Homomorphic Sensing: Sparsity and Noise

Anonymous Authors[1]

Abstract

Unlabeled sensing is a recent problem encompassing many data science and engineering applications and typically formulated as solving linear equations with its right-hand side vector undergoing 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. Under our tighter and simpler conditions than those of prior work, the homomorphic sensing problem admits a unique solution. Moreover, we find that the conditions are less demanding under a sparsity assumption, which in particular imply that the associated ℓ_0 minimization problem has a unique minimum. Furthermore, such a unique solution is locally stable under noise. We also consider this sparsity assumption in unlabeled sensing, leading to the problem of *unlabeled compressed sensing*, which we show admits a unique sparsest solution as long as there exists a sufficiently sparse solution. On the algorithmic level, we solve the unlabeled compressed sensing problem by an iterative algorithm, with its efficiency and effectiveness evidenced by synthetic data experiments. Finally, we connect several other important engineering problems to unlabeled sensing under the unified homomorphic sensing framework. (Peng, Jan 31)

1. Introduction

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 $v = Ax$ with $A \in \mathbb{R}^{m\times n}$, researchers have focused on the optimization problem

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

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, *spark* (Donoho & Elad, 2003) or *Kruskal rank* (Kruskal, 1977), and has been a major role in in theoretical foundations of compressed sensing; see Theodoridis (2020); Wright & Ma (2020). 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 is some x^* for which the sparsest solutions to (1) are not unique whenever $m < 2k < n$.

More 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 an *unknown* $m \times m$ permutation matrix Π^*, unlabeled sensing means that i) $y = \Pi^* A x^*$ and, ii) with y, A given, solving the equations[2]

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

for x, where $\mathcal{S}_m$ is the set of $m \times m$ permutation matrices. Unnikrishnan et al. (2018) proved[3] that the sufficient and necessary condition for (2) to admit x^* as the unique solution for $A \in \mathbb{R}^{m\times n}$ generic is $m \geq 2n$.

A notable development following Unnikrishnan et al. (2018) is a generalization of unlabeled sensing, 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, we have

$$y = TAx, \quad T \in \mathcal{T}, \quad x \in \mathbb{R}^n, \tag{3}$$

where we are now given the measurements as $y = T^* A x^*$ for some unknown $T^* \in \mathcal{T}$ and the goal is to solve (3) for x. Tsakiris (2018; 2020) proved that (3) admits x^* as

[1]Anonymous Institution, Anonymous City, Anonymous Region, Anonymous Country. Correspondence to: Anonymous Author <anon.email@domain.com>.

Preliminary work. Under review by the International Conference on Machine Learning (ICML). Do not distribute.

[1]For now it is safe to think of a generic A as "random" (§2.1).

[2]Its connection to compressed sensing is that the k-sparse x^* comes from the union of $\binom{n}{k}$ subspaces, while in unlabeled sensing the measurements y come from the union of $m!$ subspaces.

[3]We found that the result was also independently proved by Han et al. (2018) using different techniques.

the unique solution for a generic matrix A of size $m \times n$, whenever it holds that i) every matrix of $\mathcal{T}$ has rank at least $2n$, and ii) the algebraic-geometric *dimension*[4] of a specific set $\mathcal{U} \subset \mathbb{C}^m$ depending 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 condition of Unnikrishnan et al. (2018) which guarantees the uniqueness of the solution to (2).

Under this global picture we review related work next. We consider i) applications which motivate the unlabeled sensing problem, ii) theory, iii) algorithms developed for it.

