal{R}(\delta,2\rho)]^2}}; \tag{77}$$

$$\mu^{snhtp}(\delta,\rho) := \sqrt{\frac{2\left(\frac{\mathcal{U}(\delta,3\rho)+\mathcal{L}(\delta,3\rho)}{1-\mathcal{L}(\delta,\rho)}\right)^2}{1-\left(\frac{\mathcal{U}(\delta,2\rho)+\mathcal{L}(\delta,2\rho)}{1-\mathcal{L}(\delta,\rho)}\right)^2}}; \tag{78}$$

$$\mu^{scosamp}(\delta,\rho) := \sqrt{\frac{4[\mathcal{R}(\delta,4\rho)]^2(1+3[\mathcal{R}(\delta,4\rho)]^2)}{1-[\mathcal{R}(\delta,4\rho)]^2}}. \tag{79}$$

For SIHT with a fixed step size of $\omega^\star = 2/(2+\mathcal{U}(\delta,3\rho)-\mathcal{L}(\delta,3\rho))$, the validity of employing the asymptotic ARIP bounds was established in [23, A.4.]. For SIHT and SHTP with a fixed step size of $\omega^\star = 1$, we see that Lemma 5 establishes that $\phi_1(ck) = \max\{U_{ck}, L_{ck}\} = R_{ck}$ is a valid selection, and thus from Table I we have

$$\mu^{shtp}(k;A) = \sqrt{\frac{2[R_{3k}]^2}{1-[R_{2k}]^2}}.$$

This allows us to state the following theorem.

Theorem 6. *Suppose $A \in M(m,n)$ is drawn from the Gaussian ensemble and that A has RIP constants $R_{2k}, R_{3k} < 1$. Consider SHTP with fixed step size $\omega^\star = 1$. Then for any $\epsilon > 0$, there is overwhelming probability on the draw of A that*

$$\mu^{shtp}(k;A) < \mu^{shtp}(\delta,(1+\epsilon)\rho). \tag{80}$$

Proof: Fix $\tau < 1$ and let $\Omega = (0,\tau)^2$. For $t \in \Omega$ define

$$F[t] = \frac{2t_2^2}{1-t_1^2}.$$

Clearly, F satisfies the conditions of Lemma 7 since

$$\nabla F[t] = \left(\frac{4t_2^2 t_1}{(1-t_1^2)^2}, \frac{4t_2}{1-t_1^2}\right) > 0.$$

Now let

$$z(k) = [R_{2k}, R_{3k}]$$
$$z(\delta,\rho) = [\mathcal{R}(\delta,2\rho), \mathcal{R}(\delta,3\rho)].$$

Then with overwhelming probability on the draw of A,

$$F[z(k)] < F[z(\delta,(1+\epsilon)\rho)].$$

Finally, we see that with overwhelming probability on the draw of A,

$$\begin{aligned}\mu^{shtp}(k;A) &= \sqrt{F[z(k)]} \\ &< \sqrt{F[z(\delta,(1+\epsilon)\rho)]} \\ &= \mu^{shtp}(\delta,(1+\epsilon)\rho).\end{aligned}$$

■

The arguments establishing the validity of the bounds $\mu^{siht}(\delta,\rho)$ and $\mu^{scosamp}(\delta,\rho)$ are similar to the argument for Theorem 6 and are therefore omitted. We now establish the validity of the asymptotic bounds for the normalized algorithms, SNIHT and SNHTP. To do so, recall the ARIP function $\psi(ck)$ from Table I:

$$\psi(ck) := \frac{U_{ck}+L_{ck}}{1-L_k}. \tag{81}$$

Therefore we introduce the following functions defined on the set $\Omega = (0,\tau)^3 \times (0,\infty)^2$ for any $\tau < 1$:

$$F_2^\psi[t] = \frac{t_4+t_2}{1-t_1}; \tag{82}$$

$$F_3^\psi[t] = \frac{t_5+t_3}{1-t_1}. \tag{83}$$

These functions have nonnegative gradients since

$$\nabla F_2^\psi[t] = \left(\frac{t_4+t_2}{(1-t_1)^2}, \frac{t_4+1}{1-t_1}, 0, \frac{1+t_2}{1-t_1}, 0\right); \tag{84}$$

$$\nabla F_3^\psi[t] = \left(\frac{t_5+t_3}{(1-t_1)^2}, 0, \frac{t_5+1}{1-t_1}, 0, \frac{1+t_3}{1-t_1}\right). \tag{85}$$

