Sparse and noisy homomorphic sensing: the well-posedness for a class of inverse problems

Anonymous Authors[1]

Abstract

Homomorphic sensing is a recent formulation for a class of inverse problems, its special cases including unlabeled sensing, linear regression without correspondences, missing data recovery, real phase retrieval, mixed linear regression, and more. We provide theoretical conditions, tighter and simpler than that of prior work, under which the homomorphic sensing problem admits a unique solution. Moreover, we show that the conditions are less demanding under a sparsity assumption, and they in particular imply that the associated ℓ_0 optimization problem has a unique minimum. On the other hand, we show that the solution to the homomorphic sensing problem is locally stable under noise. We further apply those results to the above examples to guarantee the uniqueness or local stability, yielding i) known conditions typically obtained in diverse literature via diverse approaches, ii) novel conditions for sparse versions of unlabeled sensing variants, iii) better conditions for the local stability of unlabeled sensing.

December 27, 2020

1. Introduction

Many tasks in machine learning can be formulated as inverse problems. Before solving a given inverse problem, it is important to make sure that the problem is well-posed. The well-posedness, as per Hadamard, consists of three ingredients: i) the existence of a solution, ii) the uniqueness of a solution, and iii) the stability of the solution; see, e.g., (Hadamard, 1902; 1923; Arridge et al., 2019). The existence is usually justified by the belief on the generative procedure of the data and the choice of the model. The stability of the solution under noise is of practical concern, but it might be investigated only when the uniqueness is guaranteed.

[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.

To further appreciate the role of the uniqueness, consider a classic example, linear regression. With $A \in \mathbb{R}^{m\times n}$, $x^* \in \mathbb{R}^n$, $v^* = Ax^*$, one aims to solve the linear equations $v^* = Ax$ for x. If A is of full column rank ($m \geq n$) then a solution x^* is unique — we can find it via Gaussian elimination or else. On the opposite, for example when $m < n$, there are infinitely many solutions and no algorithm is expected to recover x^* — some regularization is needed. Whether A is full column rank or not has thus been a *mental decision boundary* that influences algorithmic choices.

On one side of this boundary where infinitely many solutions are presented, practitioners have assumed x^* as a sparsest one. The fundamental question then recurs: is a sparsest solution x^* unique? The answer has been a major role in theoretical foundations of compressed sensing; see, e.g., (Theodoridis, 2020; Wright & Ma, 2020) for a tutorial. Beyond that, recent years have also witnessed the pursuits of such answers in modern inverse problems, e.g., matrix completion (Tsakiris, 2020b), matrix recovery (Xu, 2018), and deep networks (Puthawala et al., 2020).

This paper focuses on the *homomorphic sensing* problem, recently posed in (Tsakiris, 2018b; 2020a) and in the expository paper (Tsakiris & Peng, 2019), which is concerned with the uniqueness for a class of generalizations of linear regression. There, the linear measurements Ax^* are corrupted by an *unknown* $r \times m$ matrix T^* with $r \leq m$, and we observe $y = T^*Ax^*$. The prior knowledge on T^* is that it comes from a finite set $\mathcal{T}$ of $r \times m$ matrices; $\mathcal{T}$ models the type of corruptions that the measurements Ax^* undergo. In other words, with $\mathcal{T}, A, y$ given and fixed, we have

$$y = TAx, \quad T \in \mathcal{T}, \quad x \in \mathbb{R}^n, \tag{1}$$

and the goal is to solve (1) for x and T. Note that x^* (with T^*) is clearly a solution to (1), the fundamental question is whether x^* is a unique solution.[1]

1.1. Examples of homomorphic sensing

Formulation (1) is able to model a class of inverse problems, depending on what kind of matrix set $\mathcal{T}$ is. For example, if $\mathcal{T}$ contains only one matrix then (1) is just about linear re-

[1]The uniqueness of T^* is not considered in this paper.

