Homomorphic Sensing: Sparsity and Noise

Liangzu Peng [1] **Boshi Wang** [1] **Manolis C. Tsakiris** [1]

Abstract

Unlabeled sensing is a recent problem encompassing many data science and engineering applications and typically formulated as solving linear equations whose right-hand side vector has undergone an unknown permutation. It was generalized to the *homomorphic sensing* problem by replacing the unknown permutation with an unknown linear map from a given finite set of linear maps. In this paper we present tighter and simpler conditions for the homomorphic sensing problem to admit a unique solution. We show that this solution is locally stable under noise, while under a sparsity assumption it remains unique under less demanding conditions. Sparsity in the context of unlabeled sensing leads to the problem of *unlabeled compressed sensing*, and a consequence of our general theory is the existence under mild conditions of a unique sparsest solution. On the algorithmic level, we solve unlabeled compressed sensing by an iterative algorithm validated by synthetic data experiments. Finally, under the unifying homomorphic sensing framework we connect unlabeled sensing to other important practical problems.

1. Introduction

1.1. Compressed Sensing

The beginning of the 21st century has witnessed the birth of *compressed sensing*, a subject, as written by Theodoridis (2020), whose starting point is *to develop conditions for the solution of an underdetermined linear system of equations*. In an attempt at finding a sparsest solution of the linear equations $y = Ax$ with $A \in \mathbb{R}^{m\times n}$, researchers have focused on the optimization problem

$$\min_{x\in\mathbb{R}^n} \|x\|_0 \quad \text{s.t.} \quad y = Ax. \tag{1}$$

Assuming the existence of a k-sparse solution x^* to (1), the first question is whether x^* is unique. The answer is typically characterized via two frequently used notions, the *spark* (Donoho & Elad, 2003) or the *Kruskal rank* (Kruskal, 1977), and has played a major role in the theoretical foundations of compressed sensing, where x^* is typically viewed as a sparse signal and A as a measurement matrix. Specifically, for a *generic*[1] $A \in \mathbb{R}^{m\times n}$, x^* is the unique sparsest solution to (1) if $m \geq 2k$. Conversely, for any $A \in \mathbb{R}^{m\times n}$, there are multiple sparsest solutions to (1) whenever $m < 2k < n$. Later efforts considered the equivalence between the ℓ_0 semi-norm minimization (1) and its ℓ_1 relaxation known as *basis pursuit*, typically studied via the *nullspace property* (Cohen et al., 2009) or the *restricted isometry property* (Candes & Tao, 2005). Built upon those theoretical grounds are efficient convex optimization algorithms that solve the ℓ_1 basis pursuit problem. For a modern account of compressed sensing the reader is referred to the relevant chapters in Theodoridis (2020); Wright & Ma (2020).

1.2. Unlabeled Sensing

Recently, increasing research efforts have concentrated on the *unlabeled sensing* problem, proposed by Unnikrishnan et al. (2015); Unnikrishnan et al. (2018) in signal processing contexts. With $\mathcal{S}_m$ the set of $m \times m$ permutation matrices, $\Pi^* \in \mathcal{S}_m$ an *unknown* permutation, and $y = \Pi^* Ax^*$, unlabeled sensing is concerned with solving the equations[2]

$$y = \Pi Ax, \quad \Pi \in \mathcal{S}_m, \quad x \in \mathbb{R}^n \tag{2}$$

in the unknown x from the given data y, A. The connection of unlabeled sensing with compressed sensing is that both problems can be cast as subspace classification problems. In both cases the given data can be thought of as a union of subspaces $\bigcup_{i\in[\ell]} \mathcal{V}_i \subset \mathbb{R}^m$ together with a point y in that union, and the problem is to determine which subspace $\mathcal{V}_i$ the point y belongs to. In compressed sensing, the $\mathcal{V}_i$'s arise as the images of the $\ell = \binom{n}{k}$ k-dimensional coordinate subspaces of $\mathbb{R}^n$ under the measurement matrix A; in unlabeled sensing there are $\ell = m!$ subspaces $\mathcal{R}(\Pi A)$ obtained as Π

[1]School of Information Science and Technology, ShanghaiTech University, Shanghai, China. Correspondence to: Liangzu Peng, Boshi Wang, Manolis C. Tsakiris <penglz,wangbsh,mtsakiris@shanghaitech.edu.cn>.

Proceedings of the 38^{th} International Conference on Machine Learning, PMLR 139, 2021.

[1]For now it is safe to think of "generic" as "random", and "for a generic $A \in \mathbb{R}^{m\times n}$" as "for almost all $A \in \mathbb{R}^{m\times n}$" (see §2.1).

[2]In fact Unnikrishnan et al. (2018) considered the even more challenging problem where only a subset of the entries of $y = \Pi^* Ax^*$ is given; see also §4.

ranges in $\mathcal{S}_m$, where $\mathcal{R}(\Pi A)$ is the column space of ΠA.

Applications. A prominent data analysis application of unlabeled sensing is *record linkage* (Fellegi & Sunter, 1969; Domingo-Ferrer & Muralidhar, 2016; Muralidhar, 2017). This relates to linking records collected from different sources, a routine operation of government agencies (e.g., see Antoni & Schnell (2019)), for the purpose of subsequent data analysis. Due to privacy concerns, it is customary for each entry of the records corresponding to some individual to not be associated with a unique identifier of this individual (e.g., the social security number). As a result, domain-specific algorithms for linking the respective entries in two (or more) records corresponding to the same individual can be error-prone, yielding imperfect data for later analysis. An alternative is to ask whether one can fit a linear regression model between two unlinked records, say $y \in \mathbb{R}^m$ and $A \in \mathbb{R}^{m\times n}$, where y records only one feature while A records n features of the same m individuals. Without linkage, the correspondences between entries of y and rows of A are unknown. Such data imperfection is naturally modeled by an unknown permutation $\Pi^* \in \mathcal{S}_m$, and this gives $y = \Pi^* A x^*$ with $x^* \in \mathbb{R}^n$ the unknown linear regression parameters. The aim is to recover x^* from y, A and this is exactly problem (2). Besides record linkage, other applications abound: signal estimation using distributed sensors (Zhu et al., 2017; Song et al., 2018; Peng et al., 2019), target localization in signal processing (Wang et al., 2020a), neuron matching in computational neuroscience (Nejatbakhsh & Varol, 2021), automated translation of medical codes (Shi et al., 2020) and flow cytometry (Abid et al., 2017; Abid & Zou, 2018) in biology, multi-target tracking (Ji et al., 2019; Xie et al., 2021) and point set registration (Pananjady et al., 2017; Lian et al., 2017) in computer vision; see Pananjady et al. (2018); Xie et al. (2021) for more applications.

Theory. Unnikrishnan et al. (2015) proved the fundamental fact that, if A is generic, then x^* is the unique solution to (2) if and only if $m \geq 2n$, a result more recently and independently also obtained by Han et al. (2018) using different techniques. While this result holds for any $x^* \in \mathbb{R}^n$, Tsakiris et al. (2020) showed that $m \geq n+1$ samples are sufficient for uniqueness, if we additionally assume that $x^* \in \mathbb{R}^m$ is generic. Once the uniqueness for (2) was settled, noisy measurements were considered. With $\overline{y} := y + \epsilon = \Pi^* A x^* + \epsilon$ for some noise ϵ, Pananjady et al. (2018) showed that the estimator

$$(\hat{x}, \hat{\Pi}) \in \underset{x\in\mathbb{R}^n,\, \Pi\in\mathcal{S}_m}{\operatorname{argmin}} \left\|\overline{y} - \Pi A x\right\|_2 \tag{3}$$

is NP-hard to compute if $n > 1$. Moreover, assuming A has i.i.d. standard Gaussian entries and ϵ has Gaussian distribution $\mathcal{N}(0, \sigma^2 I_m)$, they asserted that $\Pi^* = \hat{\Pi}$ with high probability as long as the SNR:=$\left\|x^*\right\|_2^2/\sigma^2$ is exponentially high (e.g., SNR$\geq m^c$ for some constant $c > 0$). Of similar flavor is a result of Hsu et al. (2017), where it was shown under the same setting of Pananjady et al. (2018) that x^* can not be approximately recovered, unless the SNR is larger than $c' \min\{1, n/\log\log m\}$ for some constant $c' > 0$. Later on, under the above assumptions on A and ϵ, Slawski & Ben-David (2019) showed that, if Π^* is p-sparse in the sense that Π^* permutes at most p rows of A, then the estimator

$$\underset{x\in\mathbb{R}^n}{\operatorname{argmin}} \left\|\overline{y} - Ax\right\|_1 \tag{4}$$

behaves well with high probability, in the sense that its distance to x^* is an increasing linear function of σ.

