

Thresholding Orthogonal Matching Pursuit for Imaging Applications*

Hai Le[†] and Alexei Novikov[‡]

Abstract. This paper considers the problem of solving for sparse solutions of linear systems that arise from imaging applications. The theory of compressive sensing provides the framework to solve such problem with sparse recovery algorithms. The performance of these algorithms depends on whether the columns of measurement matrices are not coherent. However, the measurement matrices coming from imaging problems may not satisfy coherence criteria. We, therefore, are interested in sparse recovery algorithms for such measurement matrices. We use the concept of *vicinity* to analyze nearby points in the imaging scene that create high coherence in the measurement matrices. We present **Thresholding Orthogonal Matching Pursuit** (TOMP), a modified version of the classical OMP with a new stopping criteria. We prove theoretical guarantees that TOMP can recover the locations of the point sources exactly even when the measurement matrices are not coherent and the data is noisy. We then demonstrate the effectiveness of TOMP with numerical simulations in passive array imaging.

16 **Key words.** Passive Array Imaging, OMP, Vicinity

17 AMS subject classifications. 68U10, 94A08

1. Introduction. This work is motivated by imaging problems that can be cast as a linear system

$$20 \quad (1.1) \qquad \qquad \qquad \mathcal{A}\bar{\rho} + e = b,$$

where $\mathcal{A} \in \mathbb{C}^{N \times K}$ is the *measurement matrix*, $\mathbf{b} \in \mathbb{C}^N$ is the *data vector*, the unknown source vector $\bar{\rho} = (\bar{\rho}_1, \dots, \bar{\rho}_k) \in \mathbb{C}^K$ represents the image, and $\mathbf{e} \in \mathbb{C}^N$ is an unknown noise vector that comes from either the environment or numerically rounding errors.

24 Typically, the system (1.1) is *underdetermined* because we can only gather a few mea-
 25 surements, so $N \ll K$. When $N \ll K$, the system (1.1) has infinitely many solutions, so
 26 it is not possible to solve this system uniquely without additional a priori information. We
 27 consider, however, imaging problems with sparse scenes. The goal is to locate the positions
 28 and amplitudes of a small number of M of point sources that illuminate an array of re-
 29 ceivers. That is we are interested in finding the support of $\bar{\rho}$, which is denoted by $\text{supp}(\bar{\rho})$,
 30 and the magnitudes of $\bar{\rho}_i$. In this context, the source vector $\bar{\rho}$ is assumed to be M -sparse,
 31 which means it has M nonzero entries and $M \ll K$. The assumption $M \ll K$ places the
 32 system (1.1) under the *compressive sensing* framework and *sparse recovery algorithms* are
 33 used to solve such problem. We refer to [17] for a comprehensive review of the development
 34 of the field. Popular methods include Basis Pursuit Denoising (BPDN) [5], LASSO [22],
 35 Orthogonal Matching Pursuit (OMP) [24], CoSaMP [21].

*Submitted to the editors DATE.

Funding: This work was funded by the grants NSF DMS-1813943 and AFOSR FA9550-20-1-0026.

[†]Department of Mathematics, Pennsylvania State University, University Park, PA (hvl2@psu.edu, novikov@psu.edu).

In theory, the effectiveness of sparse recovery algorithms in solving the system (1.1) often relies on conditions about orthogonality of columns of the matrix \mathcal{A} . For example, when $e = 0$, the unique M -sparse solution of (1.1) can be obtained with l_1 -norm minimization when \mathcal{A} is assumed to be incoherent, meaning its mutual coherence is smaller than $1/(2M)$. (The mutual coherence of \mathcal{A} is defined as $\max_{i \neq j} |\langle \mathbf{a}_i, \mathbf{a}_j \rangle|$ where $\mathbf{a}_i \in \mathbb{C}^N$ are column vectors of \mathcal{A} and are normalized to one.) The same result can be stated when \mathcal{A} satisfies a restricted isometry property (RIP) [4], which essentially states that a set of columns of \mathcal{A} behave approximately like an orthogonal system. In our imaging problems, however, these types of conditions are hardly ever satisfied. Furthermore, we may have many almost collinear columns of \mathcal{A} . The reason is that we usually want high resolution in the image, and to do so, we discretize the imaging scene with finer grids and effectively increase the coherence of \mathcal{A} . This motivates us to look for algorithms that can be guaranteed to work even when the matrix \mathcal{A} is not fully incoherent. One possible approach to relax the coherence condition on \mathcal{A} is to use the following concept of *vicinity* which was investigated in [19] (it was introduced earlier as a *coherence band* in [11]).

Definition 1.1. (*Vicinity*) For any $i \in \text{supp}(\bar{\rho})$ define the corresponding η -vicinity of \mathbf{a}_i as

$$(1.2) \quad \text{vic}_\eta(\mathbf{a}_i) = \{k : |\langle \mathbf{a}_k, \mathbf{a}_i \rangle| \geq \eta\}.$$

Here the column vectors \mathbf{a}_i are normalized to one.

Roughly speaking, a vicinity of a given vector contains vectors that are almost parallel to the given one. We then can analyze the problem by only assuming conditions on these vicinities instead of the whole set of columns of \mathcal{A} . Indeed, results in [19] and [11] demonstrate that popular methods like l_1 -norm minimization or OMP work quite well under such conditions. In the presence of noise, [19, 11] demonstrated that the locations of the M point sources can be recovered approximately within their vicinities, but the image may still be corrupted by noise outside of the vicinities.

In this paper, we present **Thresholding Orthogonal Matching Pursuit** (TOMP), an algorithm that is guaranteed to recover the support exactly even in the presence of noise. It combines the structure of OMP and a stopping procedure to effectively select the true support. To this end, we can informally state our theoretical findings as follows, and the details will be presented as [Theorem 5.2](#) in [section 5](#).

Theorem 1.2. (Exact Recovery of TOMP) If the number of sources is not too large, their locations are not coherent and the level of noise is small, then TOMP recovers the support of $\bar{\rho}$, the true solution of (1.1).

The paper is organized as follows. We explain the passive array imaging problem and how to formulate it as a linear algebra problem in [section 2](#). In [section 3](#), we review some methods that can be used to solve the linear system. We then introduce our algorithm TOMP in [section 4](#). We state theoretical results about TOMP in [section 5](#), Numerical results about TOMP will be presented in [section 6](#) and the proofs of main theorems are given in [section 7](#).

76 **2. Passive Array Imaging.** Consider the problem of imaging/locating radiative sources
 77 from passive arrays of receivers. The objective is to determine the positions \mathbf{z}_j and the
 78 complex amplitudes α_j , $j = 1, \dots, M$, of the point sources from measurements obtained
 79 on an array of receivers. The imaging system is characterized by the array aperture a , the
 80 distance L from the array to the sources, the bandwidth B and the central wavelength λ_0 .

81 The sources are located inside a region of interest called the image window (IW), which
 82 is discretized with a uniform grid of points \mathbf{y}_k , $k = 1, \dots, K$. (Here and below, we use
 83 bold letters $\mathbf{a}, \mathbf{b}, \dots$ for column vectors.) We assume that the sources lie on the grid. The
 84 unknown is the *true source vector*

$$85 \quad \bar{\boldsymbol{\rho}} = [\bar{\rho}_1, \dots, \bar{\rho}_K]^\top \in \mathbb{C}^K$$

86 whose components $\bar{\rho}_k$ correspond to the complex amplitudes of the M sources at the grid
 87 points \mathbf{y}_k , $k = 1, \dots, K$. For the true source vector we have $\bar{\rho}_k = \alpha_j$ if $\mathbf{y}_k = \mathbf{z}_j$ for some
 88 $j = 1, \dots, M$, while $\bar{\rho}_k = 0$ otherwise.

89 Denoting by $G(\mathbf{x}, \mathbf{y}; \omega)$ the Green's function for the propagation of a signal of angular
 90 frequency ω from point \mathbf{y} to point \mathbf{x} , we define the single-frequency Green's function vector
 91 that connects a point \mathbf{y} in the IW with all points \mathbf{x}_r , $r = 1, \dots, N_r$, on the array as

$$92 \quad \mathbf{g}(\mathbf{y}; \omega) = [G(\mathbf{x}_1, \mathbf{y}; \omega), G(\mathbf{x}_2, \mathbf{y}; \omega), \dots, G(\mathbf{x}_{N_r}, \mathbf{y}; \omega)]^\top \in \mathbb{C}^{N_r}$$

93 representing the signal received at the array due to a source of amplitude one, phase zero,
 94 and frequency ω at \mathbf{y} . If the medium is homogeneous, the Green's function is $G(\mathbf{x}, \mathbf{y}; \omega) =$
 95 $e^{i\omega\|\mathbf{x}-\mathbf{y}\|_2/c_0}/(4\pi\|\mathbf{x}-\mathbf{y}\|_2)$. The data for the imaging problem are the signals

$$96 \quad (2.1) \quad b(\mathbf{x}_r, \omega_l) = \sum_{j=1}^M \alpha_j G(\mathbf{x}_r, \mathbf{z}_j; \omega_l)$$

97 recorded at receiver locations \mathbf{x}_r , $r = 1, \dots, N_r$, at frequencies ω_l , $l = 1, \dots, S$. These data
 98 are stacked in a column vector

$$99 \quad (2.2) \quad \mathbf{b} = [b(\omega_1)^\top, b(\omega_2)^\top, \dots, b(\omega_S)^\top]^\top \in \mathbb{C}^N; \quad N = N_r \cdot S,$$

100 with $b(\omega_l) = [b(\mathbf{x}_1, \omega_l), b(\mathbf{x}_2, \omega_l), \dots, b(\mathbf{x}_{N_r}, \omega_l)]^\top \in \mathbb{C}^{N_r}$, $l = 1, \dots, S$. Then, the source
 101 vector $\bar{\boldsymbol{\rho}}$ solves the system

$$102 \quad (2.3) \quad \mathcal{A} \bar{\boldsymbol{\rho}} = \mathbf{b},$$

103 with \mathcal{A} being $N \times K$ matrix whose columns \mathbf{a}_k are the multiple-frequency Green's function
 104 vectors

$$105 \quad (2.4) \quad \mathbf{a}_k = [\mathbf{g}(\mathbf{y}_1; \omega_1)^\top, \mathbf{g}(\mathbf{y}_1; \omega_2)^\top, \dots, \mathbf{g}(\mathbf{y}_1; \omega_S)^\top]^\top \in \mathbb{C}^N, \quad 1 \leq k \leq K,$$

106 which are normalized to have length 1.