gression. Moreover, real-world data applications with missing values, missing correspondences, or missing signs have supplied several other choices of $\mathcal{T}$, and therefore several examples of homomorphic sensing. The applications include, e.g., record linkage for data integration (Fellegi & Sunter, 1969; Lahiri & Larsen, 2005; Slawski & Ben-David, 2019), neuron matching in computational neuroscience (Nguyen et al., 2017; Nejatbakhsh & Varol, 2019), automated translation of medical codes (Shi et al., 2020) and gated flow cytometry (Abid & Zou, 2018; Xie et al., 2020) in biology, signal estimation using distributed sensors (Zhu et al., 2017; Song et al., June 2018; Peng et al., 2019) or from rearranged and erased frame coefficients (Han & Sun, 2014) in communication networks, target localization in signal processing (Wang et al., 2020), multi-target tracking (Ji et al., 2019) and point set registration (Pananjady et al., 2017; Tsakiris & Peng, 2019) in computer vision; see, e.g., (Klibanov et al., 1995; Pananjady et al., 2018; Shi et al., 2020) for more.

As mentioned, these applications have given several examples of homomorphic sensing. We next present them from a unified perspective and thus as part of our contributions.

The recent *unlabeled sensing* problem was introduced in (Unnikrishnan et al., 2015; Unnikrishnan et al., 2018), with a connection to compressed sensing presented, and was then included in (Tsakiris, 2018b; 2020a) and also in (Tsakiris & Peng, 2019) as an example of homomorphic sensing. Here we motivate it from *record linkage* (Fellegi & Sunter, 1969), an application promoted by (Slawski & Ben-David, 2019).

Linking the records collected from different sources has been a routine operation of government agencies like the US Census Bureau, for the purpose of subsequential data analysis (e.g., computing regression coefficients). 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 compute the regression coefficients $x^* \in \mathbb{R}^n$, even without linking two given numerical records, namely the design matrix $A = [a_1, \dots, a_m]^\top \in \mathbb{R}^{m\times n}$ and measurements $y = [y_1, \dots, y_r]^\top \in \mathbb{R}^r$; here we recall $r \leq m$. In this scenario, the correspondences between the entries of y and rows of A are unknown, and there are $(m-r)$ values absent in y. Those imperfections on data might very well be modeled by an unknown $r \times m$ *selection* matrix S^*, i.e., a matrix whose rows are formed by r distinct standard basis vectors of $\mathbb{R}^m$, or equivalently a $m \times m$ permutation matrix with $(m-r)$ rows removed. That is, $y = S^*Ax^*$. The question here — or equivalently that of *unlabeled sensing* (Unnikrishnan et al., 2015; Unnikrishnan et al., 2018) — is whether a solution x^* is unique[2] to the following relation

$$y = SAx, \quad S \in \mathcal{S}_{r,m}, \quad x \in \mathbb{R}^n, \tag{2}$$

where $\mathcal{S}_{r,m}$ is the set of $r \times m$ selection matrices. For example, the following two matrices are elements of $\mathcal{S}_{2,3}$.

$$S_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad S_2 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \tag{3}$$

A special case of unlabeled sensing is where $r = m$, or equivalently where $\mathcal{S}_{r,m}$ becomes the set $\mathcal{S}_m := \mathcal{S}_{r,m}$ of $m \times m$ permutation matrices. This case is known as *linear regression without correspondences* (Hsu et al., 2017; Pananjady et al., 2018; Slawski & Ben-David, 2019; Tsakiris et al., 2020). The next example called *ordered unlabeled sensing* (Haghighatshoar & Caire, 2018) is where the unknown selection S^* is assumed as *order-preserving*, i.e., S^* preserves the relative order of the rows of A that it selects. For example, the above S_1 is order-preserving but S_2 is not.

We found the problem of *missing data recovery* (Zhang, 2006; Liu et al., 2017; Liu et al., 2019) or *signal recovery with erasures at known locations* (Han & Sun, 2014) as the next example. Describing it in mathematical terms, we aim to recover x^* from $y' = O^*Ax^* \in \mathbb{R}^m$ and A, where O^* is an unknown coordinate projection $O^* \in \mathbb{R}^{m\times m}$, i.e., a diagonal matrix with r ones and $m-r$ zeros on the diagonal. Here, the major difference from unlabeled sensing is that the positions at which the values are missing are known.