Applications. We first examine in detail a data analysis application, *record linkage* (Fellegi & Sunter, 1969; 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 subsequential data analysis. Due to privacy concerns, each entry of the records corresponding to some individual is not associated with a unique identifier of this individual (e.g., the social security number). As a result, a computer-based linkage of the respective entries in two (or more) records corresponding to the same individual can be error-prone, yielding imperfect data for later analysis. It is thus of natural interest to ask whether one can perform linear regression on $y \in \mathbb{R}^m$ and $A \in \mathbb{R}^{m\times n}$, even without linking them. Without linkage, the correspondences between entries of y and rows of A are unknown. Such data imperfection might very well be modeled by an unknown permutation $\Pi^* \in \mathcal{S}_m$, and this gives $y = \Pi^* A x^*$ with x^* unknown. The aim is to recover x^* from y. This is exactly problem (1). Besides record linkage, other applications abound: signal estimation using distributed sensors (Song et al., 2018), target localization in signal processing (Wang et al., 2020), neuron matching in computational neuroscience (Nejatbakhsh & Varol, 2021), automated translation of medical codes (Shi et al., 2020) and flow cytometry (Abid & Zou, 2018) in biology, multi-target tracking (Ji et al., 2019) and point set registration (Pananjady et al., 2017) in computer vision; see, e.g., (Pananjady et al., 2018; Xie et al., 2021) for more.

Theory. While the aforementioned result of Unnikrishnan et al. (2018) holds for any $x^* \in \mathbb{R}^n$, Tsakiris et al. (2020) showed that $m \geq n+1$ samples are sufficient for the uniqueness, for both $x^* \in \mathbb{R}^m$ and $A \in \mathbb{R}^{m\times n}$ generic.

Since the uniqueness for (2) was settled, the noise case came into picture. With $\overline{y} := y + \epsilon = \Pi^* A x^* + \epsilon$ for some noise ϵ, Pananjady et al. (2018) showed that the estimator

$$(\hat{x}, \hat{\Pi}) \in \operatorname*{argmin}_{x\in\mathbb{R}^n,\, \Pi\in\mathcal{S}_m} \|\overline{y} - \Pi A x\|_2 \tag{4}$$

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:=$\|x^*\|_2^2/\sigma^2$ 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), which showed 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 an estimation whose distance to x^* is upper bounded in terms of p, n, m with high probability can be obtained.

Algorithms[5]**.** The above sparse assumption on Π^* also leads Slawski & Ben-David (2019) to a relaxation of (4), say minimizing $\|\overline{y} - Ax\|_1$ over $x \in \mathbb{R}^n$, solvable via convex optimization, and once solved, empirically it yields an estimation close to x^* as long as no more than half of 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 reweighted 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 (4), 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 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 as initialization fed to (4) for alternating minimization. This gives an algorithm of linear complexity in m, efficient for $n \leq 5$ or intractable otherwise. The algorithm of Peng & Tsakiris (2020) is based on a concave minimization reformulation of (4) solved via branch-and-bound and it can handle the case $n \leq 8$, while otherwise intractable.

The message is that, since the assumption $m \geq 2n$, necessary for the uniqueness of x^* to (2), was (implicitly) made by all of the above algorithms, none of them works at the ill-posed region $m < 2n$, let alone when $m < n$.

1.1. Contributions of this paper

We improve and generalize prior works in several ways. We first provide tighter and simpler conditions than those of Tsakiris (2018; 2020); Tsakiris & Peng (2019) for the homomorphic sensing problem (3) to have x^* as the unique

[4]We review the notion of "dimension" in §2.1.

[5]See, e.g., algorithms of Slawski et al. (2020); Zhang & Li (2020); Jeong et al. (2020) for other types of unlabeled data.

solution (Theorem 1). We next discuss generalizations.

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

$$\min_{x\in\mathbb{R}^n} \|x\|_0 \quad \text{s.t.} \quad y = TAx, \ \ T\in\mathcal{T}. \tag{5}$$

Under the assumption that x^* is a k-sparse solution to (5), our main result is Theorem 2 which provides conditions under which (5) admits x^* as the unique solution.

Noisy homomorphic sensing. We also extend homomorphic sensing (3) to *noisy* homomorphic sensing, where we are given the noisy measurements $\overline{y} = y + \epsilon = T^* A x^* + \epsilon$. We show in Theorem 3 that, as long as $\|\epsilon\|_2$ is *sufficiently small* and that (3) admits a unique solution, the following problem (6) produces a locally stable solution $\hat{x}$.

$$(\hat{x}, \hat{T}) \in \operatorname*{argmin}_{x\in\mathbb{R}^n, T\in\mathcal{T}} \|\overline{y} - TAx\|_2. \tag{6}$$

When setting $\mathcal{T}$ to $\mathcal{S}_m$ (4), we obtain an improved result over that of Unnikrishnan et al. (2018).