107 **3. Sparsity Promoting Algorithms.** We want to benefit from our assumption that the
 108 point sources only occupy a small fraction of the image window. This means that the
 109 vector $\bar{\rho}$ is sparse and $M = |\text{supp}(\bar{\rho})| \ll K$, say, $K = N^\beta$ for some $\beta \geq 1$. (We denote
 110 $\text{supp}(\rho)$ as the set of indices of nonzero entries of a vector ρ , and call that the support
 111 of this vector.) This prior knowledge makes solving the system (2.3) fall under the scope
 112 of compressive sensing. If there is no noise, the system (2.3) can be solved by solving the
 113 Basis Pursuit problem [18]:

114 (3.1) $\rho_{l_1} = \underset{\rho}{\operatorname{argmin}} \|\rho\|_1 \quad \text{s.t. } \mathcal{A}\rho = b.$

115 where we used $\|\rho\|_p$ to denote the l_p -norm of a vector ρ .

116 It can be proven that the solution of the optimization problem (3.1) recovers the true
 117 source vector $\bar{\rho}$ exactly under certain conditions on the measurement matrix \mathcal{A} and sparsity
 118 level M [3, 6, 10, 12]. These conditions often involve assumptions on incoherence of the
 119 columns of the matrix \mathcal{A} . For example, if we assume columns of \mathcal{A} satisfy

120 (3.2) $|\langle \mathbf{a}_i, \mathbf{a}_j \rangle| < \frac{1}{2M}, \quad \text{for all } i \neq j,$

121 (where $\langle \mathbf{u}, \mathbf{v} \rangle := \bar{\mathbf{u}}^T \mathbf{v}$ is the complex inner product between two complex vectors \mathbf{u} and
 122 \mathbf{v}), then the solution of (3.1) equals the true source vector $\bar{\rho}$ [9].

123 However, in imaging applications, the condition (3.2) may be violated. Furthermore,
 124 $\langle \mathbf{a}_i, \mathbf{a}_j \rangle$ is often very close to 1 when \mathbf{a}_i and \mathbf{a}_j correspond to nearby points in the IW. The
 125 reason is that in imaging we may want high resolution, so we need to choose very small
 126 meshsize of the grid of the IW. However, making meshsize small increases coherence and
 127 also increases the number of columns K of the matrix \mathcal{A} . The theory in [19] suggests that
 128 the l_1 -method works well even when the condition (3.2) is violated and even demonstrates
 129 exact recovery in the noiseless case. The authors found that the condition (3.2) does not
 130 need to be satisfied for the whole matrix \mathcal{A} for proving theoretical results, we only need to
 131 assume incoherence on the support of the true source vector $\bar{\rho}$. This means that the point
 132 sources should stay far apart from each other. The results in [19] state that as long as
 133 the vicinities (defined in (1.2)) corresponding to points on the support do not overlap, the
 134 l_1 -method (3.1) recovers the true source vector $\bar{\rho}$ exactly. In the earlier work [11] support
 135 recovery was obtained under more restrictive conditions on vicinities. In our paper, the
 136 goal will be similar. We want to develop an algorithm that recovers exactly $\text{supp}(\rho)$ under
 137 a less stringent condition than (3.2), even in the noisy case.

138 Suppose the data vector b is now corrupted by a noise vector e , that is $b = \mathcal{A}\bar{\rho} + e$.
 139 In this case, the exact recovery of $\bar{\rho}$ is no longer possible. Nevertheless, recovering the
 140 support of $\bar{\rho}$ is still feasible under certain conditions. A popular approach is a modification
 141 of (3.1), the Basis Pursuit Denoising (BPDN) [5] or the Least Absolute Shrinkage and
 142 Selection Operator (LASSO) [22]. It is an l_2 -optimization method that promotes sparsity
 143 by penalizing l_1 -norm:

144 (3.3) Find $\rho_\lambda^{LASSO} \in \underset{\rho}{\operatorname{argmin}} \left(\frac{1}{2} \|b - \mathcal{A}\rho\|_2^2 + \lambda \|\rho\|_1 \right), \quad \lambda \geq 0.$

145 Convexity ensures that there is always a solution to (3.3). The tuning/penalty parameter
 146 λ must be appropriately chosen to obtain the desired properties of the minimizer ρ_λ^{LASSO} .
 147 As λ increases, BPDN will choose the minimizer with fewer non-zero entries. On the other
 148 hand, smaller λ will produce a solution closer to the least-squares approximation. Given
 149 the level of noise one can choose λ optimally so that the support of ρ_λ^{LASSO} recovers as
 150 much of the support of the true solution $\bar{\rho}$ as possible, see for example, [13, 23, 26]. Some
 151 examples to choose λ include *cross-validation* [22] and choosing λ *adaptively* [27, 7]. These
 152 methods may be computationally expensive and/or require additional estimates of, for
 153 example, the level of noise $\|e\|_2$.

154 *Square-Root LASSO* (SQRT-LASSO) [2] is a related version of (3.3) that does not re-
 155 quire knowledge of the level of noise in choosing parameters. SQRT-LASSO is the following
 156 optimization problem

157 (3.4)
$$\text{Find } \rho_\lambda^{SQRT} \in \operatorname{argmin} (\|\mathbf{b} - \mathcal{A}\rho\|_2 + \lambda\|\rho\|_1), \quad \lambda \geq 0.$$

158 The functionals in (3.4) and in (3.3) differ in the exponent of the l_2 -norm. This difference
 159 makes SQRT-LASSO attractive, because it allows us to choose λ independent of knowledge
 160 of noise in the Square-Root LASSO approach [2]. The authors called their method pivotal
 161 with respect to λ . Our numerical experiments in Section 6 corroborate that SQRT-LASSO
 162 can recover $\bar{\rho}$ quite effectively in the noisy case: the true support is recovered, along with
 163 very few nearby points, and the recovered image is quite precise.

164 SQRT-LASSO may be a starting point for derivation of greedy sparse recovery algo-
 165 rithms. We demonstrate it in Section 4, where we derive two such algorithms: Stagewise
 166 Orthogonal Matching Pursuit (StOMP) and Thresholding Orthogonal Matching Pursuit
 167 (TOMP). Both of them keep the pivotal property of SQRT-LASSO. Both of them work well
 168 if columns of \mathcal{A} are incoherent. StOMP is much faster than TOMP. Yet, TOMP is better
 169 for imaging, when columns of \mathcal{A} are not incoherent. We demonstrate that TOMP recovers
 170 the true support exactly when columns of \mathcal{A} are not incoherent, even in the noisy case.
 171 We rigorously show that the exact recovery happens under certain conditions on vicinities,
 172 which are less stringent than (3.2). As the name suggests, our TOMP is a modified version
 173 of the well-known Orthogonal Matching Pursuit (OMP) [24]. The main difference is while
 174 OMP is required to run in M (the sparsity level) iterations, TOMP can automatically
 175 terminate after all the locations in the true support have been recovered because of an
 176 additional pivotal stopping criteria based on thresholding.

177 **4. Thresholding Orthogonal Matching Pursuit.**

178 **4.1. From Square-Root LASSO to StOMP to TOMP.** Consider the Square-Root
 179 LASSO [2] minimization problem given in (3.4). Recall that $\bar{\rho}$ is the solution of (1.1) that
 180 we are trying to recover. In [2], the authors show that with the following choice of τ ,

181 (4.1)
$$\tau \asymp \frac{\sqrt{\log K}}{\sqrt{N}},$$

182 the solution ρ_τ^{SQRT} to (3.4) achieves the near-oracle performance, which roughly means that
 183 $\|\rho_\tau^{SQRT} - \bar{\rho}\|_2$ is small and bounded by known parameters and $\|e\|_2$. Unlike in the LASSO

184 method (3.3), the choice of the parameter τ in (4.1) does not depend on the strength of
 185 the noise $\|e\|_2$, which is not known in practice. This makes Square-Root LASSO tuning-
 186 free/pivotal with respect to noise. We apply proximal gradient descent to solve (3.4). Using
 187 ISTA [1] as an example, the method states that we can find the minimum of

188
$$\frac{1}{2}\|\mathbf{b} - \mathcal{A}\rho\|_2^2 + \tau\|\rho\|_1 =: f(\rho) + g(\rho) = F(\rho),$$

189 by iteratively running

190 (4.2)
$$\rho^0 \in \mathbb{C}^K, \quad \rho^{n+1} = \mathcal{L}_{\tau\Delta\rho}(\rho^n - \Delta\rho\nabla f(\rho^n)),$$

where $\nabla f(\rho) = \mathcal{A}^*(\mathcal{A}\rho - \mathbf{b})$, the (soft) thresholding operator

$$\mathcal{L}_\lambda(u) = \text{sign}(u) \max(|u| - \lambda, 0)$$

191 (when u is a vector, this operator acts component-wise), and stepsize $\Delta\rho > 0$. The iteration
 192 (4.2) can be seen as the usual gradient descent with an addition of thresholding procedure
 193 to promote sparsity. Under certain conditions on \mathcal{A} and $\bar{\rho}$, we can show that ρ^n converges
 194 to the minimizer of $F(\rho)$ [16]. We will apply proximal gradient descent method to solve
 195 (3.4). The resulting algorithm is [Algorithm 4.1](#).

Algorithm 4.1 Square-Root LASSO by Proximal Gradient Descent

INPUT: measurement matrix $\mathcal{A} \in \mathbb{C}^{N \times K}$,

data vector $\mathbf{b} \in \mathbb{C}^N$,

parameter $\tau > 0$,

stepsize $\Delta\rho > 0$,

number of iterations $\tilde{n} > 0$.

OUTPUT: a vector $\rho^{\tilde{n}} \in \mathbb{C}^K$.

$$\rho^0 \leftarrow 0$$

for $n = 0, 1, \dots, \tilde{n} - 1$

$$\rho^{n+1} \leftarrow \mathcal{L}_{\tau\Delta\rho} \left(\rho^n + \Delta\rho \cdot \frac{\mathcal{A}^*(\mathbf{b} - \mathcal{A}\rho^n)}{\|\mathbf{b} - \mathcal{A}\rho^n\|_2} \right).$$

end

196 The quotient term comes directly from differentiating $\|\mathbf{b} - \mathcal{A}\rho\|_2$ with respect to ρ .
 197 Theory for convergence of ρ^n to the sparse solution of (3.4) has been addressed in [15,
 198 Theorem 3.3]. We instead look for a modification to speed up [Algorithm 4.1](#). This mod-
 199 ification will also slightly change the underlying goal from signal estimation to support
 200 recovery. This means at each iteration, we will try to recover part of $\text{supp}(\bar{\rho})$ by using
 201 the thresholding procedure, and then we immediately update the data vector \mathbf{b} to remove
 202 the detected support; the process then repeats. The main advantage of our modification
 203 is to quickly identify the support at each iteration without spending time estimating the
 204 magnitude at each location. The resulting algorithm is [Algorithm 4.2](#), and it has been
 205 derived in [8] under the name of *Stagewise Orthogonal Matching Pursuit* (StOMP).