We also found that the 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; see (Grohs et al., 2020) for a vivid account. In a mathematical formulation of this problem, we have the relation

$$y = BAx, \quad B \in \mathcal{B}_m, \quad x \in \mathbb{R}^n, \tag{4}$$

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[3], the second best 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 (4); see, e.g., (Chen et al., 2019; Klusowski et al., 2019) where the connection between the two problems was discussed.

An interesting generalization which we call *unsigned unlabeled sensing* was explored in (Lv & Sun, 2018) and is a combination of real phase retrieval and unlabeled sensing. 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

$$y = CAx, \quad C \in \mathcal{S}_{r,m}\mathcal{B}_m, \quad x \in \mathbb{R}^n. \tag{5}$$

[2] Unique and approximate recovery of S^* when $r = m$ was considered in, e.g., (Pananjady et al., 2018; Zhang et al., 2019b).

[3] Both (B^*, x^*) and $(-B^*, -x^*)$ satisfy (4).

1.2. Contributions of this paper

In (Tsakiris, 2018b; 2020a) it was proved that (1) admits a unique solution for a *generic* matrix A of size $m \times n$, whenever i) every matrix of $\mathcal{T}$ has rank at least $2n$, ii) the *dimension* of a specific set $\mathcal{U} \subset \mathbb{C}^m$ depending on $\mathcal{T}$ is at most $m - n$.[4] Based on their work, we make several improvements (§2). First, in Theorem 1, we give tighter and simpler conditions than those of (Tsakiris, 2018b; 2020a) which still allow (1) to have a unique solution.

(*Sparse homomorphic sensing*) Moreover, if x^* is assumed as sparse, our conditions for the uniqueness of (1) can be less demanding (Theorem 2). As a direct consequence (Proposition 6), under our conditions, the solution to the following optimization problem (6) is unique, and is necessarily x^*.

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

(*Noisy homomorphic sensing*) Next, we consider a noisy formulation of (1), where the measurements $\overline{y} = y + \epsilon = T^*Ax^* + \epsilon$ are corrupted by additive noise ϵ. We show in Theorem 3 that the following estimator (7) produces a locally stable solution $\hat{x}$, as long as $\|\epsilon\|_2$ is *sufficiently small* and that (1) admits a unique solution.

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

In (Tsakiris, 2018b; 2020a) and (Tsakiris & Peng, 2019) the results were applied to unlabeled sensing (e.g., by setting $\mathcal{T} = \mathcal{S}_{r,m}$). In §3 we apply our results to the examples of §1.1. In so doing, in a unified way, we obtain that, with *at least* $2n$ samples, (2), (4), (5) all admit unique (or up to sign) solutions for a generic $A \in \mathbb{R}^{m \times n}$. As will be detailed in §3, these results have been shown in several prior works, e.g., (Balan et al., 2006; Unnikrishnan et al., 2018; Han et al., 2018; Lv & Sun, 2018; Dokmanic, 2019) — motivated by different applications, stated in different mathematical languages, and proved using different approaches.

Moreover, if x^* is *sufficiently sparse* and if $\mathcal{T}$ is set to be $\mathcal{S}_m, \mathcal{S}_{r,m}, \mathcal{B}_m$, or $\mathcal{S}_{r,m}\mathcal{B}_m$, then (6) has a unique (or up to sign) sparsest solution for $A \in \mathbb{R}^{m \times n}$ generic, which is necessarily x^* (or $\pm x^*$). For sparse phase retrieval ($\mathcal{B}_m$), this result was proved independently by (Wang & Xu, 2014) and (Akçakaya & Tarokh, 2013). For sparse versions of unlabeled sensing variants ($\mathcal{S}_m, \mathcal{S}_{r,m}$, and $\mathcal{S}_{r,m}\mathcal{B}_m$), these results are novel to the best of our knowledge.