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

$$\min_{x\in\mathbb{R}^n} \|x\|_0 \quad \text{s.t.} \quad y = \Pi Ax, \ \ \Pi\in\mathcal{S}_m. \tag{7}$$

Clearly, (7) is a special case of (5), 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 (7) as long as $m \geq 2k$.*

Proposition 1 is a surprise to us. Indeed, the number $2k$ is the threshold for unique recovery of x^* in compressed sensing (recall the 1st paragraph of §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 (7), which we solve via an iterative algorithm based on subgradient descend and ℓ_1 minimization (§4.1). This is the first algorithm for unlabeled sensing which works even when $m < n$, a regime unexplored in prior works. By experiments (§4.2), 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. Last but not least, 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 a connection among those problems to be established under the name homomorphic sensing. We will discuss this in detail in §5.

2. Sparse homomorphic sensing

The uniqueness for (5) involves the measurements y, where $y = T^* A x^*$ depends on an arbitrary k-sparse solution x^*, and we want to guarantee unique recovery of all possible k-sparse $x^* \in \mathbb{R}^n$. This motivates the following definition.

Definition 1 (ℓ_0^k). *Given $A \in \mathbb{R}^{m\times n}$ and $\mathcal{T} \subset \mathbb{R}^{r\times m}$, if for any $T_1, T_2 \in \mathcal{T}$ and any k-sparse $x_1, x_2 \in \mathbb{R}^n$ we have*

$$T_1 A x_1 = T_2 A x_2 \ \ \Rightarrow \ \ x_1 = x_2, \tag{8}$$

then we say that $\mathcal{T}$ and A satisfy the unique ℓ_0 recovery property for k-sparse vectors, denoted as $\ell_0^k(\mathcal{T}, A)$.

Note that $\ell_0^n(\mathcal{T}, A)$ corresponds to the uniqueness of the solution to the homomorphic sensing problem (3). In fact, Definition 1 implies the following equivalence.

Proposition 2. *The following are equivalent.*

i) We have $\ell_0^k(\mathcal{T}, A)$.

ii) For any k-sparse vector x^* as a solution to problem (7), x^* is the unique sparsest solution.

Proof. We only prove that i) implies ii); the other direction is not difficult. Let x^+ be an optimal solution, so $\|x^+\|_0 \leq \|x^*\|_0$. Then x^+ is k-sparse and $y = T^+ A x^+$ for some $T^+ \in \mathcal{T}$. Since $y = T^* A x^*$ we get that $x^* = x^+$. $\square$

Certainly, approaching $\ell_0^k(\mathcal{T}, A)$ relies on understanding the role of i) the design matrix A, of ii) the matrix set $\mathcal{T}$, and of iii) the k-sparse vectors. For i), our goal will be proving $\ell_0^k(\mathcal{T}, A)$ for $A \in \mathbb{R}^{m\times n}$ generic, and we will define "generic" using algebraic geometry in §2.1. For ii), we will consider conditions to put on $\mathcal{T}$ which allow for $\ell_0^k(\mathcal{T}, A)$ to hold for $A \in \mathbb{R}^{m\times n}$ generic in §2.2. For iii), we will present our understanding on the non-linearity brought by k-sparse vectors in §2.3. These discussions will lead us to our main theoretical results, which we present in §2.4.

2.1. Algebraic geometry

The basic object in algebraic geometry is complex (*resp.* real) *algebraic variety*. To start with, define the complex (*resp.* real) *hypersurface* $\mathcal{H}$ as a subset of $\mathbb{C}^m$ (*resp.* $\mathbb{R}^m$), which consists of the set of complex (*resp.* real) roots of a polynomial p in m variables with complex (*resp.* real) coefficients; that is, $\mathcal{H} := \{z : p(z) = 0\}$. An algebraic variety is the intersection of some hypersurfaces, that is, the common roots of some polynomials. A subvariety of an algebraic variety $\mathcal{Q}$ is a subset of $\mathcal{Q}$ and is itself an algebraic variety. For example, any line and plane of $\mathbb{R}^3$ is an algebraic variety, and any line of $\mathbb{R}^3$ is a subvariety of some 2D plane. By a *generic* matrix of $\mathbb{R}^{m\times n}$ having some property, we mean that every matrix in the complement $\mathcal{C}$ of some proper subvariety $\mathcal{P}$ of $\mathbb{R}^{m\times n}$ satisfies this property. Intuitively[6], since $\mathcal{P}$ is the intersection of some hypersurfaces, a matrix randomly chosen from $\mathbb{R}^{m\times n}$ according to some continuous probability distribution will land itself in $\mathcal{C}$, with probability 1, for the same reason that the intersection of some 2D planes of $\mathbb{R}^3$ has measure 0.