Algorithms. Under the above sparse assumption on Π^*, the relaxation (3) of Slawski & Ben-David (2019) can be solved via convex optimization. Empirically, this yields a good estimate of x^* as long as no more than half of the data are shuffled, i.e., $p/m < 0.5$. This was improved by Slawski et al. (2019); Slawski et al. (2021), who synthesized hypothesis testing, expectation maximization, and recursively reweighed least-squares into an efficient algorithm which can handle up to $p/m = 0.7$ shuffled data, with a drawback of being sensitive to the distribution of A. Tsakiris & Peng (2019), Tsakiris et al. (2020), and Peng & Tsakiris (2020) followed a very different route towards solving (3), with the aim of tackling the fully shuffled case $p/m = 1$. The two algorithms of Tsakiris & Peng (2019) are based on branch-and-bound and RANSAC respectively, and have good performance for $n \leq 4$, while intractable for $n \geq 5$. The algebraic approach of Tsakiris et al. (2020) is based on solving a system of n polynomial equations in n variables, and selects the most suitable among the finite set of roots to initialize (3) towards alternating minimization. This gives an algorithm of linear complexity in m, efficient for $n \leq 5$ and intractable otherwise. The algorithm of Peng & Tsakiris (2020) is based on a concave minimization reformulation of (3) and is solved via branch-and-bound. It can handle dimensions $n \leq 8$, which to the best of our knowledge is the largest value up to date for the successful operation of an unlabeled sensing algorithm on fully shuffled data. Let us also note that, when $m < n$, both (3) and (4) are bound to have infinitely many solutions and thus all of the above algorithms in principle break down. Finally, recent algorithms that handle other types of unlabeled data include Slawski et al. (2019); Wang et al. (2020b); Zhang et al. (2019); Slawski et al. (2020); Zhang & Li (2020); Abbasi et al. (2020); Marano & Willett (2020); Jeong et al. (2020); Jeong et al. (2021); Abbasi et al. (2021); Yao et al. (2021).

1.3. Homomorphic Sensing

Inspired by Unnikrishnan et al. (2018) a generalization of unlabeled sensing was posed by Tsakiris (2018; 2020); Tsakiris & Peng (2019) under the name *homomorphic sensing*. This generalization replaces the set $\mathcal{S}_m$ of $m \times m$

permutations with an *arbitrary* finite set $\mathcal{T}$ of $r \times m$ matrices, $r \leq m$. That is

$$y = TAx, \quad T \in \mathcal{T}, \quad x \in \mathbb{R}^n, \tag{5}$$

where we are now given the measurements as $y = T^*Ax^*$ for some unknown $T^* \in \mathcal{T}$ and the goal is to solve (5) for x. Tsakiris (2018; 2020) proved that if A is generic, then x^* is the unique solution to (5), providing that the dimension of a certain algebraic variety that depends on $\mathcal{T}$ is at most $m-n$. Tsakiris (2018; 2020); Tsakiris & Peng (2019) applied their results to unlabeled sensing (e.g., by setting $\mathcal{T}$ to be $\mathcal{S}_m$) and obtained the same conditions of Unnikrishnan et al. (2018) which guarantee the uniqueness of the solution to (2).

1.4. Contributions of this paper

We improve, generalize prior works in several ways; complete proofs of our statements are in Peng & Tsakiris (2021).

Sparse homomorphic sensing. We bring homomorphic sensing and compressed sensing together, and arrive at the problem of *sparse homomorphic sensing*. Recalling $y = T^*Ax^*$, we consider the following ℓ_0 optimization problem:

$$\min_{x \in \mathbb{R}^n} \|x\|_0 \quad \text{s.t.} \quad y = TAx, \;\; T \in \mathcal{T} \tag{6}$$

Assuming that x^* is a k-sparse solution to (6), we provide conditions under which x^* is the unique such solution (Theorem 1). When $k = n$, our conditions in particular guarantee the uniqueness of the solution to homomorphic sensing (5), and they are tighter and simpler than those of Tsakiris (2018; 2020); Tsakiris & Peng (2019) (Corollary 1).

Noisy homomorphic sensing. We also extend homomorphic sensing (5) to *noisy* homomorphic sensing, where we are given the noisy measurements $\overline{y} = y + \epsilon = T^*Ax^* + \epsilon$. We show in Theorem 2 that, as long as $\|\epsilon\|_2$ is *sufficiently small* and (5) admits a unique solution, the following problem (7) produces an estimate $\hat{x}$ close to x^*:

$$(\hat{x}, \hat{T}) \in \operatorname*{argmin}_{x \in \mathbb{R}^n, T \in \mathcal{T}} \|\overline{y} - TAx\|_2 \tag{7}$$

When setting $\mathcal{T}$ to $\mathcal{S}_m$ (3), we obtain an improved result for unlabeled sensing over that of Unnikrishnan et al. (2018).

Unlabeled compressed sensing. We propose *unlabeled compressed sensing*, where we let $y = \Pi^*Ax^*$ with x^* a k-sparse solution to the following optimization problem:

$$\min_{x \in \mathbb{R}^n} \|x\|_0 \quad \text{s.t.} \quad y = \Pi Ax, \;\; \Pi \in \mathcal{S}_m \tag{8}$$

Clearly, (8) is a special case of (6), where $\mathcal{T}$ is set to $\mathcal{S}_m$. So our theorem for sparse homomorphic sensing can be applied to unlabeled compressed sensing, and in so doing, we get:

Proposition 1. *For a generic $A \in \mathbb{R}^{m \times n}$, x^* is the unique sparsest solution to (8) as long as $m \geq 2k$.*

Proposition 1 seems interesting. Indeed, the number $2k$ is the threshold for unique recovery of x^* in compressed sensing (recall §1.1), but this number remains the same in unlabeled compressed sensing, even though there could be $m!$ choices for the potential permutations. In particular, Proposition 1 holds true even when $m \ll n$.

Computationally, we consider a relaxation of (8), which we solve via an iterative algorithm based on subgradient descend and ℓ_1 minimization (§3.2). This is the first algorithm for unlabeled sensing which works even when $m < n$, a regime unexplored in prior works. By experiments (§3.3), we empirically show that i) the algorithm returns a correct estimate as long as x^* and Π^* are both *sufficiently sparse* (i.e., p, k are small), ii) it is efficient, iii) it is robust to noise.

A broader picture. As it turns out, we find that, besides unlabeled sensing, homomorphic sensing contains as special cases other important inverse problems, such as *real phase retrieval*, *mixed linear regression*, *missing data recovery*, to name a few. This allows our theory to be further applied to those special cases, and also allows potential connections between these problems themselves via the unifying framework of homomorphic sensing. We discuss this in §4.

2. Homomorphic sensing theory

In §2.1 we give some preliminaries on algebraic geometry. In §2.2 we formally state the problem and discuss immediate observations. In §2.3 we describe conditions for the uniqueness of the solution to (6). In §2.4 and §2.5, we introduce Theorems 1 and 2 for sparse and noisy homomorphic sensing, respectively. Along the way we give intuition on the proofs when space allows and we refer the reader to Peng & Tsakiris (2021) for the complete arguments.

2.1. Preliminaries on algebraic geometry

Here we review some background from algebraic geometry. A very accessible introduction to this subject is Cox et al. (2013) and a friendly one is Eisenbud (2013).

The basic object in algebraic geometry is a complex (*resp.* real) *algebraic variety*, say $\mathcal{Q}$, which is typically defined as the set of complex (*resp.* real) roots of finitely many polynomials $p_1, \ldots, p_s$ in m variables with complex (*resp.* real) coefficients; in other words, $\mathcal{Q} := \{z : p_i(z) = 0, i = 1, \ldots, s\}$. A subvariety of some algebraic variety $\mathcal{Q}$ is a subset of $\mathcal{Q}$ that is itself an algebraic variety. For example, a linear subspace of $\mathbb{R}^m$ is an algebraic variety defined by linear forms, and so is the union of finitely many linear subspaces, $\cup_{i=1}^s \mathcal{V}_i$, which is defined by the products of the defining equations of its constituent subspaces $\mathcal{V}_i$'s.