Next, when $\mathcal{T}$ is set to be $\mathcal{S}_{r,m}$, we get a SNR condition which defines a non-asymptotic regime where the local stability of $\hat{x}$ is promised, an improvement upon the asympotic result of (Unnikrishnan et al., 2018).

[4] We review the notions of "generic" and "dimension" in §2. It is now safe to think of a generic A as "random", and the dimension of a set as a number which measures how large the set is.

2. Theory

2.1. Homomorphic sensing

The uniqueness of a solution to (1) involves the measurements $y \in \mathbb{R}^r$, the design matrix $A \in \mathbb{R}^{m \times n}$, and the finite set $\mathcal{T} \subset \mathbb{R}^{r \times m}$ of matrices. Note that $y = T^*Ax^*$ depends on an arbitrary x^*, and that we want to guarantee the unique recovery of all possible $x^* \in \mathbb{R}^n$ (a sparse x^* is considered in §2.2). This motivates the following definition.

Definition 1 (hsp). *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 x_1, x_2 we have*

$$T_1Ax_1 = T_2Ax_2 \;\; \Rightarrow \;\; x_1 = x_2, \tag{8}$$

then we say that $\mathcal{T}$ and A have the homomorphic sensing property, written as $\mathrm{hsp}(\mathcal{T}, A)$. *If $T_1Ax_1 = T_2Ax_2$ only implies $x_1 = \pm x_2$, then we write* $\mathrm{hsp}_{\pm}(\mathcal{T}, A)$.

Definition 1 (hsp) encodes our desire: $\mathrm{hsp}(\mathcal{T}, A)$ holds if and only if (1) admits a unique solution for any $x^* \in \mathbb{R}^n$. Similarly, $\mathrm{hsp}_{\pm}(\mathcal{T}, A)$ is equivalent to unique up to sign recovery. Our discussion will primarily focus on $\mathrm{hsp}(\mathcal{T}, A)$, with a notice that similar results also hold for $\mathrm{hsp}_{\pm}(\mathcal{T}, A)$. We use hsp directly when there is no chance for confusion.

Clearly, $\mathrm{hsp}(\mathcal{T}, A)$ would rest on the nature of A and $\mathcal{T}$. Next we determine what kinds of A and $\mathcal{T}$ to work with, respectively in §2.1.1 and §2.1.2. This will shed light on the conditions that allow $\mathrm{hsp}(\mathcal{T}, A)$ to hold (§2.1.3).

2.1.1. Which design matrix?

Note that $A \in \mathbb{R}^{m \times n}$ might be arbitrary. If A is not of full column rank, then (1) has infinitely many solutions — regardless of $\mathcal{T}$. Thus, $\mathrm{hsp}(\mathcal{T}, A)$ can not hold for any A. The second best to hope is that $\mathrm{hsp}(\mathcal{T}, A)$ holds for a *generic* $A \in \mathbb{R}^{m \times n}$. We next explain the reason of hoping so by reviewing the algebraic-geometric notion, "generic".

The central object in algebraic geometry is a complex (*resp.* real) *algebraic variety*. To start with, let us define the complex (*resp.* real) *hypersurface* $\mathcal{H}$ to be 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; in other words, $\mathcal{H} = \{z : p(z) = 0\}$. An algebraic variety is the intersection of finitely many hypersurfaces, that is, the common roots of finitely many 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. The real hypersurface $\mathbb{R}^{m \times n}$, or the real algebraic variety defined by the zero polynomial, is of our interest. 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 $\mathbb{R}^{m \times n}$

satisfying this property. Intuitively[5], since $\mathcal{P}$ is the intersection of finitely many hypersurfaces, a matrix randomly chosen from $\mathbb{R}^{m\times n}$ will land itself in $\mathcal{C}$, with probability 1 — for the same reason that the intersection of (finitely many) 2D planes of $\mathbb{R}^3$ has measure 0. To be more specific, that a generic $A\in\mathbb{R}^{m\times n}$ satisfies $\mathrm{hsp}(\mathcal{T},A)$ implies that $\mathrm{hsp}(\mathcal{T},A)$ holds with probability 1 if the entries of A are sampled independently at random according to some continuous probability distribution. Our purpose is to show that a generic $A\in\mathbb{R}^{m\times n}$ satisfies $\mathrm{hsp}(\mathcal{T},A)$.