By the above, we justified what we meant in §1 by a generic matrix $A\in\mathbb{R}^{m\times n}$. Also, with the fact that $\ell_0^k(\mathcal{T},A)$ could not hold for any A (e.g., when A is the zero matrix), the next best to hope is that $\ell_0^k(\mathcal{T},A)$ holds for a generic $A\in\mathbb{R}^{m\times n}$ under certain conditions on $\mathcal{T}$. Finding those conditions on $\mathcal{T}$ is the main theme of the next subsection (§2.2).

2.2. Conditions on the matrix set $\mathcal{T}$

The role of $\mathcal{T}$ in $\ell_0^k(\mathcal{T},A)$ requires more efforts to penetrate, and it is better appreciated by first focusing on $\ell_0^n(\mathcal{T},A)$, a property for homomorphic sensing (3) (we will bring the sparsity k back into picture later in §2.3 and §2.4). However, unlike in unlabeled sensing (2) where the combinatorial structure of $\mathcal{S}_m$ yielded insights that assist analysis of $\ell_0^n(\mathcal{T},A)$ (see Unnikrishnan et al. (2018); Han et al. (2018)), formulation (3) only gives that $\mathcal{T}$ is a finite set of $r\times m$ matrices ($r\le m$). How can we determine what kind of $\mathcal{T}$ would satisfy $\ell_0^n(\mathcal{T},A)$ for a generic $A\in\mathbb{R}^{m\times n}$?

We first handle the finiteness of $\mathcal{T}$ by the following fact.

Proposition 3. *Suppose that $\ell_0^n(\{T_1,T_2\},A)$ holds for every $T_1,T_2\in\mathcal{T}$, for a generic $A\in\mathbb{R}^{m\times n}$, then $\ell_0^n(\mathcal{T},A)$ 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 $\ell_0^n(\{T_1,T_2\},A)$ with two matrix $T_1,T_2\in\mathcal{T}$ fixed. In what follows, we will consider when $\ell_0^n(\{T_1,T_2\},A)$ would be violated, so as to derive conditions for it to hold. One condition will be the rank constraint (10) (§2.2.1) and the other will be the *quasi-variety* constraint (11) (§2.2.2). These discussions will lead us to an improved uniqueness result for homomorphic sensing over that of Tsakiris (2018; 2020); Tsakiris & Peng (2019) (Theorem 1, §2.2.3).

[6]More technically, $\mathcal{C}$ here is a non-emtpy *Zariski open* subset of $\mathbb{R}^{m\times n}$. It is thus *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.

2.2.1. The rank constraint

By Definition 1, $\ell_0^n(\{T_1,T_2\},A)$ 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[7] linear subspace of $\mathbb{C}^m$, and therefore a complex algebraic variety. The dimension of $\mathcal{Z}_{T_1,T_2}$ influences $\ell_0^n(\mathcal{T},A)$ at least by the following way.