Declaring the subvarieties of $\mathbb{R}^{m\times n}$ as *closed* sets, we obtain the so-called *Zariski* topology, and, as usual, the complement of a closed set is called (Zariski) *open*. By a *generic* matrix of $\mathbb{R}^{m\times n}$ having some property, we mean that there is some non-empty Zariski open subset $\mathscr{O}$ of $\mathbb{R}^{m\times n}$ such that every matrix of $\mathscr{O}$ satisfies this property. A non-empty Zariski open subset is *dense*, in view of the fact that $\mathbb{R}^{m\times n}$ is *irreducible* (an irreducible algebraic variety is one which can not be written as the union of two proper subvarieties of it). The consequence of such a non-empty Zariski open set $\mathscr{O}$ being dense is that a matrix randomly chosen from $\mathbb{R}^{m\times n}$ according to some continuous probability distribution will land itself in $\mathscr{O}$, with probability one.

The final algebraic-geometric notion which we need is that of *dimension*, and for technical reasons, we consider this notion over the complex field, $\mathbb{C}$. Intuitively, the dimension of a given subset of $\mathbb{C}^m$ is a non-negative number that measures the *size* of the set. Formally, the dimension $\dim(\mathcal{Q})$ of an algebraic variety $\mathcal{Q}$ is the maximal length t of the chains $\mathcal{Q}_0 \subset \mathcal{Q}_1 \subset \cdots \subset \mathcal{Q}_t$ of distinct irreducible algebraic varieties contained in $\mathcal{Q}$; and the dimension of any set is the dimension of its *closure*, i.e., the smallest algebraic variety which contains it. For example, it is not hard to show by definition that a linear subspace has its linear-algebraic dimension equal to its algebraic-geometric dimension, and that the dimension of a finite union of linear subspaces, $\dim(\cup_{i=1}^{s}\mathcal{V}_i)$, is equal to the maximum dimension of its constituent subspaces, $\max_{i=1,\ldots,s}\dim(\mathcal{V}_i)$.

2.2. The uniqueness at first glance

The uniqueness for (6) involves the measurements $y = T^*Ax^*$, where x^* is k-sparse and can otherwise be arbitrary, and we seek conditions that guarantee that it is the only k-sparse solution to (6). This motivates Definition 1.

Definition 1 (hsp)**.** *Let $A \in \mathbb{R}^{m\times n}$, $\mathcal{T} \subset \mathbb{R}^{r\times m}$. If for any $T_1, T_2 \in \mathcal{T}$ and any k-sparse $x_1, x_2 \in \mathbb{R}^n$ we have*

$$T_1Ax_1 = T_2Ax_2 \;\;\Rightarrow\;\; x_1 = x_2,$$

then we say that $\mathcal{T}$ and A satisfy the homomorphic sensing property for k-sparse vectors, denoted as $\mathrm{hsp}(\mathcal{T}, A, k)$.

While $\mathrm{hsp}(\mathcal{T}, A) := \mathrm{hsp}(\mathcal{T}, A, n)$ is a clearly special case of $\mathrm{hsp}(\mathcal{T}, A, k)$ (with k set to be n), Definition 1 suggests that $\mathrm{hsp}(\mathcal{T}, A)$ implies $\mathrm{hsp}(\mathcal{T}, A, k)$ for any $k \le n$. Moreover, $\mathrm{hsp}(\mathcal{T}, A)$ is the same as the uniqueness of the solution to the homomorphic sensing problem (5), which, in turn, is equivalent to that x^* is the only feasible point of (6) and necessarily the unique solution to (6). More generally, Definition 1 immediately implies the following equivalence.

Proposition 2. *We have* $\mathrm{hsp}(\mathcal{T}, A, k)$ *if and only if Problem* (6) *always has a unique k-sparse solution.*

Next, our main focus will be on $\mathrm{hsp}(\mathcal{T}, A, k)$. Specifically, we will discuss the conditions to put on $\mathcal{T}$ under which $\mathrm{hsp}(\mathcal{T}, A, k)$ holds for a generic $A \in \mathbb{R}^{m\times n}$.

2.3. Conditions for $\mathrm{hsp}(\mathcal{T}, A, k)$

Recall that $\mathcal{T}$ is a finite set of matrices. We first handle the finiteness of $\mathcal{T}$ by the following fact.

Proposition 3. *Suppose that* $\mathrm{hsp}(\{T_1, T_2\}, A, k)$ *holds for every $T_1, T_2 \in \mathcal{T}$ and for a generic $A \in \mathbb{R}^{m\times n}$, then* $\mathrm{hsp}(\mathcal{T}, A, k)$ *holds for $A \in \mathbb{R}^{m\times n}$ generic.*

Proof. This follows directly from the fact that the intersection of finitely many non-empty Zariski open subsets of $\mathbb{R}^{m\times n}$ is again non-empty and Zariski open. □

Proposition 3 is intuitive, and it suggests us to focus on $\mathrm{hsp}(\mathcal{T}, A, k)$ with two matrices $T_1, T_2 \in \mathcal{T}$ fixed. In what follows we will consider when $\mathrm{hsp}(\{T_1, T_2\}, A, k)$ would be violated, so as to derive conditions for it to hold. One condition will be the rank constraint (10) (§2.3.1) and the other will be the *quasi-variety* constraint (12) (§2.3.2).

2.3.1. The rank constraint

By Definition 1, $\mathrm{hsp}(\{T_1, T_2\}, A, k)$ is tightly related to the column spaces of T_1 and T_2. This motivates us to consider

$$\mathcal{Z}_{T_1,T_2} := \{u \in \mathbb{C}^m : T_1u = T_2u\}. \tag{9}$$

Note that $\mathcal{Z}_{T_1,T_2}$ is a complex[3] linear subspace of $\mathbb{C}^m$, and therefore a complex algebraic variety. The role of $\mathcal{Z}_{T_1,T_2}$ in hsp can be seen as follows. Let x_1, x_2 be k-sparse and their images Ax_1, Ax_2 under A are vectors of $\mathcal{Z}_{T_1,T_2}$. Then $T_1Ax_1 = T_2Ax_2$ is the same as $T_1A(x_1 - x_2) = 0$. Hence, whether $T_1Ax_1 = T_2Ax_2$ implies $x_1 = x_2$ reduces to the problem of compressed sensing, and the answer is immediate from inspecting the spark or Kruskal rank of T_1A (recall §1.1). As a consequence, to fully understand $\mathrm{hsp}(\{T_1, T_2\}, A, k)$ it suffices to only focus on k-sparse vectors whose images under A are not contained in $\mathcal{Z}_{T_1,T_2}$.

We now therefore assume, *without loss of generality* (see Peng & Tsakiris (2021) for full arguments), that $\mathcal{Z}_{T_1,T_2}$ does not intersect the images of all k-sparse vectors under a generic A. Since the latter form a union of subspaces in $\mathbb{R}^m$ of dimension k, we can simply assume that $\mathcal{Z}_{T_1,T_2}$ has dimension smaller than or equal to $m - k$. Under this assumption we obtain the following characterization.

Proposition 4. *Let* $\dim(\mathcal{Z}_{T_1,T_2}) \le m - k$, $\mathrm{rank}[T_1\ T_2] < 2k$. $\mathrm{hsp}(\{T_1, T_2\}, A, k)$ *is false for $A \in \mathbb{R}^{m\times n}$ generic.*

Proof. It suffices to show that $\mathrm{hsp}(\{T_1, T_2\}, A_i, k)$ is false for $A_i \in \mathbb{R}^{m\times k}$ generic (here one could regard A_i as con-

[3] While T_1 and T_2 are real matrices, we define $\mathcal{Z}_{T_1,T_2}$ as a complex object on account of technical reasons.

sisting of k distinct columns of A). If $\mathrm{rank}(T_1 A_i) < k$ then $\mathrm{hsp}(\{T_1\}, A_i, k)$ is false. Hence let us assume $m \geq k$, $r \geq k$, $\mathrm{rank}(T_1 A_i) = k$, and similarly assume $\mathrm{rank}(T_2 A_i) = k$. But $[T_1\ T_2] \in \mathbb{R}^{r \times 2m}$ has rank smaller than $2k$, so does the $r \times 2k$ matrix $[T_1 A_i\ T_2 A_i]$. As a result, there are non-zero vectors $x_1, x_2 \in \mathbb{R}^k$ such that $T_1 A_i x_1 = T_2 A_i x_2$. Assume $x_1 = x_2$. Then $A_i x_1 = A_i x_2$ and $A_i x_1$ is an element of $\mathcal{Z}_{T_1,T_2}$. But $\mathcal{Z}_{T_1,T_2} \cap \mathbb{R}^m$ has dimension at most $m - k$, so, for a generic $A_i \in \mathbb{R}^{m \times k}$, the column space intersects $\mathcal{Z}_{T_1,T_2}$ only at zero. This gives $A_i x_1 = 0$ and $x_1 = 0$, a contradiction. Hence $x_1 \neq x_2$. □

To prevent Proposition 4 from happening, we consider:

The Rank Constraint

$$\mathrm{rank}(T) \geq 2k, \quad \forall\, T \in \mathcal{T}. \tag{10}$$

The rank constraint (10) ensures that $\mathrm{rank}[T_1\ T_2] \geq 2k$ for any $T_1, T_2 \in \mathcal{T}$, so that the bad situation in Proposition 4 would never occur. This constraint is perhaps the simplest, because it does not involve any interaction of T_1 and T_2.