2.1.2. Which matrix set?

Note that $\mathcal{T}\subset\mathbb{R}^{r\times m}$ can be arbitrary. Unlike (2)-(5) where the combinatorial structures (e.g., $\mathcal{S}_{r,m},\mathcal{B}_m$) yielded insights that assist analysis (Balan et al., 2006; Unnikrishnan et al., 2018; Han et al., 2018; Lv & Sun, 2018), formulation (1) only gives that $\mathcal{T}$ is a finite set. How can we determine what kind of $\mathcal{T}$ would satisfy $\mathrm{hsp}(\mathcal{T},A)$ for a generic A?

A very easy case is when $\mathcal{T}=\{T\}$ contains only one matrix (linear regression); then hsp (8) would require to consider only the null space of TA. A less ideal case is when $\mathcal{T}=\{T_1,T_2\}$ with T_1,T_2 invertible ($r=m$); $\mathrm{hsp}(\{T_1,T_2\},A)$ can be understood via analyzing the Jordan canonical form of $T_1^{-1}T_2$ (Tsakiris, 2018b; 2020a). However, it is *not uncommon* to have the case $r<m$, as is usual in unlabeled sensing (recall (3)); this renders any matrix non-invertible and a lemma of (Tsakiris, 2018b; 2020a) not directly applicable. In the presence of this hurdle, we proceed in the absence of the invertibility assumption.

We investigate into how $\mathcal{T}$ violates $\mathrm{hsp}(\mathcal{T},A)$. Let $T_1,T_2\in\mathcal{T}$. By Definition 1, $\mathrm{hsp}(\mathcal{T},A)$ is related to how the column spaces of T_1A and T_2A interact. It is thus natural to consider how T_1 and T_2 interact, which can be seen from the set

$$\mathcal{Z}_{T_1,T_2}:=\{w\in\mathbb{C}^m: T_1w=T_2w\}. \tag{9}$$

Note that $\mathcal{Z}_{T_1,T_2}$ is a complex linear subspace of $\mathbb{C}^m$, the null space of T_1-T_2, and therefore a complex algebraic variety. If $T_1=S_1$ and $T_2=S_2$ of (3), then $\mathcal{Z}_{T_1,T_2}$ is a line of $\mathbb{C}^3$ defined by $\{w\in\mathbb{C}^3: w_1=w_2=w_3\}$.

The importance of $\mathcal{Z}_{T_1,T_2}$ is in that it captures the *similarity* of T_1 and T_2. For example, if $\dim(\mathcal{Z}_{T_1,T_2})=m$ then $\mathcal{Z}_{T_1,T_2}=\mathbb{C}^m$ and so T_1 is the same as T_2. Moreover, if $r=m$ and T_2 is the identity matrix, then $\mathcal{Z}_{T_1,T_2}$ is the set of all eigenvectors of T_1 corresponding to eigenvalue 1; the larger the geometric multiplicity of eigenvalue 1 is, the more similar T_1 is to the identity matrix. More generally, as the dimension of $\mathcal{Z}_{T_1,T_2}$ goes larger, it is more likely for T_1 and T_2 to send a $w\in\mathbb{C}^m$ to the same destination, $T_1w=T_2w$.

[5]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.

Finally, we remark that, even if T_1 and T_2 are real matrices, we define $\mathcal{Z}_{T_1,T_2}$ as a complex object. This is because we can conveniently discuss its real counterpart ($\mathcal{Z}_{T_1,T_2}\cap\mathbb{R}^m$) whenever needed, and is on account of technical reasons.

The following is partly due to the dissimilarity of T_1,T_2.