Proposition 4. *Let $\dim(\mathcal{Z}_{T_1,T_2})\le m-n$, $\operatorname{rank}[T_1\ T_2]<2n$. Then $\ell_0^n(\{T_1,T_2\},A)$ is false for a generic $A\in\mathbb{R}^{m\times n}$.*

Proof. If $\operatorname{rank}(T_1A)<n$ then any matrix $A\in\mathbb{R}^{m\times n}$ violates $\ell_0^n(\{T_1\},A)$. Hence let us assume $m\ge n$, $r\ge n$, $\operatorname{rank}(T_1A)=n$, and similarly assume $\operatorname{rank}(T_2A)=n$. But $[T_1\ T_2]\in\mathbb{R}^{r\times 2m}$ has rank smaller than $2n$, so does the $r\times 2n$ matrix $[T_1A\ T_2A]$. As a result, there are non-zero vectors $x_1,x_2\in\mathbb{R}^n$ such that $T_1Ax_1=T_2Ax_2$. Assume for the sake of contradiction that $x_1=x_2$. Then $Ax_1=Ax_2$ and Ax_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 $m-n$, so for a generic $A\in\mathbb{R}^{m\times n}$, the column space intersects $\mathcal{Z}_{T_1,T_2}$ only at zero. This gives $Ax_1=0$ and $x_1=0$, a contradiction. Hence $x_1\ne x_2$. □

The two conditions of Proposition 4 are potential sources for $\ell_0^n(\{T_1,T_2\},A)$ to get violated. To prevent this from happening, it is the insight of Tsakiris (2018; 2020) and Tsakiris & Peng (2019) that considered:

The Rank Constraint

$$\operatorname{rank}(T)\ge 2n,\quad \forall\, T\in\mathcal{T}. \tag{10}$$

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

2.2.2. The quasi-variety constraint

The rank constraint (10), however, does not exclude all possible violations of $\ell_0^n(\{T_1,T_2\},A)$. We next introduce an algebraic-geometric object similar to $\mathcal{Z}_{T_1,T_2}$ that also accounts for $\ell_0^n(\{T_1,T_2\},A)$. For a column vector w of m variables, consider all 2×2 determinants of the $r\times 2$ matrix $[T_1w\ \ T_2w]$. 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 1. *For $m=3$, $r=2$ and*

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

[7]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.

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

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

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*.

It turns out that $\mathcal{U}_{T_1,T_2}$ might be of potential harm to $\ell_0^n(\{T_1,T_2\},A)$ for the following reason.

Proposition 5. *If the intersection of $\mathcal{U}_{T_1,T_2}$ and the column space $R(A)$ of A is not empty, then $\ell_0^n(\{T_1,T_2\},A)$ is false.*

Proof. Let $u \in R(A) \cap \mathcal{U}_{T_1,T_2}$. 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 $u \in R(A)$ we have for some $x \in \mathbb{R}^n$ with $Ax = u$ that $T_1 Ax = T_2 A(\lambda x)$ or $T_1 A(\lambda x) = T_2 Ax$. But $x \neq \lambda x$. We finished the proof. □

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 $\dim(\varnothing) := -1$. Intuitively[8], 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 $\dim(\mathcal{U}_{T_1,T_2}) \leq m - n$ for some $T_1, T_2 \in \mathcal{T} \subset \mathbb{R}^{r\times m}$. Then the column space $R(A)$ of a generic matrix $A \in \mathbb{R}^{m\times n}$ does not intersect $\mathcal{U}_{T_1,T_2}$.*

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 $R(A)$, and, as a consequence, to avoid the violation of $\ell_0^n(\{T_1,T_2\},A)$ in Proposition 5. This justifies the following constraint:

The Quasi-variety Constraint

$$\dim(\mathcal{U}_{T_1,T_2}) \leq m - n, \ \ \forall\, T_1, T_2 \in \mathcal{T}. \qquad (11)$$

[8] More technically, 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 $\mathcal{Q}$. The dimension of any set, e.g., $\mathcal{U}_{T_1,T_2}$, is the dimension of its *closure*, i.e., the smallest algebraic variety which contains it.

Remark 1. *If $\dim(\mathcal{U}_{T_1,T_2}) > m - n$, then a generic $\underline{A} \in \mathbb{C}^{m\times n}$ violates $\ell_0^n(\mathcal{T},\underline{A})$ (proved in the supplementary). In this sense, the quasi-variety constraint (11) is the tightest.*

The rank constraint (10) and quasi-variety constraint (11) are sufficient for $\ell_0^n(\mathcal{T},A)$, as we will see soon.