However, the rank constraint does not exclude all possible violations of $\mathrm{hsp}(\{T_1, T_2\}, A, k)$, as illustrated next.

Example 1. *Suppose $m = 4, r = 4, k = 2$. For any $A \in \mathbb{R}^{4 \times n}$, let $a_4^\top \in \mathbb{R}^{1 \times n}$ be the last row of A. Note that the null space of $a_4^\top$ has dimension at least $n - 1$, it must intersect any subspace of dimension 2 at some non-zero point. Hence there is a non-zero k-sparse vector u of $\mathbb{R}^n$ which satisfies $a_4^\top u = 0$. Let T_1 and T_2 be defined as below.*

$$T_1 = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 9 & 10 & 11 & 12 \\ 13 & 14 & 15 & 16 \end{bmatrix} \qquad T_2 = \begin{bmatrix} 3 & 6 & 9 & 20 \\ 15 & 18 & 21 & 40 \\ 27 & 30 & 33 & 60 \\ 39 & 42 & 45 & 80 \end{bmatrix}$$

Then we have $T_1 A(3u) = T_2 A u$, but $3u \neq u$. The rank constraint is satisfied, but $\mathrm{hsp}(\{T_1, T_2\}, A, k)$ is violated.

2.3.2. The quasi-variety constraint

Here we introduce a constraint on $\{T_1, T_2\}$ under which the situation of Example 1 would not happen. This involves the following algebraic-geometric object. For a column vector w of m variables, consider all 2×2 determinants of the $r \times 2$ matrix $[T_1 w\ \ T_2 w]$. Each such determinant is a quadratic polynomial in entries of w. Let $\mathcal{Y}_{T_1,T_2} \subset \mathbb{C}^m$ be the complex algebraic variety defined by those determinants.

Example 2. *For $m = 3$, $r = 2$ and*

$$T_1 = \begin{bmatrix} 0 & 0 & 2 \\ 2 & 4 & 1 \end{bmatrix}, \quad \textit{and} \quad T_2 = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix},$$

the variety $\mathcal{Y}_{T_1,T_2}$ consists of the complex roots of the following polynomial p in variables w_1, w_2, and w_3.

$$p = \det \begin{bmatrix} 2w_3 & w_1 + 2w_2 + 3w_3 \\ 2w_1 + 4w_2 + w_3 & 4w_1 + 5w_2 + 6w_3 \end{bmatrix}$$

Alternatively and equivalently, we might describe $\mathcal{Y}_{T_1,T_2}$ as the set of vectors u's of $\mathbb{C}^m$ for which $T_1 u$ and $T_2 u$ are linearly dependent. Observing that $\mathcal{Z}_{T_1,T_2}$ of (9) is a subvariety of $\mathcal{Y}_{T_1,T_2}$, we define the following set

$$\mathcal{U}_{T_1,T_2} := \mathcal{Y}_{T_1,T_2} \backslash \mathcal{Z}_{T_1,T_2} \tag{11}$$

to be the set-theoretical difference between two varieties $\mathcal{Y}_{T_1,T_2}$ and $\mathcal{Z}_{T_1,T_2}$, with one containing the other. Based on the definition, $\mathcal{U}_{T_1,T_2}$ is usually named as a *quasi-variety*.

Letting $b := \binom{n}{k}$, $[b] := \{1, \ldots, b\}$, and denoting the set of submatrices of A of size $m \times k$ by $\{A_i\}_{i=1}^b$, we show that $\mathcal{U}_{T_1,T_2}$ is of potential harm to $\mathrm{hsp}(\{T_1, T_2\}, A, k)$:

Proposition 5. *If for any $i \in [b]$, the intersection of $\mathcal{U}_{T_1,T_2}$ and the column space $R(A_i)$ of A_i is not empty, then $\mathrm{hsp}(\{T_1, T_2\}, A, k)$ is not true.*

Proof. Suppose that $R(A_i) \cap \mathcal{U}_{T_1,T_2}$ is not empty and let $u \in R(A_i) \cap \mathcal{U}_{T_1,T_2}$. Then there is some $\lambda \in \mathbb{R}$ such that $T_1 u = \lambda T_2 u$ or $\lambda T_1 u = T_2 u$. Since $u \in \mathcal{U}_{T_1,T_2}$, we have $u \neq 0$ and $\lambda \neq 1$. Since $u \in R(A_i)$ we have for some $x \in \mathbb{R}^n$ with $A_i x = u$ that $T_1 A_i x = T_2 A_i (\lambda x)$ or $T_1 A_i (\lambda x) = T_2 A_i x$. But $x \neq \lambda x$. □

Proposition 5 suggests that the bad event where $\mathcal{U}_{T_1,T_2}$ intersects $R(A)$ must be prevented. We then expect the quasi-variety $\mathcal{U}_{T_1,T_2}$ to be as of small *size* as possible. Its size can be modeled by *dimension*, an algebraic-geometric notion that assigns to each subset of $\mathbb{C}^m$ a non-negative integer with the convention $\dim(\varnothing) := -1$ (recall §2.1). To say that $\dim(\mathcal{U}_{T_1,T_2})$ is small is to say that $\mathcal{U}_{T_1,T_2}$ is small, which in turn implies that it is unlikely for $\mathcal{U}_{T_1,T_2}$ to intersect $R(A)$. We formalize this intuition below.

Proposition 6. *Suppose $\dim(\mathcal{U}_{T_1,T_2}) \leq m - k$. Then, for a generic matrix $A \in \mathbb{R}^{m \times n}$, the column space $R(A_i)$ of A_i does not intersect $\mathcal{U}_{T_1,T_2}$ for any $i \in [b]$.*

As per Proposition 6, enforcing $\mathcal{U}_{T_1,T_2}$ to have small dimension is indeed an effective means to exclude the bad event of $\mathcal{U}_{T_1,T_2}$ intersecting $R(A_i)$; that is, with $\dim(\mathcal{U}_{T_1,T_2}) \leq k$ the violation of $\mathrm{hsp}(\{T_1, T_2\}, A, k)$ in Proposition 5 would not happen. This justifies the following constraint:

The Quasi-variety Constraint

$$\dim(\mathcal{U}_{T_1,T_2}) \leq m - k, \quad \forall\, T_1, T_2 \in \mathcal{T}. \tag{12}$$

Remark 1. *The dimension* $\dim(\mathcal{U}_{T_1,T_2})$ *that appeared in* (12) *can be computed as follows. In commutative algebra terms (Cox et al., 2013; Eisenbud, 2013),* $\dim(\mathcal{U}_{T_1,T_2})$ *is equal to the (Krull) dimension of the vanishing ideal of the closure of* $\mathcal{U}_{T_1,T_2}$*. This vanishing ideal, by definition* (11)*, is the saturation of the vanishing ideal of* $\mathcal{Y}_{T_1,T_2}$ *with respect to the vanishing ideal of* $\mathcal{Z}_{T_1,T_2}$*. The saturation, and thus its Krull dimension, can be computed directly based on the definitions of* $\mathcal{Y}_{T_1,T_2}$ *and* $\mathcal{Z}_{T_1,T_2}$*, using commutative algebra software, e.g.,* `Macaulay2`*.*

Example 3. *In Example 1, we have* $\dim(\mathcal{U}_{T_1,T_2}) = 3$ *which is larger than* $m-k=2$*. As a result,* T_1 *and* T_2 *of Example 1 violate the quasi-variety constraint.*

The rank constraint (10) and quasi-variety constraint (12) are sufficient for $\mathrm{hsp}(\mathcal{T}, A, k)$, as we will see soon.

2.4. Unique recovery in sparse homomorphic sensing

Theorem 1. *If a finite set* $\mathcal{T} \subset \mathbb{R}^{r\times m}$ *of matrices satisfies the rank constraint* (10) *and quasi-variety constraint* (12)*, we have* $\mathrm{hsp}(\mathcal{T}, A, k)$ *for a generic* $A \in \mathbb{R}^{m\times n}$*.*

Constraints (10), (12) of Theorem 1 guarantee the uniqueness of the solution to sparse homomorphic sensing (6).