Proposition 1. *Suppose for some $T_1,T_2\in\mathcal{T}\subset\mathbb{R}^{r\times m}$ that $\mathrm{rank}[T_1\ T_2]<2n$ and $\dim(\mathcal{Z}_{T_1,T_2})\le m-n$. Then a generic $A\in\mathbb{R}^{m\times n}$ violates $\mathrm{hsp}(\mathcal{T},A)$.*

To understand why the condition $\mathrm{rank}[T_1\ T_2]<2n$ of Proposition 1 is the potential source of violating $\mathrm{hsp}(\mathcal{T},A)$, consider the $r\times 2n$ matrix $[T_1A\ T_2A]$. Note that the property $\mathrm{hsp}(\{T_1,T_2\},A)$ gets violated only when this matrix has rank smaller than $2n$.[6] The condition $\mathrm{rank}[T_1\ T_2]<2n$, which implies $\mathrm{rank}[T_1A\ T_2A]<2n$, thus gives a chance for $\mathrm{hsp}(\{T_1,T_2\},A)$ to be violated. This chance then becomes a truth, as per Proposition 1, if T_1 and T_2 are not similar in the sense that $\dim(\mathcal{Z}_{T_1,T_2})\le m-n$.

How to prevent the violation of hsp in Proposition 1 from happening? It is the insight of (Tsakiris, 2018b; 2020a) and (Tsakiris & Peng, 2019) that considered:

The Rank Constraint

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

The rank constraint (10) to put on matrices of $\mathcal{T}$ ensures $\mathrm{rank}[T_1\ T_2]>2n$ for any $T_1,T_2\in\mathcal{T}$, so that the above violation would never happen. This constraint is perhaps the simplest, as it does not involve any interaction of T_1,T_2.

The linear-algebraic rank constraint (10), however, does not exclude all possible violations of hsp. We next introduce an algebraic-geometric object that also accounts for the similarity of T_1 and T_2 — but in a different if not converse way. For a column vector w of m variables, consider all 2×2 determinants of the $r\times 2$ matrix $[T_1w\ \ T_2w]$. Since each determinant is a quadratic polynomial in entries of w, we obtain $\binom{r}{2}$ polynomials in total. Let $\mathcal{Y}_{T_1,T_2}\subset\mathbb{C}^m$ be the complex algebraic variety defined by those polynomials.

Example 1. *For $m=3$, $r=2$ and*

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

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.

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

If $T_1=S_1$ and $T_2=S_2$ of (3), then $\mathcal{Y}_{T_1,T_2}$ is a hypersurface of $\mathbb{C}^3$ defined by the complex roots of $w_1w_2-w_3^2=0$.

[6]If $[T_1A\ T_2A]$ is of full rank $2n$ then $T_1Ax_1=T_2Ax_2$ with any $x_1,x_2\in\mathbb{R}^n$ and any $T_1,T_2\in\mathcal{T}$ implies $x_1=x_2=0$.

Alternatively, we might describe $\mathcal{Y}_{T_1,T_2}$ as the set of vectors u's of $\mathbb{C}^m$ such that $[T_1u \;\; T_2u]$ has rank at most one, i.e.,

$$\mathcal{Y}_{T_1,T_2} = \{u \in \mathbb{C}^m : \text{rank}[T_1u \;\; T_2u] \leq 1\}.$$

Equivalently, $\mathcal{Y}$ consists of all vectors u's of $\mathbb{C}^m$ for which T_1u and T_2u are linearly dependent. Observe that $\mathcal{Z}_{T_1,T_2}$ of (9) is a subvariety of $\mathcal{Y}_{T_1,T_2}$. Define the 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*.

It can be verified that, for every $u \in \mathcal{U}_{T_1,T_2}$, we have $T_1u = \lambda T_2u$ or $\lambda T_1u = T_2u$, for some $\lambda \in \mathbb{C}$ with $\lambda \neq 1$. In words, T_1 and T_2 send u to the same destination, e.g., $\lambda T_1u = T_2u$, up to a multiplicative factor $\lambda \neq 1$. This is a potential harm to hsp for the following reason.