2.2.3. Unique recovery in 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 (11), we have $\ell_0^n(\mathcal{T},A)$ for a generic $A \in \mathbb{R}^{m\times n}$.*

Constraints (10), (11) of Theorem 1 guarantee the uniqueness of the solution to the homomorphic sensing problem (3). Moreover, since $\ell_0^n(\mathcal{T},A)$ implies $\ell_0^k(\mathcal{T},A)$ and in view of Proposition 2, the two constraints are in fact sufficient for the uniqueness of the solution to (5). Thus Theorem 1 can be thought of as a result for sparse homomorphic sensing.

To compare, note that Tsakiris (2018; 2020); Tsakiris & Peng (2019) used the same rank constraint (10) and a different quasi-variety constraint for $\ell_0^n(\mathcal{T},A)$. We claim that our quasi-variety constraint (11) is simpler and tighter than theirs; recall that (11) is the tightest in the sense of Remark 1. We also claim that the proof techniques for Theorem 1 are quite different from those of Tsakiris (2018; 2020). Moreover, our proof introduces several novel ideas. We will validate the claims in the supplementary.

Summary. In this subsection we considered $\ell_0^n(\mathcal{T},A)$ (Theorem 1), where the major players are $\mathcal{T}$ and (the genericity of) A. We next bring the sparsity back onto the ground and focus on $\ell_0^k(\mathcal{T},A)$ for any $k \leq n$ (§2.3, §2.4).

2.3. The non-linear structure of k-sparse vectors

One major hurdle towards $\ell_0^k(\mathcal{T},A)$ is the non-linearity introduced by the set $\mathcal{K}$ of k-sparse vectors of $\mathbb{R}^n$. Indeed, $\mathcal{K}$ is the *union* of $\binom{n}{k}$ coordinate subspaces of $\mathbb{R}^n$, each spanned by k distinct standard basis vectors of $\mathbb{R}^n$. Moreover, those subspaces might have non-zero intersections; in fact, any two such subspaces intersect at dimension at least $\max\{2k-n, 0\}$. If $\mathcal{T}$ contained only one matrix (the case of compressed sensing), then the effects of those intersections on $\ell_0^k(\mathcal{T},A)$ can be understood via spark or Kruskal rank (recall §1). However, this does not apply to the case where $\mathcal{T}$ has two or more matrices in an obvious way.

Our main insight for overcoming this hurdle is as follows. Let n lines $\ell_1,\dots,\ell_n$ be spanned respectively by n standard basis vectors $e_1,\dots,e_n$ of $\mathbb{R}^n$, i.e., $\ell_i = \mathrm{Span}(e_i)$. Then we can view $\mathcal{K}$ as a *structured composition* of $\ell_1,\dots,\ell_n$ via the following lens. Indeed, any k out of the n lines sum to a k-dimensional subspace, and all possible summations give rise to $\binom{n}{k}$ coordinate subspaces, which in turn compose the union $\mathcal{K}$. Dissecting $\mathcal{K}$ in this way allows us to understand it through the n independent lines rather than through the potentially dependent subspaces. Overall, this view on $\mathcal{K}$, combined with Theorem 1 and tools from algebraic geometry, suffices to approach our main result, presented next.

2.4. Unique recovery in sparse homomorphic sensing

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

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

Then we have $\ell_0^k(\mathcal{T}, A)$ for a generic $A \in \mathbb{R}^{m\times n}$.

The conclusion of Theorem 2, $\ell_0^k(\mathcal{T}, A)$, implies the uniqueness of the solution to (5); recall Proposition 2. From Theorem 1 to Theorem 2, we changed n into k in both the conditions and conclusions. Whenever practical applications promise $k \ll n$, the conditions of Theorem 2 are less demanding for guaranteeing unique ℓ_0 recovery (5). This improvement is because Theorem 2 makes use of the prior knowledge that x^* is k-sparse, while Theorem 1 does not.

Now we derive Proposition 1 from Theorem 2. Since $\mathcal{S}_m$ enjoys more algebraic properties than $\mathcal{T}$, the rank and quasi-variety constraints might be simplified. 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 for the unlabeled sensing problem. 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:

Proposition 7. *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 Theorem 1 with Proposition 7 gives:

Corollary 1. *The following is true for $A \in \mathbb{R}^{m\times n}$ generic:*

$$m \geq 2n \Rightarrow \ell_0^k(\mathcal{S}_m, A).$$

We remark that Corollary 1 is the same as Proposition 1.