There are two major hurdles towards proving Theorem 1. One is the non-linearity brought by the set $\mathcal{K}$ of k-sparse vectors. The other is from the matrix set $\mathcal{T}$. Indeed, matrices of $\mathcal{T}$ might be non-diagonalizable and even non-invertible (e.g., see §4); this makes analysis here harder than in unlabeled sensing. Moreover, we have to consider the interaction of $\mathcal{K}$ and $\mathcal{T}$. Overcoming these hurdles requires several novel technical ideas; see Peng & Tsakiris (2021) for details.

We remark that (12) is the tightest in the following sense:

Proposition 7. *If* $\dim(\mathcal{U}_{T_1,T_2}) > m-k$*, then a generic* $\underline{A} \in \mathbb{C}^{m\times n}$ *violates* $\mathrm{hsp}(\mathcal{T}, \underline{A}, k)$*.*

2.4.1. Unique recovery in homomorphic sensing

Theorem 1 implies the following result (with k set to be n).

Corollary 1. *Suppose that a finite set* $\mathcal{T} \subset \mathbb{R}^{r\times m}$ *satisfies that* $\mathrm{rank}(T) \geq 2n$ *for any* $T \in \mathcal{T}$ *and that*

$$\dim(\mathcal{U}_{T_1,T_2}) \leq m-n, \ \ \forall\, T_1, T_2 \in \mathcal{T}. \tag{13}$$

Then we have $\mathrm{hsp}(\mathcal{T}, A)$ *for a generic* $A \in \mathbb{R}^{m\times n}$*.*

As mentioned, $\mathrm{hsp}(\mathcal{T}, A)$ is equivalent to the uniqueness of the solution to the homomorphic sensing problem (5).

To compare, note that Tsakiris (2018; 2020); Tsakiris & Peng (2019) used the same rank constraint $\mathrm{rank}(T) \geq 2n$ and a different quasi-variety constraint for $\mathrm{hsp}(\mathcal{T}, A)$. The quasi-variety constraint of Tsakiris (2018; 2020); Tsakiris & Peng (2019) is $\dim(\mathcal{U}_{PT_1,T_2}) \leq m-n$, where P is some *unknown* projection onto the column space of T_2.[4] It is easy to verify that $\mathcal{U}_{T_1,T_2}$ is a subset of $\mathcal{U}_{PT_1,T_2}$ and so $\dim(\mathcal{U}_{T_1,T_2}) \leq \dim(\mathcal{U}_{PT_1,T_2})$. Thus, in comparison, constraint (13) is tighter. In fact, it is the tightest in the sense of Proposition 7. Finally, constraint (13) is also simpler because it dispenses with the unknown projection matrix P.

2.5. Noisy homomorphic sensing

We consider the homomorphic sensing problem in the presence of noise $\epsilon \in \mathbb{R}^r$. Let $\overline{y} := y + \epsilon = T^*Ax^* + \epsilon$ be our measurements. The questions are i) how we can estimate x^*, given $\overline{y}, \mathcal{T}, A$, and ii) how good the estimate is.

For i), we shortly mention that we can in principle solve (7) to obtain an estimate $(\hat{x}, \hat{T})$ of interest via exhaustive search. Indeed, for each $T_0 \in \mathcal{T}$ compute the least-squares solution $x_0 := (T_0A)^\dagger \overline{y}$ which minimizes $\left\|\overline{y} - T_0Ax\right\|_2$ over $\mathbb{R}^n$ in variables x, where we used $(\cdot)^\dagger$ to denote the pseudoinverse of a matrix. Among all least-squares solutions, then, take $\hat{x}$ which causes the minimum residual error.

Question ii), or more specifically whether $\hat{x}$ is close to x, is our main focus. This question is naturally *discrete* for the following reason. For arbitrary noise ϵ, the optimal $\hat{T}$ can be any matrix of $\mathcal{T}$. Since $\mathcal{T}$ is an arbitrary discrete set of matrices, the corresponding $\hat{x}$ can be arbitrarily far from x^*.

We handle this discreteness by identifying "nice" matrices contained in $\mathcal{T}$; by "nice" we mean a subset $\mathcal{T}_1$ of $\mathcal{T}$ such that each matrix of $\mathcal{T}_1$ will yield a least-squares solution which is close to x^*. With $R(\cdot)$ denoting the column space of a matrix, a concrete definition of $\mathcal{T}_1$ is given as

$$\mathcal{T}_1 = \big\{T \in \mathcal{T} : y \in R(TA)\big\}.$$

With $\sigma(\cdot)$ denoting the largest singular value of a matrix, the next proposition explains why $\mathcal{T}_1$ is a "nice" set.

Proposition 8. *Assume* $T_0 \in \mathcal{T}_1$ *and that* $\mathrm{hsp}(\mathcal{T}, A)$ *holds for some* $A \in \mathbb{R}^{m\times n}$*. Then* $x_0 - x^* = (T_0A)^\dagger \epsilon$ *where* $x_0 = (T_0A)^\dagger \overline{y}$*. In particular* $\left\|x_0 - x^*\right\|_2 \leq \sigma((T_0A)^\dagger)\left\|\epsilon\right\|_2$*.*

Under the uniqueness assumption for the homomorphic sensing problem ($\mathrm{hsp}(\mathcal{T}, A)$), Proposition 8 states that any $T_0 \in \mathcal{T}$ results a stable least-squares estimate x_0, whose distance to x^* can be upper bounded in terms of noise and data. As for the estimate $(\hat{x}, \hat{T})$ of (7), the remaining question is whether $\hat{T}$ is a "nice" matrix contained in $\mathcal{T}_1$.

First note that $\mathcal{T}_1$ is not empty because $y = T^*Ax^*$ and $T^* \in \mathcal{T}_1$. Also, if $\mathcal{T}_1 = \mathcal{T}$ then $\hat{T}$ is of course an element of $\mathcal{T}_1$. In fact, our next claim is that $\hat{T}$ is always "nice" (i.e., $\hat{T} \in \mathcal{T}_1$) in presence of sufficiently small noise.

[4] Here $\mathcal{U}_{PT_1,T_2}$ is defined by replacing T_1 of $\mathcal{U}_{T_1,T_2}$ with PT_1.

Proposition 9. *We have $\hat{T} \in \mathcal{T}_1$ whenever $\mathcal{T}_1 = \mathcal{T}$ or*

$$\|\epsilon\|_2 < \|y\|_2 \Big(1 - \max_{T \in \mathcal{T} \setminus \mathcal{T}_1, x \in \mathbb{R}^n} \frac{y^\top TAx}{\|y\|_2 \|TAx\|_2}\Big). \quad (14)$$

Since for every $T' \in \mathcal{T} \setminus \mathcal{T}_1$, the column space of $T'A$ does not contain y, the maximization term of (14) is strictly smaller than 1. Hence, the right-hand side of (14) is positive.

From Propositions 8 and 9, we are ready to draw a local stability result for noisy homomorphic sensing.

Theorem 2. *Suppose i)* $\mathrm{hsp}(\mathcal{T}, A)$ *holds true, ii)* $\mathcal{T}_1 = \mathcal{T}$ *or* (14) *holds, then* $\hat{x} - x^* = (\hat{T}A)^\dagger \epsilon$*, and in particular* $\|\hat{x} - x^*\|_2 \leq \sigma((\hat{T}A)^\dagger)\|\epsilon\|_2$.

Condition (14) defines a non-asymptotic regime, where the local stability of $\hat{x}$ is guaranteed (Theorem 2). In particular, if $\mathcal{T} = \mathcal{S}_m$, Theorem 2 is an improvement over the asymptotic result of Unnikrishnan et al. (2018).

3. Unlabeled compressed sensing

3.1. Theory

Recall Proposition 1 where the uniqueness for unlabeled compressed sensing is guaranteed. Here we derive Proposition 1 from Corollary 1. When $\mathcal{T}$ is $\mathcal{S}_m$, some algebraic properties of permutations might be utilized to simplify the rank and quasi-variety constraints. Indeed, every permutation of $\mathcal{S}_m$ has rank m, so the rank constraint becomes $m \geq 2k$, a requirement on the number of samples. Moreover, inspired by Tsakiris (2018; 2020); Tsakiris & Peng (2019), an interesting result is that, whenever the rank constraint is fulfilled, the quasi-variety constraint is automatically satisfied (this is not always true for a set of arbitrary matrices $\mathcal{T}$; see Example 1):

Proposition 10. *For two permutation matrices $\Pi_1, \Pi_2 \in \mathcal{S}_m$, we have $\dim(\mathcal{U}_{\Pi_1,\Pi_2}) \leq m - k$ as long as $m \geq 2k$.*

Combining Corollary 1 with Proposition 10 gives:

Corollary 2. *The following is true for $A \in \mathbb{R}^{m \times n}$ generic:*

$$m \geq 2k \Rightarrow \mathrm{hsp}(\mathcal{S}_m, A, k)$$

Recalling Proposition 2, we note that, for a generic $A \in \mathbb{R}^{m \times n}$, the condition $m \geq 2k$ of Corollary 2 implies the uniqueness of the solution to unlabeled compressed sensing 8. Thus, Corollary 2 is the same as Proposition 1.