Algorithm 4.2 Stagewise Orthogonal Matching Pursuit (StOMP)**INPUT:** measurement matrix $\mathcal{A} \in \mathbb{C}^{N \times K}$,data vector $\mathbf{b} \in \mathbb{C}^N$,parameter $\tau > 0$.**OUTPUT:** a set Ω of indices of columns of \mathcal{A} .
$$\Omega^0 \leftarrow \emptyset, \mathbf{b}^0 \leftarrow \mathbf{b} \quad (\text{Initialization})$$
repeat

$$\tilde{\Omega}^{n+1} \leftarrow \{1 \leq i \leq K : |\langle \mathbf{a}_i, \mathbf{b}^n \rangle| > \tau \|\mathbf{b}^n\|_2\} \quad (\text{Thresholding})$$

$$\text{If } \tilde{\Omega}^{n+1} = \emptyset, \text{ then break;} \quad (\text{Stopping Criterion})$$

$$\Omega^{n+1} \leftarrow \Omega^n \cup \tilde{\Omega}^{n+1} \quad (\text{Support Update})$$

$$\mathbf{b}^{n+1} \leftarrow \mathbf{b} - \mathcal{A}_{\Omega^{n+1}} \mathcal{A}_{\Omega^{n+1}}^\dagger \mathbf{b} \quad (\text{Projection})$$
until stopped

206 We denote the pseudoinverse of a full-rank matrix \mathcal{A} as $\mathcal{A}^\dagger := (\mathcal{A}^* \mathcal{A})^{-1} \mathcal{A}^*$, where for
 207 any S , set of indices, \mathcal{A}_S is the matrix with columns of \mathcal{A} drawn from the set S . Note
 208 that the matrix operator $\mathcal{A}_S \mathcal{A}_S^\dagger$ represents the orthogonal projection onto the vector space
 209 spanned by columns of \mathcal{A} indexed by the set S .

210 Note that the output of [Algorithm 4.2](#) at each iteration is a set of indices Ω^n instead of a
 211 vector $\rho^n \in \mathbb{C}^K$ that approximates $\bar{\rho}$. In other words, [Algorithm 4.2](#) focuses on recovering
 212 the support of $\bar{\rho}$ rather than recovering $\bar{\rho}$ itself. One could certainly find ρ^n once its support
 213 is known. We derived [Algorithm 4.2](#) essentially by setting $\rho^n = 0$ in the thresholding term
 214 in [Algorithm 4.1](#) and looking at the support of the vector $\mathcal{L}_{\tau \Delta \rho}(\Delta \rho \mathcal{A}^* \mathbf{b} / \|\mathbf{b}\|_2)$, which will
 215 be included in the discovered set of indices. Moreover, at each iteration, instead of \mathbf{b} , we
 216 will use \mathbf{b}^n that is obtained by orthogonally projecting \mathbf{b} in order to remove dependence
 217 on the indices discovered from the previous iteration. We then see that the stepsize $\Delta \rho$ is
 218 now not needed and this explains how the set $\tilde{\Omega}^{n+1}$ is defined in the **Thresholding** step
 219 of [Algorithm 4.2](#).

220 The role of the Thresholding step is also to control the noise. As we demonstrate in
 221 our proofs, one could choose τ so that noise cannot corrupt the image with probability
 222 arbitrarily close to one.

223 StOMP, however, may and will fail to find the correct support of $\bar{\rho}$ when columns of
 224 \mathcal{A} are not incoherent. It happens because the inner products $|\langle \mathbf{a}_i, \mathbf{b}^n \rangle|$ approximately have
 225 the same value when vectors \mathbf{a}_i correspond to points that lie inside a vicinity of a true
 226 source. Therefore, at each iteration, StOMP will find many false sources in vicinities of
 227 each true source. We claim we can overcome this issue, if we only pick the index with the
 228 largest inner product $|\langle \mathbf{a}_i, \mathbf{b}^n \rangle|$. That is, instead of recovering multiple locations at each
 229 iteration like in [Algorithm 4.2](#), we choose only the one that passes the thresholding step
 230 and correlates the most with \mathbf{b}^n . The **Projection** step will then remove the impact of
 231 this source. It implies that none of the points in its vicinity will be chosen in the next
 232 iterations. Hence no false sources will be found. The role of the Thresholding step is again
 233 to control the noise.

234 **4.2. Description of TOMP.** The Thresholding Orthogonal Matching Pursuit has the
 235 following structure.

Algorithm 4.3 Thresholding Orthogonal Matching Pursuit (TOMP)

INPUT: measurement matrix $\mathcal{A} \in \mathbb{C}^{N \times K}$, data vector $\mathbf{b} \in \mathbb{C}^N$, parameter $\tau > 0$.

OUTPUT: a set Ω of indices of columns of \mathcal{A} .

$\Omega^0 \leftarrow \emptyset, \mathbf{b}^0 \leftarrow \mathbf{b}$ repeat $\tilde{\Omega}^{n+1} \leftarrow \{1 \leq i \leq K : \langle \mathbf{a}_i, \mathbf{b}^n \rangle > \tau \ \mathbf{b}^n\ _2\}$ If $\tilde{\Omega}^{n+1} = \emptyset$, then break ; $i_{\max} \leftarrow \operatorname{argmax}_{i \in \tilde{\Omega}^{n+1}} \langle \mathbf{a}_i, \mathbf{b}^n \rangle $ $\Omega^{n+1} \leftarrow \Omega^n \cup \{i_{\max}\}$ $\mathbf{b}^{n+1} \leftarrow \mathbf{b} - \mathcal{A}_{\Omega^{n+1}} \mathcal{A}_{\Omega^{n+1}}^\dagger \mathbf{b}$ until stopped	(Initialization) (Thresholding) (Stopping Criterion) (Selection) (Support Update) (Projection)
---	---

236 We explain each step in details:

- 237 • **Initialization:** The first approximation Ω^0 to the true support is the empty set.
- 238 • **Thresholding:** This step selects all indices that make the inner product $|\langle \mathbf{a}_i, \mathbf{b}^n \rangle|$
 239 bigger than $\tau \|\mathbf{b}^n\|_2$. The idea is that those column vectors with indices belonging
 240 to $\operatorname{supp}(\bar{\rho})$ will have large correlations with \mathbf{b}^n . The thresholding parameter τ is
 241 to control how large the inner product we want to select.
- 242 • **Stopping Criterion:** The algorithm will stop if nothing new is detected in the
 243 thresholding step.
- 244 • **Selection:** The index i_{\max} that makes the largest inner product $|\langle \mathbf{a}_i, \mathbf{b}^n \rangle|$ will be
 245 selected.
- 246 • **Support Update:** New index set Ω^{n+1} is created by merging the old index set
 247 Ω^n and the new location i_{\max} .
- 248 • **Projection:** We need to remove the part of data vector \mathbf{b} that approximates the
 249 locations that are already detected. This is done by projecting \mathbf{b} onto the space
 250 that is orthogonally complement to the space spanned by columns of $\mathcal{A}_{\Omega^{n+1}}$.

251 We can obtain the recovered signal with the support Ω by computing $\mathcal{A}_\Omega^\dagger \mathbf{b}$, which is the
 252 solution to the l_2 -minimization problem $\min_{\rho} \|\mathcal{A}_\Omega \rho - \mathbf{b}\|_2$.

253 We remark that [Algorithm 4.3](#) is also a simple modification of OMP. It adds a thresh-
 254 olding step before selecting the index with largest inner product. We will demonstrate that
 255 this step yields two advantages over OMP: 1) We do not need to specify how many iter-
 256 ations it will run. Unlike OMP where it performs optimally when the number of iterations
 257 is M (sparsity level), which is not known a priori in practice, our algorithm automatically
 258 stops when every nonzero location has been recovered; 2) The thresholding step acts as
 259 a noise filter. In the noisy case, this step helps the algorithm not create false discoveries
 260 caused by the noise. For the above two reasons, choosing τ appropriately is important for
 261 the performance of [Algorithm 4.3](#). We will demonstrate in [section 5](#) how to choose this
 262 thresholding parameter independently of the level of noise.

263 **5. Main Result.** Our main contribution is a theorem that guarantees TOMP is able
 264 to recover signals in coherent settings with or without noise. To simplify notations, we will
 265 denote $\rho = (\rho_1, \dots, \rho_K)$ (instead of $\bar{\rho}$) the true source vector that we are trying to recover.

266 We restate the definition of vicinity of readers' convenience

267 **Definition 5.1.** (*Vicinity*) For any $i \in \text{supp}(\bar{\rho})$ define the corresponding η -vicinity of \mathbf{a}_i
 268 as

269 (5.1) $\text{vic}_\eta(\mathbf{a}_i) = \{k : |\langle \mathbf{a}_k, \mathbf{a}_i \rangle| \geq \eta\} .$

270 We then list the following conditions on the vicinities and the size of the noise

- 271 1. For $i_1 \neq i_2$, $i_1, i_2 \in \text{supp}(\rho)$, and $j \in \text{vic}_\eta(\mathbf{a}_{i_1})$, we have

272 (5.2) $|\langle \mathbf{a}_j, \mathbf{a}_{i_2} \rangle| \leq \frac{c}{\sqrt{N}} \quad \text{for some } c > 0.$

- 273 2. For $j \in \text{vic}_\eta(\mathbf{a}_i)$, where $i \in \text{supp}(\rho)$, we have

274 (5.3) $|\langle \mathbf{a}_j, \mathbf{a}_i \rangle| < 1 - \gamma \left(\frac{2c_1 M}{\sqrt{N}} + \frac{8c_1^2 M^2}{3N} \right) - \frac{1}{\rho_{\min}} \frac{2c_0 \sqrt{\log N}}{\sqrt{N}} \|\mathbf{e}\|_2$

275 with c_0 specified later, c_1 being the smallest c that (5.2) holds, $\gamma := \rho_{\max}/\rho_{\min}$
 276 where $\rho_{\min} := \min_{i \in \text{supp}(\rho)}(|\rho_i|)$ and $\rho_{\max} := \max_{i \in \text{supp}(\rho)}(|\rho_i|)$.

- 277 3. The noise level satisfies

278 (5.4) $\|\mathbf{e}\|_2 \leq \frac{\sqrt{N} \rho_{\min}}{2c_0 \sqrt{\log N}} \left(\frac{2}{3} - M\eta - \frac{4\eta^2 M^2}{3} \right),$

279 with the same c_0 as in (5.3).