For the proofs of both of the following theorems, define

$$z(k) := [L_k, L_{2k}, L_{3k}, U_{2k}, U_{3k}], \tag{86}$$

$$z(\delta,\rho) := [\mathcal{L}(\delta,\rho), \mathcal{L}(\delta,2\rho), \mathcal{L}(\delta,3\rho), \mathcal{U}(\delta,2\rho), \mathcal{U}(\delta,3\rho)]. \tag{87}$$

Theorem 7. *Suppose $A \in M(m,n)$ is drawn from the Gaussian ensemble and that A has ARIP constants L_k, L_{3k}, U_{3k}. Then for any $\epsilon > 0$, there is overwhelming probability on the draw of A that*

$$\mu^{sniht}(k;A) < \mu^{sniht}(\delta,(1+\epsilon)\rho). \tag{88}$$

Proof: Fix $\tau < 1$ and let $\Omega = (0,\tau)^3 \times (0,\infty)^2$. From (81), (83), and (86), we see that

$$\psi(3k) = F_3^\psi[z(k)],$$

and from (76), (83), and (87), we have

$$F_3^\psi[z(\delta,\rho)] = \frac{1}{2}\mu^{sniht}(\delta,\rho).$$

(85) establishes that F_3^ψ satisfies the conditions to invoke Lemma 7. Thus, with overwhelming probability on the draw of A,

$$\begin{aligned}\mu^{sniht}(k;A) &= 2\psi(3k) = 2F_3^\psi[z(k)] \\ &< 2F_3^\psi[z(\delta,(1+\epsilon)\rho)] = \mu^{sniht}(\delta,(1+\epsilon)\rho).\end{aligned}$$

■

Theorem 8. *Suppose $A \in M(m,n)$ is drawn from the Gaussian ensemble and that A has ARIP constants*

$L_k, L_{2k}, L_{3k}, U_{2k}, U_{3k}$ with $U_{2k} + L_{2k} + L_k < 1$. Then for any $\epsilon > 0$, there is overwhelming probability on the draw of A that

$$\mu^{snhtp}(k; A) < \mu^{snhtp}(\delta, (1+\epsilon)\rho). \tag{89}$$

Proof: Fix $\tau < 1$ and let $\Omega = (0, \tau)^3 \times (0, \infty)^2$. Restrict the domain to the set $\tilde{\Omega} = \{t \in \Omega : F_2^{\psi}[t] < 1\}$. Now define

$$F^H[t] = \frac{2(F_3^{\psi}[t])^2}{1 - (F_2^{\psi}[t])^2}.$$

For $i = 2, 4$,

$$\frac{\partial}{\partial t_i} F^H[t] = \frac{4F_2^{\psi}[t](F_3^{\psi}[t])^2 \left(\frac{\partial}{\partial t_i} F_2^{\psi}[t]\right)}{\left(1 - (F_2^{\psi}[t])^2\right)^2} > 0.$$

For $i = 3, 5$,

$$\frac{\partial}{\partial t_i} F^H[t] = \frac{4F_3^{\psi}[t] \left(\frac{\partial}{\partial t_i} F_3^{\psi}[t]\right)}{1 - (F_2^{\psi}[t])^2} > 0.$$

Finally,

$$\frac{\partial}{\partial t_1} F^H[t] = \frac{4F_3^{\psi}[t] \left(\frac{\partial}{\partial t_1} F_3^{\psi}[t]\right) + 4F_3^{\psi}[t] F_2^{\psi}[t]\Psi[t]}{1 - (F_2^{\psi}[t])^2},$$

where

$$\begin{aligned}\Psi[t] &= F_3^{\psi}[t]\left(\frac{\partial}{\partial t_1} F_2^{\psi}[t]\right) - F_2^{\psi}[t]\left(\frac{\partial}{\partial t_1} F_3^{\psi}[t]\right) \\ &= \frac{t_5 + t_3}{1 - t_1}\left(\frac{t_4 + t_2}{(1 - t_1)^2}\right) - \frac{t_4 + t_2}{1 - t_1}\left(\frac{t_5 + t_3}{(1 - t_1)^2}\right) \\ &= 0.\end{aligned}$$