3. 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 (6) 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 $\|\overline{y} - T_0Ax\|_2$ over $x \in \mathbb{R}^n$, 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. We note that that 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 $\hat{x}$ could be arbitrarily far from x^*.

We handle this discreteness by identifying "nice" matrices in $\mathcal{T}$; by "nice" we mean a subset $\mathcal{T}_1$ of $\mathcal{T}$ so 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, we set:

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

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 that $\ell_0^n(\mathcal{T}, A)$ holds for some $A \in \mathbb{R}^{m\times n}$ and that $T_0 \in \mathcal{T}_1$. Then $x_0 - x^* = (T_0A)^\dagger\epsilon$ where $x_0 = (T_0A)^\dagger\overline{y}$ and thus $\|x_0 - x^*\|_2 \leq \sigma((T_0A)^\dagger)\|\epsilon\|_2$.*

Under the uniqueness assumption for the homomorphic sensing problem ($\ell_0^n(\mathcal{T}, A)$), 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 (6), 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.

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

$$\|\epsilon\|_2 < \|y\|_2 \Big(1 - \max_{T\in\mathcal{T}\setminus\mathcal{T}_1, x\in\mathbb{R}^n} \frac{y^\top TAx}{\|y\|_2\|TAx\|_2}\Big). \tag{12}$$

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 (12) is strictly smaller than 1. Hence, the right-hand side of (12) is positive.

From Theorem 1 and Propositions 8,9, we are ready to draw a local stability result for noisy homomorphic sensing.

Theorem 3. *Suppose i) $\ell_0^n(\mathcal{T}, A)$ holds true, ii) $\mathcal{T}_1 = \mathcal{T}$ or (12) 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 (12) defines a non-asymptotic regime, where the local stability of $\hat{x}$ is guaranteed (Theorem 3). In particular, if $\mathcal{T} = \mathcal{S}_m$, Theorem 3 is an improvement over the asymptotic result of Unnikrishnan et al. (2018).

4. Unlabeled compressed sensing

We presented that (7) admits a unique solution for $A \in \mathbb{R}^{m\times n}$ generic, as long as $m \geq 2k$ (Proposition 1 and Corollary 1). Here we make an attempt at solving (7).

4.1. 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 following problem

$$\min_{x \in \mathbb{R}^n} \|y - Ax\|_1 \quad \text{s.t.} \quad \|x\|_0 \leq k. \tag{13}$$

The objective function of (13) is about an old problem, *least absolute deviation*, also known as *sparse error correction*; see Kendall (1960); Candes & Tao (2005). The next natural choice is further relaxing the sparsity constraint of (13), so as to arrive at the convex problem[9] of minimizing $\|y - Ax\|_1 + \|x\|_1$ in n variables $x \in \mathbb{R}^n$. But such relaxation does not yield satisfactory performance for our purpose.

We solve (13) 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:

$$x^{(t+1)} \leftarrow \text{Proj}_{\mathcal{K}}\big(x^{(t)} - \mu A^\top \text{sgn}(Ax^{(t+1)} - y)\big) \tag{14}$$

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

$$x_J^{(t+1)} \leftarrow \operatorname*{argmin}_{x \in \mathbb{R}^n} \|y - A_J x\|_1 \tag{15}$$

In (14), we note that i) $A^\top \text{sgn}(Ax^{(t+1)} - y)$ is a subgradient of $\|y - Ax\|_1$ with sgn: $\mathbb{R}^n \to \mathbb{R}^n$ sending $[v_1, \dots, v_n]^\top$ 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) $\text{Proj}_{\mathcal{K}}(\cdot)$ projects a vector to its closest k-sparse counterpart. In (15), 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 (15) they solved a least-squares problem. Another difference is that they run the algorithm by one iteration. We solve (15) by invoking an ADMM algorithm implemented in the `FOM` toolbox of Beck & Guttmann-Beck (2019).

4.2. 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. We have not known obvious baselines or other approaches for the task of interest.

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 (13).[10] 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 $(-\infty, 10^{-10}]$.