280 We also define the mutual coherence parameter on the support of ρ ,

281 (5.5) $\mu := \max_{i,j \in \text{supp}(\rho), i \neq j} |\langle \mathbf{a}_i, \mathbf{a}_j \rangle|.$

282 We are ready to state our main theorem as follows.

283 **Theorem 5.2. (Exact Recovery of TOMP)** Let $\mathcal{A} \in \mathbb{C}^{N \times K}$, $K = N^\beta$. Let ρ be an
 284 M -sparse solution of $\mathcal{A}\rho = \mathbf{b}$. Let the noise \mathbf{e} be such that $\mathbf{e}/\|\mathbf{e}\|_2$ is uniformly distributed
 285 on the unit sphere \mathbb{S}^{N-1} . Denote Ω to be the output of TOMP with inputs measurement
 286 matrix \mathcal{A} and data vector $\mathbf{b} + \mathbf{e}$. For any $\kappa > 0$, setting $c_0 = \sqrt{2(\beta + \kappa)}$, if the conditions
 287 (5.2), (5.3), (5.4) hold, the sparsity level M satisfies

288 (5.6) $M \leq \min \left\{ \frac{1}{4\mu}, \frac{\sqrt{N}}{5c_0 \sqrt{\log N}} \right\},$

290 and the thresholding parameter is

291
$$\tau := \frac{c_0 \sqrt{\log N}}{\sqrt{N}},$$

292 then $\Omega = \text{supp}(\rho)$ with probability $1 - 4/N^\kappa$.

We remark that the theorem still holds when noise is not rotationally invariant, for example, if \mathbf{e} is assumed to be sub-gaussian. We just need to know the parameters for the tail behavior of $|\langle \mathbf{a}, \mathbf{e} \rangle|$. The proof of [Theorem 5.2](#) relies on the following proposition about the set Ω^n at each iteration.

[Proposition 5.3](#). *Assume $\mathcal{A}, \mathbf{b}, \kappa, c_0, \mathbf{e}$ as in [Theorem 5.2](#) with conditions [\(5.2\)](#), [\(5.3\)](#), [\(5.4\)](#), and the sparsity level satisfies*

$$M \leq \frac{1}{4\mu},$$

then, with probability $1 - 4/N^\kappa$, at every $(n + 1)^{\text{th}}$ -iteration ($n \geq 0$), with $(\Omega^n)^c = \text{supp}(\boldsymbol{\rho}) \setminus \Omega^n$, we have the followings hold

- *For $i \in (\Omega^n)^c$ and $j \in \text{vic}_\eta(\mathbf{a}_i)$, $j \neq i$, we have*

$$(5.7) \quad |\langle \mathbf{a}_j, \mathbf{b}^n \rangle| < |\langle \mathbf{a}_i, \mathbf{b}^n \rangle|.$$

- *Let $i^* \in (\Omega^n)^c$ be an index such that $|\rho_{i^*}| = \max_{i \in (\Omega^n)^c} (\rho_i)$. For j such that $j \notin \text{vic}_\eta(\mathbf{a}_i)$ for all $i \in \text{supp}(\boldsymbol{\rho})$, we have*

$$(5.8) \quad |\langle \mathbf{a}_j, \mathbf{b}^n \rangle| < |\langle \mathbf{a}_{i^*}, \mathbf{b}^n \rangle|.$$

Roughly speaking, the inequality [\(5.7\)](#) guarantees that TOMP selects some indices from vicinities at the Thresholding step, then the Selection step will choose among them an index from the support of the signal. On the other hand, inequality [\(5.8\)](#) implies that TOMP never selects an index that is not inside a vicinity of a true index. In other words, at every n^{th} -iteration, the set $\Omega^n \subset \text{supp}(\boldsymbol{\rho})$. The proofs of [Theorem 5.2](#) and [Proposition 5.3](#) will be given in [section 7](#).

6. Numerical Performance. In this section, we aim to demonstrate the advantages of TOMP [\(4.3\)](#) over Kirchhoff migration (KM)

$$315 \quad (6.1) \quad \boldsymbol{\rho}_{l_2} = \mathcal{A}^* \mathbf{b}.$$

[\(KM](#) is a standard array imaging method, which we use here only for comparison.), SQRT-LASSO [\(4.1\)](#), and StOMP [\(4.2\)](#). The settings of our imaging applications are as follows. We run numerics for a passive array imaging problem in a homogeneous medium. The array has $N = 25$ receivers and an aperture $a = 12.5$ cm. We consider central frequency $f_0 = 240$ GHz corresponding to $\lambda_0 = 1.75$ mm. We collect measurements corresponding to $S = 25$ equally spaced frequencies spanning a bandwidth $B = 60$ GHz. The distance from the array to the center of the imaging window is $L = 25$ cm. The size of the imaging window is 15×15 cm. It is discretized using a uniform grid of $K = 3721$ points. We seek to image a source vector with sparsity $M = 8$ with amplitudes distributing according to $N(0, 1)$. We run Kirchhoff migration, SQRT-LASSO, StOMP, and TOMP for this imaging problem at different levels of noise $\delta = \|\mathbf{e}\|_2/\|\mathbf{b}\|_2$: $\delta = 0$ (noiseless), $\delta = 0.5$ (moderate level of noise), and $\delta = 1$ (high level of noise). We also show the vicinities (with different colors) defined in [\(1.2\)](#) (with $\eta = 1/(4M)$) for each experiment. We also present the vicinity map

in each setting. The results are presented in Figures 1, 2, and 3. We can see that the vicinities overlap, and certainly, the assumptions $M < 1/(4\mu)$ and (5.3) may not satisfy. However, the numerics shows that TOMP still recovers the support exactly even when these conditions do not hold. We remark that our numerical results are not specialized to a particular physical regime. They only aim to illustrate the effectiveness of TOMP when the columns of matrices are not incoherent.

For optimal results, instead of using $\tau = 2\sqrt{\log N}/\sqrt{N}$, we calibrate τ to be the smallest number such that TOMP will output empty set when being fed with pure noise. This is done similarly as in [20] when τ is calibrated to satisfy the No Phantom Signal theorem in that paper. We perform the calibration as follows. We start with $\tau_{\text{failure}} = 0$ and $\tau_{\text{success}} = 1$ then calculate the average $\tau_{\text{mid}} = (\tau_{\text{failure}} + \tau_{\text{success}})/2$. We run the Thresholding step in Algorithm 4.3 at $\tau = \tau_{\text{mid}}$ with inputs \mathcal{A} constructed in (2.4) and $\mathbf{b} \in \mathbb{R}^N$ being Gaussian vector 50 times. If the outputs are the empty set for all of 50 times, τ_{mid} will be assigned to τ_{success} , and τ_{failure} otherwise. We stop the search when $|\tau_{\text{success}} - \tau_{\text{failure}}| \leq 10^{-3}$, and the final τ_{success} will be the thresholding parameter used for SQRT-LASSO, StOMP, and TOMP.

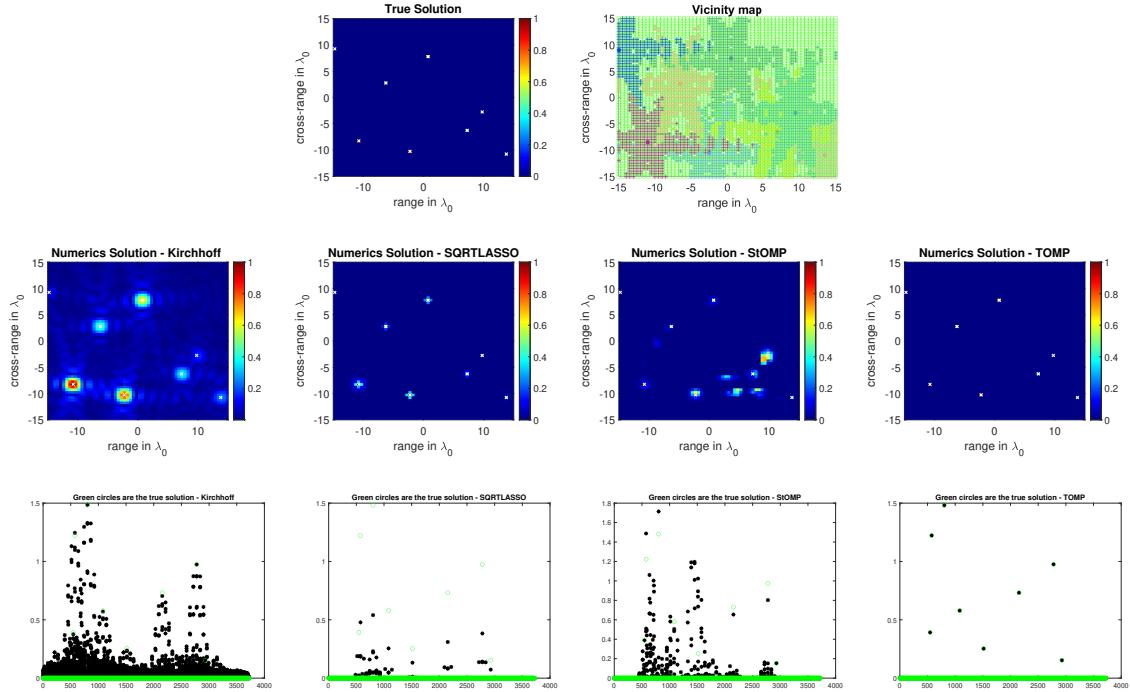


Figure 1. (The noise level is $\delta = 0$, there is no noise). In the top row, from left to right, the true image and the vicinity map. In the middle row, from left to right, the images recovered by Kirchhoff migration, SQRT-LASSO, StOMP, and TOMP. In the bottom row, from left to right, the solution vectors recovered by Kirchhoff migration, SQRT-LASSO, StOMP, and TOMP.