Hence $\nabla F^H[t] > 0$ and thus $F^H[t]$ satisfies the conditions to invoke Lemma 7. Thus, with overwhelming probability on the draw of A,

$$\begin{aligned}\mu^{snhtp}(k; A) &= \sqrt{F^H[z(k)]} \\ &< \sqrt{F^H[z(\delta, (1+\epsilon)\rho)]} = \mu^{snhtp}(\delta, (1+\epsilon)\rho).\end{aligned}$$

■

The preceding discussion establishes the validity of employing the bounds in Definition 3. Therefore for each algorithm, we establish a probabilistic lower bound on the region of the phase space in which a Gaussian matrix will satisfy the sufficient condition $\mu^{alg} < 1$. Following the work in [23], defining $\rho_S^{alg}(\delta)$ as the solution to the equation $\mu^{alg}(\delta, \rho) = 1$, if $\rho < (1-\epsilon)\rho_S^{alg}(\delta)$ for any $\epsilon > 0$, $\mu^{alg}(\delta, \rho) < 1$. Since $\mu^{alg}(k; A) < \mu^{alg}(\delta, \rho) < 1$ with overwhelming probability on the draw of A from the Gaussian ensemble, then with the same probability the sufficient ARIP condition is satisfied. The curves $\rho_S^{alg}(\delta)$ are displayed in Figure 1. Here we include level sets for both the convergence factors $\mu^{alg}(\delta, \rho)$ and the stability factor $\frac{\xi^{alg}}{1-\mu^{alg}}$ in Figure 5. The computations required to demonstrate the validity of employing the bounds on the stability functions ξ^{alg} have been omitted.

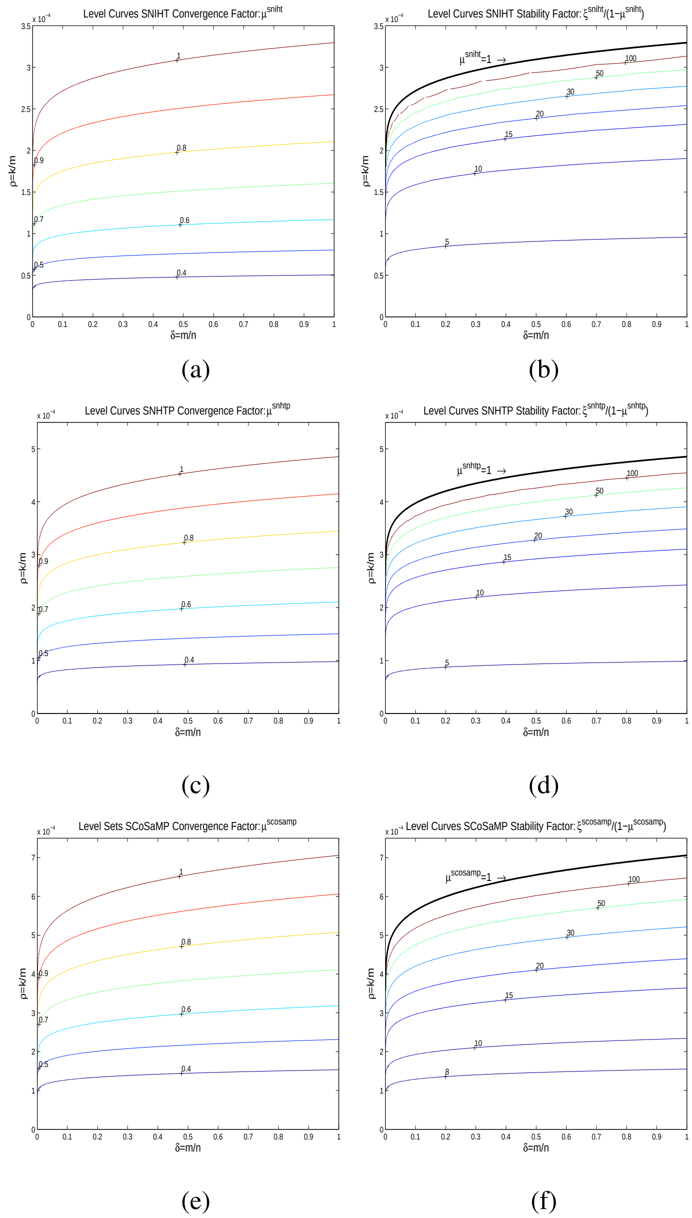


Fig. 5. Level sets for the convergence factors $\mu^{alg}(\delta, \rho)$ and the stability factors $\frac{\xi^{alg}}{1-\mu^{alg}}(\delta, \rho)$, in the left and right panels respectively: SNIHT (a),(b); SNHTP (c),(d); SCoSaMP (e),(f).