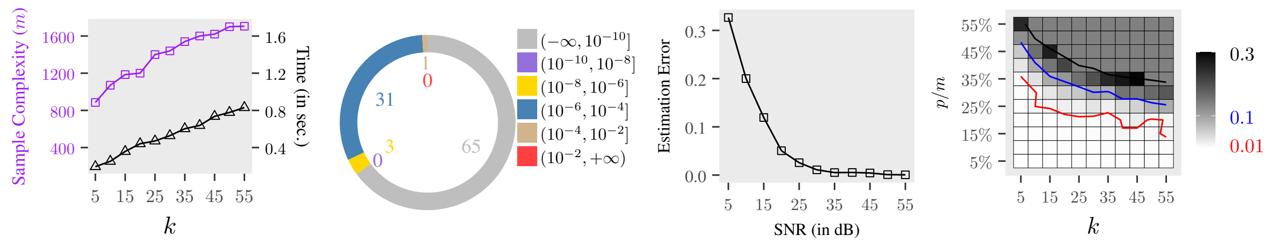
Keeping $m = 1400, n = 2000, k = 25, p = \lfloor 0.2m \rfloor$ fixed, we furthermore evaluated the robustness of the algorithm to noise. We added noise to the measurements y as per the SNR, run the algorithm, and the result was in Figure 1c (100 trials). As the SNR condition improved, the estimation 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.

5. A broader picture

The matrix set $\mathcal{T}$ in (sparse) homomorphic sensing ((3), (5)) provides some flexibility to model other important inverse problems than unlabeled sensing. We next present several

[9]This problem was considered by Wright & Ma (2009) in the context of *dense error correction*, where the authors assumed the ground-truth signal x^* has non-negative entries.

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



(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.

other choices of $\mathcal{T}$ than $\mathcal{S}_m$ that arise from data applications.

Unlabeled sensing with missing entries. An extension of unlabeled sensing was considered by Unnikrishnan et al. (2018), 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 $y = S^*Ax^* \in \mathbb{R}^r$ with some unknown *selection* matrix S^*, i.e., S^* is a permutation matrix with $(m-r)$ rows removed, and ii) one aims to solve

$$y = SAx, \quad S \in \mathcal{S}_{r,m}, \quad x \in \mathbb{R}^n, \tag{16}$$

for x, where $\mathcal{S}_{r,m}$ is the set of $r \times m$ selection matrices. 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 (3). It is the same as (16) except that the positions at which the entries are missing are known.

Real phase retrieval. We also find that the perhaps more familiar problem of *real phase retrieval* (Lv & Sun, 2018) 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

$$y = BAx, \quad B \in \mathcal{B}_m, \quad x \in \mathbb{R}^n, \tag{17}$$

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[11], 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 (17); 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 (16) and (17), 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

$$y = CAx, \quad C \in \mathcal{S}_{r,m}\mathcal{B}_m, \quad x \in \mathbb{R}^n. \tag{18}$$

To summarize, the above problems are concerned with missing correspondences, missing values, sign corruptions, or their combinations thereof, and they are actually of the same type, where the linear measurements Ax^* have further undergone an 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 the supplementary we will show 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.

6. Discussion and future work

On the theoretical ground, we presented conditions guaranteeing the uniqueness for sparse homomorphic sensing, which involves a certain ℓ_0 minimization program and from which a uniqueness result for unlabeled compressed sensing follows. A historical lesson from compressed sensing suggests to find conditions under which the corresponding ℓ_1 relaxation (e.g., (13)) 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. In fact, all one has to do is to work out a high-probability lower bound of the right-hand side of (12).

On the algorithmic front, we initiated a computational investigation into unlabeled compressed sensing. Future improvements might include reducing the sample complexity, tackling the case where more data are shuffled, dispensing with the hyper-parameters, improving the running time, etc..

We presented a broader picture in §5 using the homomorphic sensing framework. Tools from other fields might be key to advancing the research for unlabeled (compressed) sensing.

[11] Both (B^*, x^*) and $(-B^*, -x^*)$ satisfy (17).

References

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.

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.

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

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.

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.

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.

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.

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.

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., 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, 2020.

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

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.

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

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