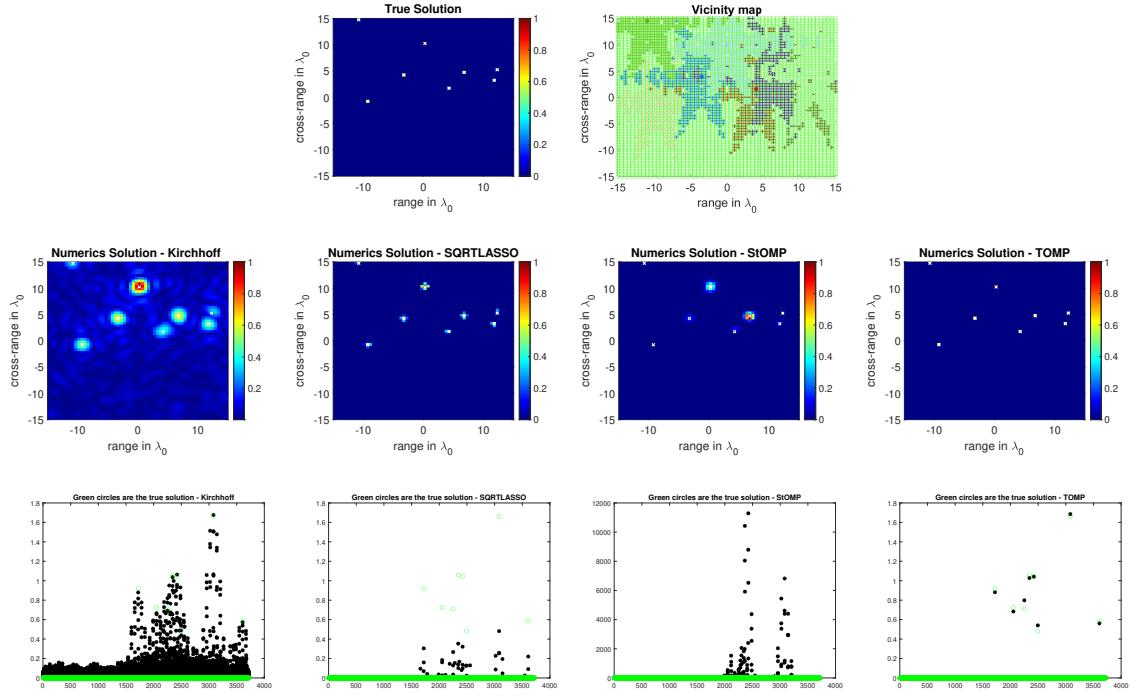


Figure 2. (The noise level is $\delta = 0.5$, so $SNR \approx 3dB$). In the top row, from left to right, the true image and the vicinity map. In the middle row, from left to right, the images recovered by Kirchhoff migration, SQRT-LASSO, StOMP, and TOMP. In the bottom row, from left to right, the solution vectors recovered by Kirchhoff migration, SQRT-LASSO, StOMP, and TOMP.

In all cases, KM produces very blurry images, StOMP and SQRT-LASSO have images with higher resolution with SQRT-LASSO being better by having less false discoveries, and TOMP recovers the image exactly. We also observe that even though the vicinities in each example overlap significantly, TOMP still performs perfectly. This supports our theory that TOMP is effective at detecting the locations of the sources while removing the effects of the vicinities.

350 7. Proofs of the Main Results.

351 7.1. Proof of Proposition 5.3.

352 *Proof.* Consider the events

$$353 \quad (7.1) \quad \mathcal{O}_1 = \left\{ \max_{1 \leq i \leq K} |\langle \mathbf{a}_i, \mathbf{e} \rangle| \leq c_0 \frac{\sqrt{\log N}}{\sqrt{N}} \|\mathbf{e}\|_2 \right\}$$

354 and

$$355 \quad (7.2) \quad \mathcal{O}_2 = \left\{ \max_{1 \leq i \leq K} |\langle (I - \mathcal{A}_{\text{supp}(\rho)} \mathcal{A}_{\text{supp}(\rho)}^\dagger) \mathbf{a}_i, \mathbf{e} \rangle| \leq c_0 \frac{\sqrt{\log N}}{\sqrt{N}} \|\mathbf{e}\|_2 \right\},$$

356 where I is the identity matrix.

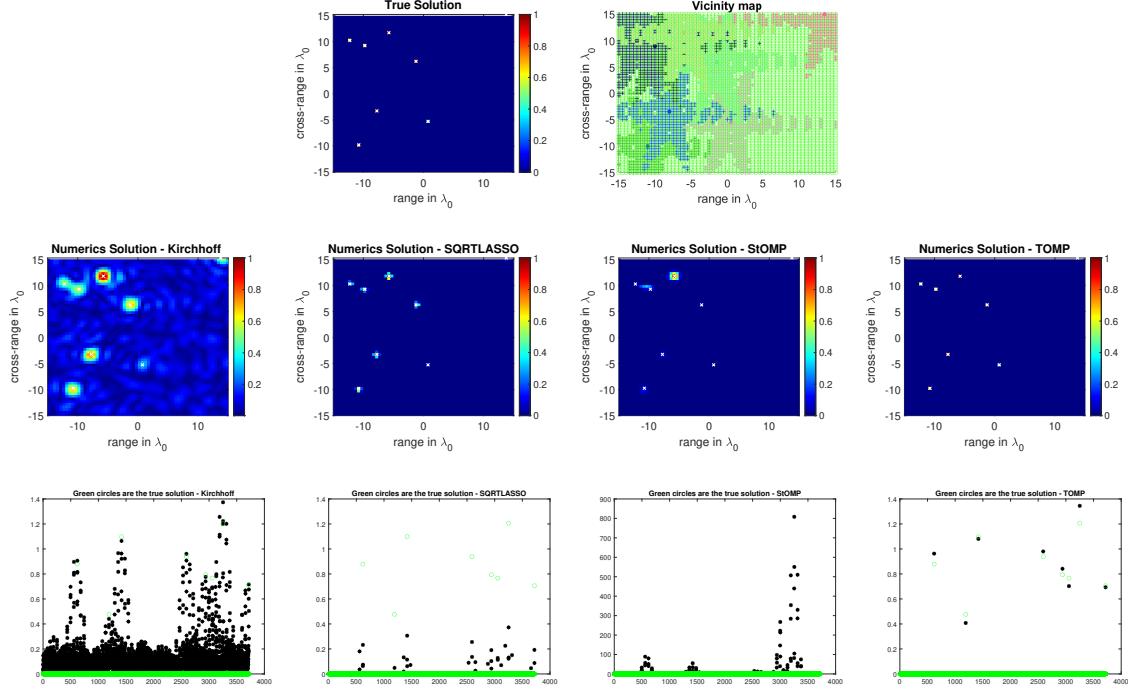


Figure 3. (The noise level is $\delta = 1$, so $SNR = 0dB$). In the top row, from left to right, the true image and the vicinity map. In the middle row, from left to right, the images recovered by Kirchhoff migration, SQRT-LASSO, StOMP, and TOMP. In the bottom row, from left to right, the solution vectors recovered by Kirchhoff migration, SQRT-LASSO, StOMP, and TOMP.

357 By independence, we have that $\mathbb{P}\left(|\langle \mathbf{a}_i, \mathbf{e} \rangle| \geq t\|\mathbf{e}\|_2/\sqrt{N}\right) \leq 2\exp(-t^2/2)$ for each \mathbf{a}_i .
 358 Here, we make use of the fact that uniformly distributed vectors in high dimension behave
 359 like Gaussian [25]. Consequently, the union bound gives

$$360 \quad \mathbb{P}\left(\max_{1 \leq i \leq K} |\langle \mathbf{a}_i, \mathbf{e} \rangle| \geq t\|\mathbf{e}\|_2/\sqrt{N}\right) \leq 2K\exp(-t^2/2) \leq 2N^\beta \exp(-t^2/2).$$

361 For $t = c_0\sqrt{\log N}$, the right-hand side becomes $2N^{\beta - c_0^2/2}$. Therefore,

$$362 \quad (7.3) \quad \mathbb{P}\left(\max_{1 \leq i \leq K} |\langle \mathbf{a}_i, \mathbf{e} \rangle| \leq c_0 \frac{\sqrt{\log N}}{\sqrt{N}} \|\mathbf{e}\|_2\right) \geq 1 - \frac{2}{N^{c_0^2/2 - \beta}}.$$

363 Choosing $c_0 = \sqrt{2(\beta + \kappa)}$, we deduce that event \mathcal{O}_1 holds with probability $1 - 2/N^\kappa$.
 364 Similarly, \mathcal{O}_2 occurs with probability $1 - 2/N^\kappa$. Therefore, the event $\mathcal{O} = \mathcal{O}_1 \cap \mathcal{O}_2$ occurs
 365 with probability $1 - 4/N^\kappa$. Suppose the event \mathcal{O} occurs, and the following analysis is
 366 deterministic on this event.

367 Our proof is to use induction on the number of iterations. Consider the first iteration
 368 ($n = 0$), then $\mathbf{b}^0 = \sum_{i \in \text{supp}(\boldsymbol{\rho})} \rho_i \mathbf{a}_i + \mathbf{e}$. First, we will prove (5.7) when $n = 0$. We have the

369 left-hand side of (5.7) can be estimated as

$$370 \quad |\langle \mathbf{a}_j, \mathbf{b}^0 \rangle| \leq |\rho_i| |\langle \mathbf{a}_j, \mathbf{a}_i \rangle| + \sum_{k \in \text{supp}(\boldsymbol{\rho}) \setminus \{i\}} |\rho_k| |\langle \mathbf{a}_j, \mathbf{a}_k \rangle| + |\langle \mathbf{a}_j, \mathbf{e} \rangle|,$$

371 and the right-hand side can be estimated as

$$372 \quad |\langle \mathbf{a}_i, \mathbf{b}^0 \rangle| \geq |\rho_i| - \sum_{k \in \text{supp}(\boldsymbol{\rho}) \setminus \{i\}} |\rho_k| |\langle \mathbf{a}_i, \mathbf{a}_k \rangle| - |\langle \mathbf{a}_i, \mathbf{e} \rangle|.$$

373 We want to show that

$$374 \quad (7.4) \quad \sum_{k \in \text{supp}(\boldsymbol{\rho}) \setminus \{i\}} |\rho_k| |\langle \mathbf{a}_j, \mathbf{a}_k \rangle| + \sum_{k \in \text{supp}(\boldsymbol{\rho}) \setminus \{i\}} |\rho_k| |\langle \mathbf{a}_i, \mathbf{a}_k \rangle| + |\langle \mathbf{a}_j, \mathbf{e} \rangle| + |\langle \mathbf{a}_i, \mathbf{e} \rangle| \leq |\rho_i| (1 - |\langle \mathbf{a}_j, \mathbf{a}_i \rangle|).$$

375 We have that

$$376 \quad |\rho_i| (1 - |\langle \mathbf{a}_j, \mathbf{a}_i \rangle|) \geq \rho_{\min} (1 - |\langle \mathbf{a}_j, \mathbf{a}_i \rangle|).$$

377 Using condition (5.2), we have that

$$378 \quad \sum_{k \in \text{supp}(\boldsymbol{\rho}) \setminus \{i\}} |\rho_k| |\langle \mathbf{a}_j, \mathbf{a}_k \rangle| \leq \rho_{\max} \frac{c_1 M}{\sqrt{N}} \quad \text{and} \quad \sum_{k \in \text{supp}(\boldsymbol{\rho}) \setminus \{i\}} |\rho_k| |\langle \mathbf{a}_i, \mathbf{a}_k \rangle| \leq \rho_{\max} \frac{c_1 M}{\sqrt{N}}.$$

379 We also have that

$$380 \quad |\langle \mathbf{a}_j, \mathbf{e} \rangle| \leq c_0 \|\mathbf{e}\|_2 \frac{\sqrt{\log N}}{\sqrt{N}} \quad \text{and} \quad |\langle \mathbf{a}_i, \mathbf{e} \rangle| \leq c_0 \|\mathbf{e}\|_2 \frac{\sqrt{\log N}}{\sqrt{N}}.$$

381 Therefore, inequality (5.7) will hold if the following condition holds:

$$382 \quad 2\rho_{\max} \frac{c_1 M}{\sqrt{N}} + 2c_0 \|\mathbf{e}\|_2 \frac{\sqrt{\log N}}{\sqrt{N}} < \rho_{\min} (1 - |\langle \mathbf{a}_j, \mathbf{a}_i \rangle|)$$

383 or

$$384 \quad |\langle \mathbf{a}_j, \mathbf{a}_i \rangle| < 1 - \gamma \frac{2c_1 M}{\sqrt{N}} - \frac{1}{\rho_{\min}} 2c_0 \|\mathbf{e}\|_2 \frac{\sqrt{\log N}}{\sqrt{N}}$$

385 which follows from condition (5.3).