C. Weak Phase Transitions

1) Optimal Step Size Selection: The increasing performance improvement of the normalized versions of Algorithms 1 and 2 as the number of jointly sparse vectors increases was discussed for Gaussian matrices A in Section III-B1. The improvement is more pronounced when A is constructed by randomly selecting m rows of an $n \times n$ discrete cosine transform matrix (DCT). In this case, we say A is drawn from the DCT ensemble. Figure 6 includes the performance comparison of the fixed step size variants of the algorithms versus the optimal step size (normalized) variants. For comparison, both the DCT ensemble and the Gaussian ensemble are included. For SIHT, the step size is fixed at $\omega = .65$ while the step size is fixed at $\omega = 1$ for SHTP. In the SMV setting SHTP ($\omega = 1$)

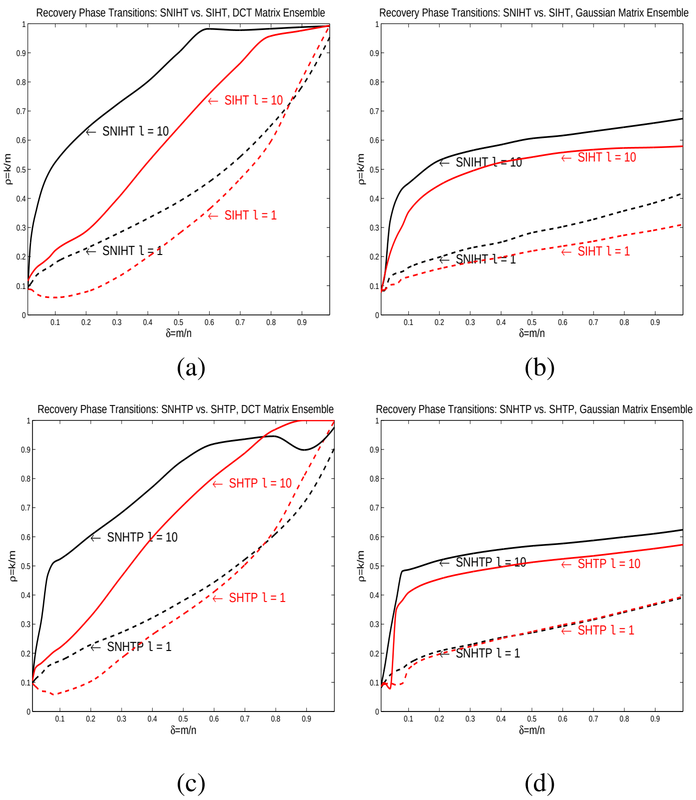


Fig. 6. Empirical weak recovery phase transitions: fixed versus optimal Step Size. SIHT versus SNIHT (a),(b) and SHTP versus SNHTP (c),(d). Matrix ensembles DCT (left panels) and Gaussian (right panels).

and SNHTP have similar performance under the Gaussian ensemble; when A is drawn from the DCT ensemble there is a more pronounced improvement of SNHTP over SHTP . For A drawn from the DCT ensemble, both SNIHT and SNHTP show an increased performance improvement over SIHT and SHTP, respectively, than for A drawn from the Gaussian ensemble.

2) Exact Recovery: For the exact recovery scenario, all three algorithms, SNIHT, SHTP, SCoSaMP show improved performance when A is drawn from the DCT ensemble rather than the Gaussian ensemble. The lone exception to this observation is SCoSaMP in the region $m/n \to 0$. Figure 7 shows the empirical weak phase transitions for both the DCT and Gaussian ensembles under the same experimental set-up as described in Section III-B. For all three algorithms, the ratio of the area below the recovery phase transition curves for $l = 2, 5, 10$ compared to the area below the curve for $l = 1$ are given in Table II.