3.2. Algorithm

Besides the k-sparsity assumption on x^*, we also assume that, in light of Slawski & Ben-David (2019), the ground-truth permutation matrix Π^* is p-sparse, i.e., $\|y - Ax\|_0 \leq p$ (see §1). This naturally leads us to the problem

$$\min_{x \in \mathbb{R}^n} \|y - Ax\|_0 \quad \text{s.t.} \quad \|x\|_0 \leq k. \quad (15)$$

Problem (15) is in general NP-hard, so we relax it into

$$\min_{x \in \mathbb{R}^n} \|y - Ax\|_1 \quad \text{s.t.} \quad \|x\|_0 \leq k. \quad (16)$$

The objective function of (16) is about an old problem, *least absolute deviation*, also known as *sparse error correction*; see, e.g., Kendall (1960); Candes & Tao (2005). The next natural choice is further relaxing the sparsity constraint of (16), so as to arrive at the convex problem[5] of minimizing $\|y - Ax\|_1 + \lambda \|x\|_1$ in n variables $x \in \mathbb{R}^n$ with some hyper-parameter $\lambda > 0$. But such relaxation does not yield satisfactory performance for our purpose.

We solve (16) using the idea of *hard thresholding pursuit* (Foucart, 2011; Cai et al., 2020). Following Cai et al. (2020), we assume that k is known in advance, and use $x^{(0)} := 0$ as initialization. The iterative update is given as:

$$x^{(t+1)} \leftarrow \mathrm{Proj}_{\mathcal{K}}\big(x^{(t)} - \mu A^\top \mathrm{sgn}(Ax^{(t+1)} - y)\big) \quad (17)$$

$$J \leftarrow \text{the support } \{i : x_i^{(t+1)} \neq 0\} \text{ of } x^{(t+1)}$$

$$x_J^{(t+1)} \leftarrow \operatorname*{argmin}_{x \in \mathbb{R}^n} \|y - A_J x\|_1 \quad (18)$$

In (17), we note that i) $A^\top \mathrm{sgn}(Ax^{(t+1)} - y)$ is a subgradient of $\|y - Ax\|_1$ at the point $x^{(t+1)}$, where sgn: $\mathbb{R}^n \to \mathbb{R}^n$ sends $[v_1, \ldots, v_n]^\top$ to a vector whose i-th entry is 1 if $v_i \geq 0$, or -1 otherwise, ii) μ is a step size to be determined, iii) $\mathrm{Proj}_{\mathcal{K}}(\cdot)$ projects a vector to its closest k-sparse counterpart. In (18), we update the non-zero entries $x_J^{(t+1)}$ of $x^{(t+1)}$ by solving the convex optimization problem, where A_J is the column-submatrix of A with its columns indexed by J; A_J is a tall matrix under the tacit assumption $m \geq 2k$. We note two differences of the algorithm from (Cai et al., 2020). First, instead of (18) they solved a least-squares problem. Second, they run the algorithm by one iteration.

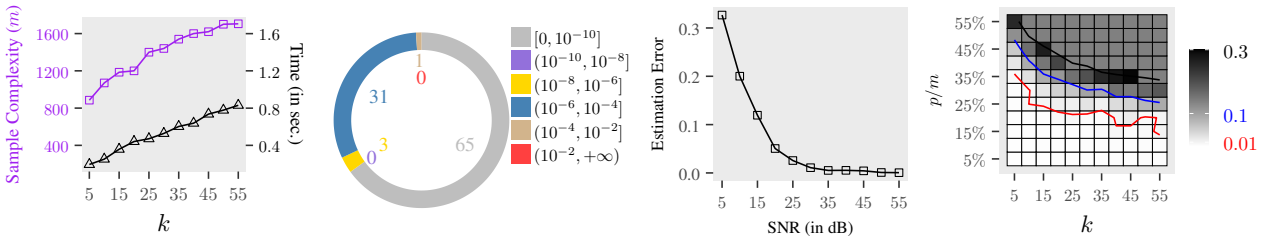
While (17) is straightforward to compute, we solve (18) by invoking an ADMM algorithm implemented in the `FOM` toolbox of Beck & Guttmann-Beck (2019).

3.3. Experiments

We evaluate the algorithm with $\mu := 10^{-4}$ and with the number of iterations set to $T := 20$ on an Intel(R) i7-8650 U, 1.9 GHz, 16 GB machine.[6] We have not known obvious baselines or other approaches for the task of interest.

[5] This problem was considered by Wright & Ma (2009) with $\lambda = 1$ in the context of *dense error correction*, where they assumed that the ground-truth signal x^* has non-negative entries.

[6] Other choices of μ did not yield significantly better results. The algorithm usually converges in about 10 iterations.



(a) sample/time complexity (b) error distribution (c) robustness to noise (d) phase transition under noise

Figure 1. The performance of the algorithm on synthetic data.

Data generation. We generate data by i) randomly sampling the entries of $A \in \mathbb{R}^{m\times n}$ from the standard normal distribution $\mathcal{N}(0,1)$, ii) randomly selecting a support of the k-sparse $x^* \in \mathbb{R}^n$ whose non-zero entries are randomly sampled also from $\mathcal{N}(0,1)$, iii) randomly producing a p-sparse permutation Π^*, and iv) computing $y = \Pi^* A x^*$.

Evaluation metrics. One evaluation metric which we use is the estimation error, computed as $\|x^* - x^{(\text{opt})}\|_2 / \|x^*\|_2$, where $x^{(\text{opt})}$ is among $\{x^{(1)}, \dots, x^{(T)}\}$ which minimizes (16).[7] Inspired by Netrapalli et al. (2013); Netrapalli et al. (2015), the other evaluation metric is the (empirical) sample complexity. Similar to Netrapalli et al. (2013), the algorithm is said to *succeed* if the estimation error is smaller than 0.01. The sample complexity of the algorithm is then the smallest among $\{2k, 3k, \dots\}$ for which the algorithm always succeeds over 100 trials for a fixed k.

Results. Figure 1 depicts the performance of the algorithm on synthetic data, with $n = 2000$ fixed. In Figure 1a we set $p := \lfloor 0.2m \rfloor$, and observed that the sampling complexity m increased as the sparsity k grew, which in turn entailed an increased running time. For example, when $n = 2000$ and $k = 25$, it took $m = 1400$ samples for the algorithm to succeed and 0.47 seconds to finish computation. Zooming in on the 100 trials at $k = 25$ of Figure 1a yields Figure 1b, where the estimation errors for the 100 trials were summarized. We saw that the estimation error is no more than 10^{-10} for 65 trials, and 96% of the 100 errors fall into the intervals $(10^{-6}, 10^{-4}]$ and $(-\infty, 10^{-10}]$.

Keeping $m = 1400, n = 2000, k = 25, p = \lfloor 0.2m \rfloor$ fixed, we furthermore evaluated the robustness of the algorithm to noise. We added noise to the measurements y as per the SNR, run the algorithm, and the result was in Figure 1c (100 trials). As the SNR condition improved, the estimation error declined, from 0.3268 (5dB) to 0.011 (30dB) and further to 0.0005 (55dB). Finally, a holistic understanding on the algorithm might be obtained via Figure 1d, where we fixed $m = 1400, n = 2000$ and SNR= 40dB and presented the estimation errors with the two sparsity levels k and p varying (100 trials). We observed that the algorithm consistently made errors smaller than 0.01 in the presence of $\leq 20\%$ shuffled data and $k \leq 35$. In the extremely sparse case $k = 5$, the algorithm could tolerate up to 45% shuffled data (with errors no more than 0.1). On the other hand, the algorithm could fail in an attempt at working at the challenging high-p, high-k region. To summarize, the algorithm was shown to be time-efficient, robust to noise, and to succeed when the ground-truth x^* and Π^* are both sufficiently sparse.

[7] Recall that a subgradient might not give a descent direction.