Proposition 2. *Consider $T_1, T_2 \in \mathcal{T} \subset \mathbb{R}^{r\times m}$ and any $A \in \mathbb{R}^{m\times n}$. If the intersection of $\mathcal{U}_{T_1,T_2}$ and the column space $R(A)$ of A is not empty, then $\text{hsp}(\mathcal{T}, A)$ is violated.*

Proof. Let $u \in \mathbb{R}^m$ in the intersection. Then T_1u and T_2u are linearly dependent. So there is some $\lambda \in \mathbb{R}$ such that $T_1u = \lambda T_2u$ or $\lambda T_1u = T_2u$. Since $u \in \mathcal{U}_{T_1,T_2}$, definition (11) implies that $u \neq 0$ and $\lambda \neq 1$. Since $u \in R(A)$ we have for some $x \in \mathbb{R}^n$ with $Ax = u$ that $T_1Ax = T_2A(\lambda x)$ or $T_1A(\lambda x) = T_2Ax$. But $x \neq \lambda x$. □

Proposition 2 presents a violation of hsp, where we began to consider the interaction of T_1 and T_2, as encoded in $\mathcal{U}_{T_1,T_2}$, and the interaction of $\mathcal{U}_{T_1,T_2}$ and $R(A)$. As a result, 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* $\dim(\mathcal{U}_{T_1,T_2})$, an algebraic-geometric notion that assigns to each subset of $\mathbb{C}^m$ a non-negative integer with the convention $\dim(\varnothing) := -1$. Intuitively[7], to say that $\dim(\mathcal{U}_{T_1,T_2})$ is large is to say that $\mathcal{U}_{T_1,T_2}$ is large, and a larger $\mathcal{U}_{T_1,T_2}$ implies a higher risk of $\mathcal{U}_{T_1,T_2}$ intersecting $R(A)$, as formalized below.

Proposition 3. *Let $T_1, T_2 \in \mathcal{T}$. Whenever $\dim(\mathcal{U}_{T_1,T_2}) > m - n \geq 0$, for a generic $\underline{A} \in \mathbb{C}^{m\times n}$ the intersection of $\mathcal{U}_{T_1,T_2}$ and the column space $R(\underline{A})$ of $\underline{A}$ is not empty.*

To explain Proposition 3, recall that, in $\mathbb{R}^m$, any linear subspace of dimension n (e.g., $R(A)$) has non-trivial intersection with another subspace of dimension larger than $m - n$. Proposition 3 is of the similar flavor except that $\mathcal{U}_{T_1,T_2}$ is a quasi-variety, to which this linear-algebraic argument can not be applied. To overcome this difficulty[8], we used tools from commutative algebra; see the supplementary.

Conversely, if $\dim(\mathcal{U}_{T_1,T_2}) \leq m - n$, could the intersection be empty? Linear-algebraic intuition suggests that, in $\mathbb{R}^m$, the column space of a generic matrix $A \in \mathbb{R}^{m\times n}$ intersects a fixed linear subspace of dimension at most $m - n$ only at zero. Does this remain true if replacing the fixed real subspace by a fixed complex quasi-variety (e.g., $\mathcal{U}_{T_1,T_2}$)? We answer this question in the next proposition.

Proposition 4. *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 4, enforcing $\mathcal{U}_{T_1,T_2}$ to have small dimension is 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 hsp in Proposition 2. This justifies:

The Quasi-variety Constraint

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

The quasi-variety constraint (12) and the rank constraint (10), once combined together, are able to sidestep all possible violations of hsp, as we will soon see (§2.1.3).

2.1.3. Recovery guarantees

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

Theorem 1 portrays a mental decision boundary (as elaborated in §1) for homomorphic sensing, where constraints (10) and (12) provide a mathematical awareness about the uniqueness of (1). It encourages pragmatic practitioners to make their algorithmic choices that respect (10) and (12).[9]

To compare, note that (Tsakiris, 2018b; 2020a; Tsakiris & Peng, 2019) used the same rank constraint (10) and a different and more complicated quasi-variety constraint. Since our quasi-variety constraint (12) is the tightest in the sense of Proposition 3, it is proven tighter than that of (Tsakiris, 2018b; 2020a; Tsakiris & Peng, 2019). We argue that the proof techniques for Theorem 1 are quite different from those of (Tsakiris, 2018b; 2020a) and themselves non-trivial. A detailed comparison and proofs are in the supplementary.