For A from the DCT ensemble, the discrepancy between the three algorithms' performance is reduced as shown in Figure 8. All three algorithms behave similarly through most of the phase space although SCoSaMP demonstrates a difficulty for $m/n \to 0$, a finding consistent with that in [24].

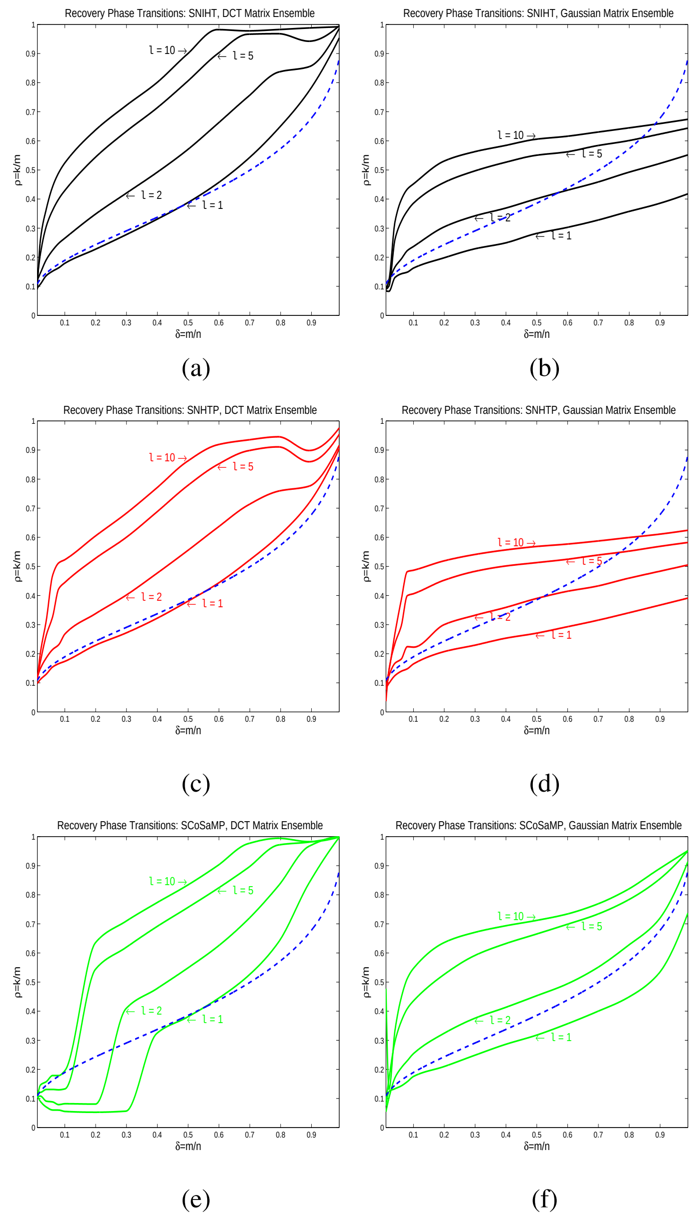


Fig. 7. Empirical weak recovery phase transitions for various joint sparsity levels with matrix ensembles DCT (left panels) and Gaussian (right panels). SNIHT (a),(b); SNHTP (c),(d); SCSMPSP (e),(f).

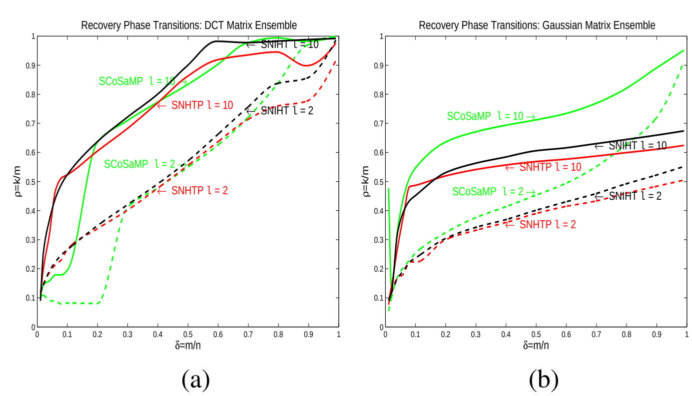


Fig. 8. Weak Recovery Phase Transitions with joint sparsity levels $l = 2, 10$ with matrix ensembles DCT (left panels) and Gaussian (right panels).