386 Next, we will prove (5.8) when $n = 0$. We have the left-hand side of (5.8) can be
387 estimated as

$$388 \quad |\langle \mathbf{a}_j, \mathbf{b}^0 \rangle| < M \rho_{\max} \eta + c_0 \|\mathbf{e}\|_2 \frac{\sqrt{\log N}}{\sqrt{N}},$$

389 and the right-hand side can be estimated as

$$390 \quad |\langle \mathbf{a}_{i^*}, \mathbf{b}^0 \rangle| > \rho_{\max} - \frac{M \rho_{\max}}{4M} - c_0 \|\mathbf{e}\|_2 \frac{\sqrt{\log N}}{\sqrt{N}} = \frac{3 \rho_{\max}}{4} - c_0 \|\mathbf{e}\|_2 \frac{\sqrt{\log N}}{\sqrt{N}}.$$

391 We then have inequality (5.8) because we assume condition (5.4). We proceed to the
392 induction step. Suppose we are at the $(n+1)$ -th iteration ($n \geq 1$), and TOMP has already

393 recovered $\Omega := \Omega^n \subset \text{supp}(\boldsymbol{\rho})$. If $\Omega = \text{supp}(\boldsymbol{\rho})$, then $\mathbf{b}^n = (I - \mathcal{A}_{\text{supp}(\boldsymbol{\rho})}\mathcal{A}_{\text{supp}(\boldsymbol{\rho})}^\dagger)\mathbf{e}$. Since
 394 event \mathcal{O}_2 holds, we have that

$$395 \quad |\langle \mathbf{a}_j, \mathbf{b}^n \rangle| = |\langle (I - \mathcal{A}_{\text{supp}(\boldsymbol{\rho})}\mathcal{A}_{\text{supp}(\boldsymbol{\rho})}^\dagger)\mathbf{a}_j, \mathbf{e} \rangle| \leq c_0 \frac{\sqrt{\log N}}{\sqrt{N}} \|\mathbf{e}\|_2$$

396 for all $1 \leq j \leq K$. Therefore, TOMP will terminate at the Thresholding step. We
 397 consider the case $\Omega^c := \text{supp}(\boldsymbol{\rho}) \setminus \Omega \neq \emptyset$. Observe that by projecting \mathbf{b} onto the orthogonal
 398 complement of the vector space spanned by $\{\mathbf{a}_k : k \in \Omega\}$, we can write

$$399 \quad \mathbf{b}^n = \mathbf{b} - \mathcal{A}_\Omega \mathcal{A}_\Omega^\dagger \mathbf{b} = \rho_i \mathbf{a}_i + \mathcal{A}_{\Omega^c \setminus \{i\}} \boldsymbol{\rho}_{\Omega^c \setminus \{i\}} - \mathcal{A}_\Omega \mathcal{A}_\Omega^\dagger (\mathcal{A}_{\Omega^c} \boldsymbol{\rho}_{\Omega^c}) + (\mathbf{e} - \mathcal{A}_\Omega \mathcal{A}_\Omega^\dagger \mathbf{e}).$$

400 First, we proceed to prove (5.7) when $n \geq 1$. We have the following estimates

$$401 \quad |\langle \mathbf{a}_j, \mathbf{b}^n \rangle| \leq |\rho_i| |\langle \mathbf{a}_j, \mathbf{a}_i \rangle| + |\langle \mathbf{a}_j, \mathcal{A}_{\Omega^c \setminus \{i\}} \boldsymbol{\rho}_{\Omega^c \setminus \{i\}} \rangle| + |\langle \mathbf{a}_j, \mathcal{A}_\Omega \mathcal{A}_\Omega^\dagger (\mathcal{A}_{\Omega^c} \boldsymbol{\rho}_{\Omega^c}) \rangle| + |\langle \mathbf{a}_j, \mathbf{e} - \mathcal{A}_\Omega \mathcal{A}_\Omega^\dagger \mathbf{e} \rangle|,$$

402 and

$$403 \quad |\langle \mathbf{a}_i, \mathbf{b}^n \rangle| \geq |\rho_i| - |\langle \mathbf{a}_i, \mathcal{A}_{\Omega^c \setminus \{i\}} \boldsymbol{\rho}_{\Omega^c \setminus \{i\}} \rangle| - |\langle \mathbf{a}_i, \mathcal{A}_\Omega \mathcal{A}_\Omega^\dagger (\mathcal{A}_{\Omega^c} \boldsymbol{\rho}_{\Omega^c}) \rangle| - |\langle \mathbf{a}_i, \mathbf{e} - \mathcal{A}_\Omega \mathcal{A}_\Omega^\dagger \mathbf{e} \rangle|.$$

404 Using condition (5.2), we have that

$$405 \quad |\langle \mathbf{a}_j, \mathcal{A}_{\Omega^c \setminus \{i\}} \boldsymbol{\rho}_{\Omega^c \setminus \{i\}} \rangle| \leq \rho_{\max} \frac{c_1 M}{\sqrt{N}} \quad \text{and} \quad |\langle \mathbf{a}_i, \mathcal{A}_{\Omega^c \setminus \{i\}} \boldsymbol{\rho}_{\Omega^c \setminus \{i\}} \rangle| \leq \rho_{\max} \frac{c_1 M}{\sqrt{N}}.$$

406 On the other hand, we have that

$$407 \quad |\langle \mathbf{a}_j, \mathcal{A}_\Omega \mathcal{A}_\Omega^\dagger (\mathcal{A}_{\Omega^c} \boldsymbol{\rho}_{\Omega^c}) \rangle| = |\langle \mathcal{A}_\Omega^* \mathbf{a}_j, (\mathcal{A}_\Omega^* \mathcal{A}_\Omega)^{-1} \mathcal{A}_\Omega^* \mathcal{A}_{\Omega^c} \boldsymbol{\rho}_{\Omega^c} \rangle| \\ 408 \quad (7.5) \quad \leq \|\mathcal{A}_\Omega^* \mathbf{a}_j\|_2 \|(\mathcal{A}_\Omega^* \mathcal{A}_\Omega)^{-1}\| \|\mathcal{A}_\Omega^* \mathcal{A}_{\Omega^c} \boldsymbol{\rho}_{\Omega^c}\|_2,$$

410 where $\|\mathcal{A}\| := \max_{\|\boldsymbol{\rho}\|_2=1} \|\mathcal{A}\boldsymbol{\rho}\|_2$ is the operator norm of \mathcal{A} . We will estimate each term on
 411 the right-hand side of (7.5). The first term is as follows

$$412 \quad \|\mathcal{A}_\Omega^* \mathbf{a}_j\|_2 = \sqrt{\sum_{k \in \Omega} |\langle \mathbf{a}_k, \mathbf{a}_j \rangle|^2} \leq \sqrt{M \cdot \frac{c_1^2}{N}} = \frac{c_1 \sqrt{M}}{\sqrt{N}}.$$

413 The second term is estimated by using Gershgorin circle theorem (see [14]), which yields

$$414 \quad \|(\mathcal{A}_\Omega^* \mathcal{A}_\Omega)^{-1}\| \leq \frac{4}{3}.$$

415 For the third term, each entry of the vector $\mathcal{A}_\Omega^* \mathcal{A}_{\Omega^c} \boldsymbol{\rho}_{\Omega^c}$ is less than, in absolute value,

$$416 \quad \frac{c}{\sqrt{N}} (|\rho_1| + \dots + |\rho_{|\Omega^c|}|) \leq \rho_{\max} \frac{c_1 M}{\sqrt{N}}.$$

417 Therefore, we have that

$$418 \quad (7.6) \quad \|\mathcal{A}_\Omega^* \mathcal{A}_{\Omega^c} \boldsymbol{\rho}_{\Omega^c}\|_2 \leq \rho_{\max} \frac{c_1 M \sqrt{M}}{\sqrt{N}}.$$

419 Overall, we obtain the following bound

$$420 \quad |\langle \mathbf{a}_j, \mathcal{A}_\Omega \mathcal{A}_\Omega^\dagger (\mathcal{A}_{\Omega^c} \boldsymbol{\rho}_{\Omega^c}) \rangle| \leq \rho_{\max} \frac{4c_1^2 M^2}{3N}.$$

421 Similarly, we also have that

$$422 \quad |\langle \mathbf{a}_i, \mathcal{A}_\Omega \mathcal{A}_\Omega^\dagger (\mathcal{A}_{\Omega^c} \boldsymbol{\rho}_{\Omega^c}) \rangle| \leq \rho_{\max} \frac{4c_1^2 M^2}{3N}.$$

423 Finally, for the terms involving the noise \mathbf{e} , we have that

$$424 \quad |\langle \mathbf{a}_j, \mathbf{e} - \mathcal{A}_\Omega \mathcal{A}_\Omega^\dagger \mathbf{e} \rangle| \leq c_0 \|\mathbf{e}\|_2 \frac{\sqrt{\log N}}{\sqrt{N}} \quad \text{and} \quad |\langle \mathbf{a}_i, \mathbf{e} - \mathcal{A}_\Omega \mathcal{A}_\Omega^\dagger \mathbf{e} \rangle| \leq c_0 \|\mathbf{e}\|_2 \frac{\sqrt{\log N}}{\sqrt{N}}.$$

425 Overall, inequality (5.7) will hold if the following condition holds:

$$426 \quad 2\rho_{\max} \left(\frac{c_1 M}{\sqrt{N}} + \frac{4c_1^2 M^2}{3N} \right) + \frac{2c_0 \sqrt{\log N}}{\sqrt{N}} \|\mathbf{e}\|_2 < \rho_{\min} (1 - |\langle \mathbf{a}_j, \mathbf{a}_i \rangle|)$$

427 or

$$428 \quad |\langle \mathbf{a}_j, \mathbf{a}_i \rangle| < 1 - \gamma \left(\frac{2c_1 M}{\sqrt{N}} + \frac{8c_1^2 M^2}{3N} \right) - \frac{1}{\rho_{\min}} \frac{2c_0 \sqrt{\log N}}{\sqrt{N}} \|\mathbf{e}\|_2$$

429 which is precisely the condition (5.3).