4. A broader picture

The matrix set $\mathcal{T}$ in (sparse) homomorphic sensing ((5), (6)) provides some flexibility to model other important inverse problems than unlabeled sensing. We next present several other choices of $\mathcal{T}$ than $\mathcal{S}_m$ that arise from data applications.

Unlabeled sensing with missing entries. In fact, (Unnikrishnan et al., 2018) considered a more general version of unlabeled sensing, where some entries of y are missing and the positions of missing entries in y are unknown. In other words, it means that i) one is given $y = S^* A x^* \in \mathbb{R}^r$ with some unknown *selection* matrix S^*, i.e., S^* is a permutation matrix with $(m-r)$ rows removed, and ii) one aims to solve

$$y = SAx, \quad S \in \mathcal{S}_{r,m}, \quad x \in \mathbb{R}^n, \tag{19}$$

for x, where $\mathcal{S}_{r,m}$ is the set of $r \times m$ selection matrices.[8] This was known by Tsakiris (2018; 2020); Tsakiris & Peng (2019) as an example of homomorphic sensing.

Missing data recovery. We find that the problem of *missing data recovery* (Zhang, 2006; Liu et al., 2017; Liu et al., 2019) or of *signal recovery with erasures at known locations* (Han & Sun, 2014) is also a special case of homomorphic sensing (5). In this problem one aims to recover x^* from $y = O^* A x^*$, where O^* is some $m \times m$ diagonal matrix

[8] In Unnikrishnan et al. (2018), (19) was called as unlabeled sensing and (2) as an important special case. The follow-up works referred to (2) as unlabeled sensing, or as *linear regression without correspondences*, or as *shuffled linear regression*; see, e.g., Hsu et al. (2017). In this paper we used unlabeled sensing for (2) and *unlabeled sensing with missing entries* for (19).

with 0 or 1 on its diagonal. Its differences from (19) are that, i) in general, the positions at which the entries are missing are understood from the positions of non-zero entries of y, and ii) there is no unknown permutation involved.

Real phase retrieval. We also find that the perhaps more familiar problem of *real phase retrieval* is another homomorphic sensing example. This problem can be traced back to the 1910s when the research on *X-ray crystallography* was launched, and has been receiving increasing attention in recent years; see, e.g., Grohs et al. (2020) for a vivid account. In this problem, we are given

$$y = BAx, \quad B \in \mathcal{B}_m, \quad x \in \mathbb{R}^n, \tag{20}$$

where $y = B^*Ax^*$, $B^* \in \mathcal{B}_m$, and $\mathcal{B}_m$ is the set of $m \times m$ *sign* matrices, i.e., diagonal matrices with ± 1 on the diagonal. Since uniquely recovering x^* is impossible[9], the goal then becomes unique recovery of x^* up to sign. The problem of *symmetric mixture of two linear regressions* (Balakrishnan et al., 2017) also admits formulation (20); see, e.g., Chen et al. (2019); Klusowski et al. (2019) for a discussion which connects the two problems.

The final example is a combination of (19) and (20), explored by Lv & Sun (2018). This involves the matrix set $\mathcal{S}_{r,m}\mathcal{B}_m := \{SB : S \in \mathcal{S}_{r,m}, B \in \mathcal{B}_m\}$ and the relation

$$y = CAx, \quad C \in \mathcal{S}_{r,m}\mathcal{B}_m, \quad x \in \mathbb{R}^n.$$

To summarize, the above problems are concerned with missing correspondences, missing values, missing signs, or combinations thereof, and they are actually of the same type, where the linear measurements Ax^* have further undergone some unknown linear map belonging to a specific set of maps, e.g., $\mathcal{S}_m$, $\mathcal{S}_{r,m}$, $\mathcal{B}_m$, and $\mathcal{S}_{r,m}\mathcal{B}_m$. In Peng & Tsakiris (2021) we present the applications of our theory to those examples, which yield either i) known results from prior works, e.g., Balan et al. (2006); Unnikrishnan et al. (2018); Han et al. (2018); Lv & Sun (2018); Dokmanic (2019); Akçakaya & Tarokh (2014); Wang & Xu (2014), or ii) even novel results for those examples. Finally, it is natural to consider our theory as having potential wider applicability to new examples of homomorphic sensing yet to discover.

5. Discussion and future work

On the theoretical part, we presented conditions guaranteeing the uniqueness for sparse homomorphic sensing (6), from which a uniqueness result for unlabeled compressed sensing follows. The next step for research is to find conditions under which the corresponding ℓ_1 relaxation (e.g., (16)) has a unique solution, which we leave as future work. Taking noise into consideration, we provided a deterministic condition for the local stability in homomorphic sensing, from which a probabilistic condition might be derived.

On the algorithmic front, we initiated a computational investigation into unlabeled compressed sensing. Future work might include developing theoretical guarantees for the algorithm of §3.2, tackling the case where more data are shuffled, dispensing with the hyper-parameters, and so on. Finally, we presented a broader picture in §4 using the homomorphic sensing framework. Tools from other fields might be key to advancing the research for unlabeled (compressed) sensing.

[9] Both (B^*, x^*) and $(-B^*, -x^*)$ satisfy (20).

References

Abbasi, A., Tasissa, A., and Aeron, S. Unlabeled sensing with local permutations. Technical report, arXiv:1911.06382v3 [eess.SP], 2020.

Abbasi, A., Tasissa, A., and Aeron, S. R-local sensing: A novel graph matching approach for multiview unlabeled sensing under local permutations. *IEEE Open Journal of Signal Processing*, pp. 1–1, 2021.

Abid, A. and Zou, J. A stochastic expectation-maximization approach to shuffled linear regression. In *Annual Allerton Conference on Communication, Control, and Computing*, pp. 470–477, 2018.

Abid, A., Poon, A., and Zou, J. Linear regression with shuffled labels. *arXiv preprint arXiv:1705.01342*, 2017.

Akçakaya, M. and Tarokh, V. New conditions for sparse phase retrieval. Technical report, arXiv:1310.1351v2 [cs.IT], 2014.

Antoni, M. and Schnell, R. The past, present and future of the german record linkage center. *Jahrbücher für Nationalökonomie und Statistik*, 239(2):319 – 331, 2019.

Balakrishnan, S., Wainwright, M. J., and Yu, B. Statistical guarantees for the em algorithm: From population to sample-based analysis. *Annals of Statistics*, 45(1):77–120, 2017.

Balan, R., Casazza, P., and Edidin, D. On signal reconstruction without phase. *Applied and Computational Harmonic Analysis*, 20(3):345 – 356, 2006.

Beck, A. and Guttmann-Beck, N. Fom – a matlab toolbox of first-order methods for solving convex optimization problems. *Optimization Methods and Software*, 34(1): 172–193, 2019.

Cai, J.-F., Li, J., Lu, X., and You, J. Sparse signal recovery from phaseless measurements via hard thresholding pursuit. Technical report, arXiv:2005.08777v2 [math.NA], 2020.

Candes, E. J. and Tao, T. Decoding by linear programming. *IEEE Transactions on Information Theory*, 51(12):4203–4215, 2005.

Chen, Y., Chi, Y., Fan, J., and Ma, C. Gradient descent with random initialization: fast global convergence for nonconvex phase retrieval. *Mathematical Programming*, 176(1):5–37, 2019.

Cohen, A., Dahmen, W., and DeVore, R. Compressed sensing and best k-term approximation. *Journal of the American mathematical society*, 22(1):211–231, 2009.

Cox, D., Little, J., and OShea, D. *Ideals, Varieties, and Algorithms: An Introduction to Computational Algebraic geometry and Commutative Algebra*. Springer Science & Business Media, 2013.

Dokmanic, I. Permutations unlabeled beyond sampling unknown. *IEEE Signal Processing Letters*, 26(6):823–827, 2019.

Domingo-Ferrer, J. and Muralidhar, K. New directions in anonymization: Permutation paradigm, verifiability by subjects and intruders, transparency to users. *Information Sciences*, 337-338:11 – 24, 2016.

Donoho, D. L. and Elad, M. Optimally sparse representation in general (nonorthogonal) dictionaries via l1 minimization. *Proceedings of the National Academy of Sciences*, 100(5):2197–2202, 2003.

Eisenbud, D. *Commutative Algebra: with a view toward algebraic geometry*, volume 150. Springer Science & Business Media, 2013.

Fellegi, I. P. and Sunter, A. B. A theory for record linkage. *Journal of the American Statistical Association*, 64(328): 1183–1210, 1969.

Foucart, S. Hard thresholding pursuit: An algorithm for compressive sensing. *SIAM Journal on Numerical Analysis*, 49(6):2543–2563, 2011.