430 Next, we will prove (5.8) when $n \geq 1$. We have that

$$431 \quad |\langle \mathbf{a}_j, \mathbf{b}^n \rangle| \leq M |\rho_{i^*}| \eta + |\langle \mathbf{a}_j, \mathcal{A}_\Omega \mathcal{A}_\Omega^\dagger (\mathcal{A}_{\Omega^c} \boldsymbol{\rho}_{\Omega^c}) \rangle| + c_0 \|\mathbf{e}\|_2 \frac{\sqrt{\log N}}{\sqrt{N}}.$$

432 and

$$433 \quad |\langle \mathbf{a}_{i^*}, \mathbf{b}^n \rangle| \geq \frac{3|\rho_{i^*}|}{4} - |\langle \mathbf{a}_{i_{\Omega^c}^*}, \mathcal{A}_\Omega \mathcal{A}_\Omega^\dagger (\mathcal{A}_{\Omega^c} \boldsymbol{\rho}_{\Omega^c}) \rangle| - c_0 \|\mathbf{e}\|_2 \frac{\sqrt{\log N}}{\sqrt{N}}.$$

434 Following the estimate (7.5) and noting that

$$435 \quad \|\mathcal{A}_\Omega^* \mathbf{a}_j\|_2 \leq \eta \sqrt{M} \quad \text{and} \quad \|\mathcal{A}_\Omega^* \mathcal{A}_{\Omega^c} \boldsymbol{\rho}_{\Omega^c}\|_2 \leq M \eta |\rho_i| \sqrt{M},$$

436 we have

$$437 \quad |\langle \mathbf{a}_j, \mathcal{A}_\Omega \mathcal{A}_\Omega^\dagger (\mathcal{A}_{\Omega^c} \boldsymbol{\rho}_{\Omega^c}) \rangle| \leq \eta \sqrt{M} \cdot \frac{4}{3} \cdot M \eta |\rho_i| \sqrt{M} = \frac{4\eta^2 M^2 |\rho_{i^*}|}{3}.$$

438 Similarly, we also have

$$439 \quad |\langle \mathbf{a}_{i_{\Omega^c}^*}, \mathcal{A}_\Omega \mathcal{A}_\Omega^\dagger (\mathcal{A}_{\Omega^c} \boldsymbol{\rho}_{\Omega^c}) \rangle| \leq \frac{|\rho_{i^*}|}{12}.$$

440 Overall, to verify inequality (5.8), we need to show that

$$441 \quad M |\rho_{i^*}| \eta + \frac{4\eta^2 M^2 |\rho_{i^*}|}{3} + c_0 \|\mathbf{e}\|_2 \frac{\sqrt{\log N}}{\sqrt{N}} < \frac{3|\rho_{i^*}|}{4} - \frac{|\rho_{i^*}|}{12} - c_0 \|\mathbf{e}\|_2 \frac{\sqrt{\log N}}{\sqrt{N}}$$

442 or equivalently,

$$443 \quad \|\mathbf{e}\|_2 \leq \frac{\sqrt{N} |\rho_{i^*}|}{2c_0 \sqrt{\log N}} \left(\frac{2}{3} - M \eta - \frac{4\eta^2 M^2}{3} \right).$$

444 This is implied from the condition (5.4). The proof is complete. ■

7.2. Proof of Theorem 5.2.

Proof. Similarly as in the proof of Proposition 5.3, we assume the event $\mathcal{O} = \mathcal{O}_1 \cap \mathcal{O}_2$ (defined in (7.1) and (7.2)) occurs, then the following analysis is deterministic on this event.

Assume $M = 0$. Since the event \mathcal{O}_1 holds, with $\tau = c_0\sqrt{\log N}/\sqrt{N}$, TOMP will terminate right after the first Thresholding step. Therefore, the output will be the empty set.

When $M \geq 1$, the objective is to demonstrate that at least one nonzero entry of ρ is detected at every iteration. It suffices to look at the nonzero entry of ρ with the largest magnitude. Consider the $(n + 1)$ -th iteration, and assume $\Omega := \Omega^n$ is strictly contained in the support of ρ (from [Proposition 5.3](#), we already have that $\Omega^n \subset \text{supp}(\rho)$ and when $\Omega^n = \text{supp}(\rho)$, TOMP terminates.) Let $\Omega^c = \text{supp}(\rho) \setminus \Omega$, the set of undetected indices. Without loss of generality, assume that the first entry ρ_1 is the nonzero entry with the largest magnitude among $\{\rho_k, k \in \Omega^c\}$. We want to show that

$$(7.7) \quad |\langle a_1, b^n \rangle| > \tau \|b^n\|_2.$$

Decompose \mathbf{b} into $\rho_1 \mathbf{a}_1 + \mathcal{A}_{\Omega^c \setminus \{1\}} \boldsymbol{\rho}_{\Omega^c \setminus \{1\}} + \mathcal{A}_\Omega \boldsymbol{\rho}_\Omega + \mathbf{e}$. Notice that, by projecting \mathbf{b} onto the orthogonal complement of the vector space spanned by $\{\mathbf{a}_k : k \in \Omega\}$, we have

$$\mathbf{b}^n = \mathbf{b} - \mathcal{A}_\Omega \mathcal{A}_\Omega^\dagger \mathbf{b} = \rho_1 \mathbf{a}_1 + \mathcal{A}_{\Omega^c \setminus \{1\}} \boldsymbol{\rho}_{\Omega^c \setminus \{1\}} - \mathcal{A}_\Omega \mathcal{A}_\Omega^\dagger (\mathcal{A}_{\Omega^c} \boldsymbol{\rho}_{\Omega^c}) + (\mathbf{e} - \mathcal{A}_\Omega \mathcal{A}_\Omega^\dagger \mathbf{e}).$$

For convenience, let us now set $\mathbf{v} := \mathcal{A}_{\Omega^c \setminus \{1\}} \rho_{\Omega^c \setminus \{1\}} - \mathcal{A}_\Omega \mathcal{A}_\Omega^\dagger (\mathcal{A}_{\Omega^c} \rho_{\Omega^c}) + (\mathbf{e} - \mathcal{A}_\Omega \mathcal{A}_\Omega^\dagger \mathbf{e})$. We then observe that

$$\begin{aligned} |\langle \mathbf{a}_1, \mathbf{b}^n \rangle|^2 - \tau \|\mathbf{b}^n\|_2^2 &= |\langle \mathbf{a}_1, \mathbf{a}_1\rho_1 + \mathbf{v} \rangle|^2 - \tau^2 \|\mathbf{a}_1\rho_1 + \mathbf{v}\|_2^2 \\ &= (1 - \tau^2) |\rho_1 + \langle \mathbf{a}_1, \mathbf{v} \rangle|^2 + \tau^2 |\langle \mathbf{a}_1, \mathbf{v} \rangle|^2 - \tau^2 \|\mathbf{v}\|_2^2 \\ &> (1 - \tau^2) |\rho_1 + \langle \mathbf{a}_1, \mathbf{v} \rangle|^2 - \tau^2 \|\mathbf{v}\|_2^2. \end{aligned}$$

By triangle inequality, we have $|\rho_1 + \langle \mathbf{a}_1, \mathbf{v} \rangle| \geq |\rho_1| - |\langle \mathbf{a}_1, \mathbf{v} \rangle|$. Therefore, it suffices to show that $|\rho_1| - |\langle \mathbf{a}_1, \mathbf{v} \rangle| > 0$ and

$$(7.8) \quad (1 - \tau^2) (|\rho_1| - |\langle \mathbf{a}_1, \mathbf{v} \rangle|)^2 > \tau^2 \|\mathbf{v}\|_2^2.$$

Let us estimate $|\langle \mathbf{a}_1, \mathbf{v} \rangle|$. We have that

$$|\langle \mathbf{a}_1, \mathbf{v} \rangle| \leq |\langle \mathbf{a}_1, \mathcal{A}_{\Omega^c \setminus \{1\}} \rho_{\Omega^c \setminus \{1\}} \rangle| + |\langle \mathbf{a}_1, \mathcal{A}_\Omega \mathcal{A}_\Omega^\dagger (\mathcal{A}_{\Omega^c} \rho_{\Omega^c}) \rangle| + |\langle \mathbf{a}_1, e - \mathcal{A}_\Omega \mathcal{A}_\Omega^\dagger e \rangle|.$$

As similarly done in the proof of Proposition 5.3, we have that

$$|\langle \mathbf{a}_1, \mathbf{v} \rangle| \leq \frac{|\rho_1|}{4} + \frac{|\rho_1|}{12} + \tau \|\mathbf{e}\|_2.$$

Consequently, we have that

$$|\rho_1| - |\langle \mathbf{a}_1, \mathbf{v} \rangle| \geq |\rho_1| - \frac{|\rho_1|}{4} - \frac{|\rho_1|}{12} - \tau \|\mathbf{e}\|_2 = \frac{2|\rho_1|}{3} - \tau \|\mathbf{e}\|_2 > 0.$$

477 Therefore, the left-hand side of (7.8) is bigger than $(1 - \tau^2)(2|\rho_1|/3 - \tau\|e\|_2)^2$. Now, we
 478 look at the right-hand side. We have that

$$479 \quad \|v\|_2 \leq \|\mathcal{A}_{\Omega^c \setminus \{1\}} \rho_{\Omega^c \setminus \{1\}}\|_2 + \|\mathcal{A}_\Omega \mathcal{A}_\Omega^\dagger (\mathcal{A}_{\Omega^c} \rho_{\Omega^c})\|_2 + \|e - \mathcal{A}_\Omega \mathcal{A}_\Omega^\dagger e\|_2.$$

480 We estimate each term as follows

- 481 • For $\|\mathcal{A}_{\Omega^c \setminus \{1\}} \rho_{\Omega^c \setminus \{1\}}\|_2$:

482 We have that

$$483 \quad \|\mathcal{A}_{\Omega^c \setminus \{1\}} \rho_{\Omega^c \setminus \{1\}}\|_2^2 = \sum_{i \in \Omega^c \setminus \{1\}} |\rho_i|^2 + 2 \sum_{i < j, i, j \in \Omega^c \setminus \{1\}} \operatorname{Re}(\langle \mathbf{a}_i, \mathbf{a}_j \rangle \bar{\rho}_i \rho_j)$$

$$484 \quad \leq (M-1)|\rho_1|^2 + 2 \cdot \frac{1}{4M} \cdot \frac{(M-1)(M-2)}{2} |\rho_1|^2$$

$$485 \quad = \left(\frac{5}{4}M - \frac{7}{4} + \frac{1}{2M} \right) |\rho_1|^2$$

486 where $\operatorname{Re}(x)$ is the real part of a complex number x . Thus

$$488 \quad \|\mathcal{A}_{\Omega^c \setminus \{1\}} \rho_{\Omega^c \setminus \{1\}}\|_2 \leq \sqrt{\frac{5}{4}M - \frac{7}{4} + \frac{1}{2M}} |\rho_1|.$$

- 489 • For $\|\mathcal{A}_\Omega \mathcal{A}_\Omega^\dagger (\mathcal{A}_{\Omega^c} \rho_{\Omega^c})\|_2$:

490 We have that

$$\begin{aligned} 491 \quad \|\mathcal{A}_\Omega \mathcal{A}_\Omega^\dagger (\mathcal{A}_{\Omega^c} \rho_{\Omega^c})\|_2^2 &= |\langle \mathcal{A}_\Omega^* \mathcal{A}_{\Omega^c} \rho_{\Omega^c}, (\mathcal{A}_\Omega^* \mathcal{A}_\Omega)^{-1} \mathcal{A}_\Omega^* \mathcal{A}_{\Omega^c} \rho_{\Omega^c} \rangle| \\ &\leq \|\mathcal{A}_\Omega^* \mathcal{A}_{\Omega^c} \rho_{\Omega^c}\|_2 \|(\mathcal{A}_\Omega^* \mathcal{A}_\Omega)^{-1}\| \|\mathcal{A}_\Omega^* \mathcal{A}_{\Omega^c} \rho_{\Omega^c}\|_2 \\ &\leq \frac{\sqrt{M}|\rho_1|}{4} \cdot \frac{4}{3} \cdot \frac{\sqrt{M}|\rho_1|}{4} \\ &\leq \frac{M|\rho_1|^2}{12}. \end{aligned}$$

- 492 • For $\|e - \mathcal{A}_\Omega \mathcal{A}_\Omega^\dagger e\|_2$:

493 Since $e - \mathcal{A}_\Omega \mathcal{A}_\Omega^\dagger e$ is an orthogonal projection of e , we have that

$$494 \quad \|e - \mathcal{A}_\Omega \mathcal{A}_\Omega^\dagger e\|_2 \leq \|e\|_2.$$

495 Overall, in order to show (7.8), it suffices to show that

$$496 \quad (1 - \tau^2) \left(\frac{2|\rho_1|}{3} - \tau\|e\|_2 \right)^2 > \tau^2 \left[\left(\sqrt{\frac{5}{4}M - \frac{7}{4} + \frac{1}{2M}} + \frac{1}{\sqrt{12}}\sqrt{M} \right) |\rho_1| + \|e\|_2 \right]^2,$$

497 or equivalently,

$$498 \quad (7.9) \quad \underbrace{\left[\frac{2}{3} \sqrt{\frac{1 - \tau^2}{\tau^2}} - \left(\sqrt{\frac{5}{4}M - \frac{7}{4} + \frac{1}{2M}} + \frac{1}{\sqrt{12}}\sqrt{M} \right) \right]}_F |\rho_1| > \underbrace{\left(\sqrt{1 - \tau^2} + 1 \right)}_G \|e\|_2.$$

499 We choose

500 (7.10)
$$f(M, \tau) = \frac{F}{G}.$$

501 We note that $F > 0$ for $M \geq 1$ using condition (5.6). We then have if $\|e\|_2 \leq f(M, \tau)\rho_{\min}$,
 502 then the inequality (7.9) holds. Consequently, $|\rho_1| - |\langle a_1, v \rangle| > 0$, and then the inequality
 503 (7.8) holds, and thus (7.7) holds. The proof is complete. ■

504 **8. Conclusions.** In this paper, we analyze Thresholding Orthogonal Matching Pursuit
 505 (TOMP), a modified version of OMP with a new stopping criteria. The stopping criteria
 506 is based on analyzing inner products of vectors in high-dimensional spaces. We demon-
 507 strate that TOMP is effective for recovering imaging scences even when the measurement
 508 matrices are highly coherent and the data are noisy. The idea is to use the concept of
 509 vicinity/coherence band to analyze nearby points whose corresponding signal vectors are
 510 coherent. We remark that TOMP works well under even less stringent conditions than
 511 what are assumed in [Theorem 5.2](#) as shown by numerics. Possible future research direc-
 512 tions are to improve [Theorem 5.2](#) by assuming weaker conditions and modify TOMP so
 513 multiple locations may be recovered at each iteration.

514 **Acknowledgments.** We are grateful to Anna Gilbert for suggesting to apply the or-
 515 thogonal projection update of CoSaMP to develop our algorithm. We thank Stephen White
 516 for suggesting doing binary search for the calibration of τ .

517

REFERENCES

- 518 [1] A. BECK AND M. TEBOULLE, *A fast iterative shrinkage-thresholding algorithm with application to*
 519 *wavelet-based image deblurring*, IEEE International Conference on Acoustics, Speech and Signal
 520 Processing, (2009), pp. 693–696.
- 521 [2] A. BELLONI, V. CHERNOZHUKOV, AND L. WANG, *Square-root lasso: pivotal recovery of sparse signals*
 522 *via conic programming*, Biometrika, 98 (2011), pp. 791–806.
- 523 [3] E. J. CANDES, J. K. ROMBERG, AND T. TAO, *Robust uncertainty principles: Exact signal recon-*
 524 *struction from highly incomplete frequency information*, IEEE Trans. Inform. Theory, 52 (2006),
 525 pp. 489–509.
- 526 [4] E. J. CANDES, J. K. ROMBERG, AND T. TAO, *Stable signal recovery from incomplete and inaccurate*
 527 *measurements*, Comm. Pure Appl. Math., 59 (2006), pp. 1207–1223.
- 528 [5] S. CHEN AND D. DONOHO, *Basis pursuit*, Proceedings of Asilomar Conference on Signals, Systems
 529 and Computers, 1 (1994), pp. 41–44.
- 530 [6] S. S. CHEN, D. L. DONOHO, AND M. A. SAUNDERS, *Atomic decomposition by basis pursuit*, SIAM
 531 Review, 43 (2001), pp. 129–159.
- 532 [7] M. CHICHIGNODU, J. LEDERER, M. J. WAINWRIGHT, AND F. BACH, *A practical scheme and fast*
 533 *algorithm to tune the lasso with optimality guarantees*, J. Mach. Learn. Res., 17 (2016), pp. 1–17.
- 534 [8] Y. T. DAVID DONOHO, I. DRORI, AND J. L. STARCK, *Sparse solution of underdetermined systems of*
 535 *linear equations by stagewise orthogonal matching pursuit*, IEEE Trans. Inform. Theory, 58 (2012),
 536 pp. 1094–1121.
- 537 [9] D. DONOHO AND M. ELAD, *Optimally sparse representation in general (nonorthogonal) dictionaries*
 538 *via l_1 minimization*, PNAS, 100 (2003), pp. 2197–2202.
- 539 [10] M. ELAD AND A. M. BRUCKSTEIN, *A generalized uncertainty principle and sparse representation in*
 540 *pairs of bases*, IEEE Trans. Inform. Theory, 48 (2002), pp. 2558–2567.
- 541 [11] A. FANNJIANG AND W. LIAO, *Coherence pattern-guided compressive sensing with unresolved grids*,
 542 SIAM J. Imaging Sci., 5 (2012), pp. 179–202.

- 543 [12] A. FEUER AND A. NEMIROVSKI, *On sparse representation in pairs of bases*, IEEE Trans. Inform. Theory, 49 (2003), pp. 1579–1581.
- 544 [13] J. J. FUCHS, *Recovery of exact sparse representations in the presence of bounded noise*, IEEE Trans. Inform. Theory, 51 (2005), pp. 3601–3608.
- 545 [14] G. GOLUB AND C. V. LOAN, *Matrix Computations*, Johns Hopkins Univ. Press, 2012.
- 546 [15] X. LI, H. JIANG, J. HAUPt, R. ARORA, H. LIU, M. HONG, AND T. ZHAO, *On fast convergence of proximal algorithms for sqrt-lasso optimization: Don't worry about its nonsmooth loss function.*, Annual Conference on Uncertainty in Artificial Intelligence (UAI), (2019).
- 547 [16] A. MALEKI, *Coherence analysis of iterative thresholding algorithms*, 47th Annual Allerton Conference on Communication, Control, and Computing (Allerton), (2009), pp. 236–243.
- 548 [17] E. C. MARQUES, N. MACIEL, L. NAVINER, H. CAI, AND J. YANG, *A review of sparse recovery* algorithms, IEEE Access, 7 (2019), pp. 1300–1322.
- 549 [18] R. MODELING WITH ERRATIC DATA, *Jon f claerbout and francis muir*, IEEE Trans. Inform. Theory, 52 (2006), pp. 489–509.
- 550 [19] M. MOSCOSO, A. NOVIKOV, G. PAPANICOLAOU, AND C. TSOGKA, *Imaging with highly incomplete* and corrupted data, Inverse Problems, 36 (2020), p. 035010.
- 551 [20] M. MOSCOSO, A. NOVIKOV, G. PAPANICOLAOU, AND C. TSOGKA, *The noise collector for sparse* recovery in high dimensions, PNAS, 117 (2020), pp. 11226–11232.
- 552 [21] D. NEEDELL AND J. A. TROPP, *Cosamp: Iterative signal recovery from incomplete and inaccurate* samples, Appl. Comput. Harmon. Anal., 26 (2009), pp. 301–321.
- 553 [22] R. TIBSHIRANI, *Regression shrinkage and selection via the lasso*, J. R. Stat. Soc. Ser. B. Stat. Methodol., 58 (1996), pp. 267–288.
- 554 [23] J. A. TROPP, *Just relax: Convex programming methods for identifying sparse signals in noise*, IEEE Trans. Inform. Theory, 52 (2006), pp. 1030–1051.
- 555 [24] J. A. TROPP AND A. C. GILBERT, *Signal recovery from random measurements via orthogonal matching* pursuit, IEEE Trans. Inform. Theory, 53 (2007), pp. 4655–4666.
- 556 [25] R. VERSHYNIN, *High-Dimensional Probability*, Cambridge Univ. Press, 2008.
- 557 [26] M. J. WAINWRIGHT, *Sharp thresholds for high-dimensional and noisy sparsity recovery using l_1 -constrained quadratic programming (lasso)*, IEEE Trans. Inform. Theory, 55 (2009), pp. 2183–2202.
- 558 [27] H. ZOU, *The adaptive lasso and its oracle properties*, J. Amer. Statist. Assoc., 101 (2006), pp. 1418–1429.