Grohs, P., Koppensteiner, S., and Rathmair, M. Phase retrieval: Uniqueness and stability. *SIAM Review*, 62(2): 301–350, 2020.

Han, D. and Sun, W. Reconstruction of signals from frame coefficients with erasures at unknown locations. *IEEE Transactions on Information Theory*, 60(7):4013–4025, 2014.

Han, D., Lv, F., and Sun, W. Recovery of signals from unordered partial frame coefficients. *Applied and Computational Harmonic Analysis*, 44(1):38 – 58, 2018.

Hsu, D., Shi, K., and Sun, X. Linear regression without correspondence. In *Advances in Neural Information Processing Systems*, 2017.

Jeong, M., Dytso, A., Cardone, M., and Poor, H. V. Recovering data permutations from noisy observations: The linear regime. *IEEE Journal on Selected Areas in Information Theory*, 1(3):854–869, 2020.

Jeong, M., Dytso, A., and Cardone, M. Retrieving data permutations from noisy observations: High and low noise asymptotics. Technical report, arXiv:2105.03015v1 [cs.IT], 2021.

Ji, R., Liang, Y., Xu, L., and Zhang, W. A concave optimization-based approach for joint multi-target track initialization. *IEEE Access*, 7:108551–108560, 2019.

Kendall, M. G. Studies in the history of probability and statistics. where shall the history of statistics begin? *Biometrika*, 47(3/4):447–449, 1960.

Klusowski, J. M., Yang, D., and Brinda, W. D. Estimating the coefficients of a mixture of two linear regressions by expectation maximization. *IEEE Transactions on Information Theory*, 65(6):3515–3524, 2019.

Kruskal, J. B. Three-way arrays: rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics. *Linear Algebra and its Applications*, 18(2):95 – 138, 1977.

Lian, W., Zhang, L., and Yang, M. An efficient globally optimal algorithm for asymmetric point matching. *IEEE Trans. Pattern Anal. Mach. Intell.*, 39(7):1281–1293, 2017.

Liu, G., Liu, Q., and Yuan, X. A new theory for matrix completion. In *Advances in Neural Information Processing Systems*, pp. 785–794, 2017.

Liu, G., Liu, Q., Yuan, X. T., and Wang, M. Matrix completion with deterministic sampling: Theories and methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1–1, 2019.

Lv, F. and Sun, W. Real phase retrieval from unordered partial frame coefficients. *Advances in Computational Mathematics*, 44(3):879–896, 2018.

Marano, S. and Willett, P. Making decisions by unlabeled bits. *IEEE Transactions on Signal Processing*, 68:2935–2947, 2020.

Muralidhar, K. Record re-identification of swapped numerical microdata. *Journal of Information Privacy and Security*, 13(1):34–45, 2017.

Nejatbakhsh, A. and Varol, E. Neuron matching in c. elegans with robust approximate linear regression without correspondence. In *IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 2837–2846, 2021.

Netrapalli, P., Jain, P., and Sanghavi, S. Phase retrieval using alternating minimization. In *Advances in Neural Information Processing Systems*, pp. 2796–2804, 2013.

Netrapalli, P., Jain, P., and Sanghavi, S. Phase retrieval using alternating minimization. *IEEE Transactions on Signal Processing*, 63(18):4814–4826, 2015.

Pananjady, A., Wainwright, M. J., and Courtade, T. A. Denoising linear models with permuted data. In *IEEE International Symposium on Information Theory*, pp. 446–450, June 2017.

Pananjady, A., Wainwright, M. J., and Courtade, T. A. Linear regression with shuffled data: Statistical and computational limits of permutation recovery. *IEEE Transactions on Information Theory*, 64(5):3286–3300, 2018.

Peng, L. and Tsakiris, M. C. Linear regression without correspondences via concave minimization. *IEEE Signal Processing Letters*, 27:1580–1584, 2020.

Peng, L. and Tsakiris, M. C. Homomorphic sensing of subspace arrangements. Technical report, arXiv:2006.05158v3 [cs.LG], 2021.

Peng, L., Song, X., Tsakiris, M. C., Choi, H., Kneip, L., and Shi, Y. Algebraically-initialized expectation maximization for header-free communication. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 5182–5186, 2019.

Shi, X., Li, X., and Cai, T. Spherical regression under mismatch corruption with application to automated knowledge translation. *Journal of the American Statistical Association*, 0(0):1–12, 2020.

Slawski, M. and Ben-David, E. Linear regression with sparsely permuted data. *Electronic Journal of Statistics*, 13(1):1–36, 2019.

Slawski, M., Diao, G., and Ben-David, E. A pseudo-likelihood approach to linear regression with partially shuffled data. Technical report, arXiv:1910.01623 [stat.ME], 2019.

Slawski, M., Rahmani, M., and Li, P. A sparse representation-based approach to linear regression with partially shuffled labels. In *Conference on Uncertainty in Artificial Intelligence*, pp. 38–48, 2019.

Slawski, M., Ben-David, E., and Li, P. Two-stage approach to multivariate linear regression with sparsely mismatched data. *Journal of Machine Learning Research*, 21 (204):1–42, 2020.

Slawski, M., Diao, G., and Ben-David, E. A pseudo-likelihood approach to linear regression with partially shuffled data. *Journal of Computational and Graphical Statistics*, 0(0):1–31, 2021.

Song, X., Choi, H., and Shi, Y. Permuted linear model for header-free communication via symmetric polynomials. In *IEEE International Symposium on Information Theory*, pp. 661–665, 2018.

Theodoridis, S. *Machine Learning: A Bayesian and Optimization Perspective*. Academic Press, 2020.

Tsakiris, M. C. Eigenspace conditions for homomorphic sensing. Technical report, arXiv:1812.07966v1 [math.CO], 2018.

Tsakiris, M. C. Determinantal conditions for homomorphic sensing. Technical report, arXiv:1812.07966v6 [math.CO], 2020.

Tsakiris, M. C. and Peng, L. Homomorphic sensing. In *International Conference on Machine Learning*, 2019.

Tsakiris, M. C., Peng, L., Conca, A., Kneip, L., Shi, Y., and Choi, H. An algebraic-geometric approach for linear regression without correspondences. *IEEE Transactions on Information Theory*, 66(8):5130–5144, 2020.

Unnikrishnan, J., Haghighatshoar, S., and Vetterli, M. Unlabeled sensing: Solving a linear system with unordered measurements. In *Annual Allerton Conference on Communication, Control, and Computing*, pp. 786–793, 2015.

Unnikrishnan, J., Haghighatshoar, S., and Vetterli, M. Unlabeled sensing with random linear measurements. *IEEE Transactions on Information Theory*, 64(5):3237–3253, May 2018.

Wang, G., Marano, S., Zhu, J., and Xu, Z. Target localization by unlabeled range measurements. *IEEE Transactions on Signal Processing*, 68:6607–6620, 2020a.

Wang, Y. and Xu, Z. Phase retrieval for sparse signals. *Applied and Computational Harmonic Analysis*, 37(3): 531 – 544, 2014.

Wang, Z., Ben-David, E., and Slawski, M. Estimation in exponential family Regression based on linked data contaminated by mismatch error. Technical report, arXiv:2010.00181v2 [stat.ME], 2020b.

Wright, J. and Ma, Y. Dense error correction via l1-minimization. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 3033–3036, 2009.

Wright, J. and Ma, Y. *High-Dimensional Data Analysis with Low-Dimensional Models: Principles, Computation, and Applications*. Cambridge University Press, 2020.

Xie, Y., Mao, Y., Zuo, S., Xu, H., Ye, X., Zhao, T., and Zha, H. A hypergradient approach to robust regression without correspondence. In *International Conference on Learning Representations*, 2021.

Yao, Y., Peng, L., and Tsakiris, M. C. Unlabeled Principal Component Analysis. Technical report, arXiv:2101.09446v1 [cs.LG], 2021.

Zhang, H. and Li, P. Optimal estimator for unlabeled linear regression. In *International Conference on Machine Learning*, pp. 11153–11162, 2020.

Zhang, H., Slawski, M., and Li, P. Permutation recovery from multiple measurement vectors in unlabeled sensing. In *IEEE International Symposium on Information Theory*, pp. 1857–1861, 2019.

Zhang, Y. When is missing data recoverable? Technical report, 2006.

Zhu, J., Cao, H., Song, C., and Xu, Z. Parameter estimation via unlabeled sensing using distributed sensors. *IEEE Communications Letters*, 21(10):2130–2133, 2017.