Finally, we present the flexibility of Theorem 1 by extending it for $\text{hsp}_{\pm}(\mathcal{T}, A)$; recall that $\text{hsp}_{\pm}(\mathcal{T}, A)$ comes into picture whenever unique recovery is not possible (e.g., in

[7] 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.

[8] Proposition 3 is for a $m \times n$ generic complex matrix. We conjecture that the same holds true for a $m \times n$ generic real matrix.

[9] We discuss how constraint (12) behaves in applications (§3).

real phase retrieval $\mathcal{T} = \mathcal{B}_m$). Defining $\mathcal{Z}^{\pm}_{T_1,T_2} := \{w \in \mathbb{C}^m : T_1 w = T_2 w \text{ or } T_1 w = -T_2 w\}$ as the union of two linear subspaces, it is the quasi-variety

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

that replaces the role of $\mathcal{U}_{T_1,T_2}$ to control $\mathrm{hsp}_{\pm}(\mathcal{T}, A)$.

Proposition 5. *Suppose that the rank constraint* (10) *holds. Then* $\mathrm{hsp}_{\pm}(\mathcal{T}, A)$ *holds for a generic* $A \in \mathbb{R}^{m\times n}$ *whenever*

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

2.2. Sparse homomorphic sensing

Tracing the trajectory traversed in compressed sensing (see also §1), here we assume that the ground-truth $x^* \in \mathbb{R}^n$ is k-sparse, i.e., it has at most k non-zero entries. The impact of this assumption is the shrinkage of the searching space of solutions to (1): now we need only to consider the set of k-sparse vectors of $\mathbb{R}^n$. Thus we revise Definition 1 of hsp.

Definition 2 (sparse-hsp)**.** *Given* $A \in \mathbb{R}^{m\times n}$ *and* $\mathcal{T} \subset \mathbb{R}^{r\times m}$*, if* (8) *holds for any matrices* $T_1, T_2 \in \mathcal{T}$ *and any* k*-sparse vectors* $x_1, x_2 \in \mathbb{R}^n$*, then we say that* $\mathcal{T}$ *and* A *have the sparse homomorphic sensing property, written as* $\text{sparse-hsp}(\mathcal{T}, A)$*. If* $T_1 A x_1 = T_2 A x_2$ *only implies* $x_1 = \pm x_2$*, then we write* $\text{sparse-hsp}_{\pm}(\mathcal{T}, A)$.

Clearly, $\mathrm{hsp}(\mathcal{T}, A)$ implies $\text{sparse-hsp}(\mathcal{T}, A)$, and so Theorem 1 applies directly for $\text{sparse-hsp}(\mathcal{T}, A)$. Our purpose here is to give a tighter condition than that of Theorem 1 for sparse-hsp. The major hurdle towards this end 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 $\mathcal{T}$ contains only one matrix (the case of sparse linear regression), then the effects of those intersections on sparse-hsp can be understood via *spark* ((Donoho & Elad, 2003)) or *Kruskal rank* ((Kruskal, 1977)), two widely used notations in compressed sensing. However, this does not apply to the case where $\mathcal{T}$ contains two or more matrices, at least in an obvious way.

Fortunately, we have known that the rank and quasi-variety constraints (10), (12) are key to approaching hsp for a finite set $\mathcal{T}$ (Theorem 1) and that the crucial idea in the quasi-variety constraint is to prevent $\mathcal{U}_{T_1,T_2}$ from intersecting the column space of A, i.e., the image of $\mathbb{R}^n$ under the linear map $\tau_A : \mathbb{R}^n \to \mathbb{R}^m$ which sends $x \in \mathbb{R}^n$ to Ax. This motivates us to analyze whether $\mathcal{U}_{T_1,T_2}$ intersects $\tau_A(\mathcal{K})$.

Our main insight for such analysis is viewing $\mathcal{K}$ as a *structured composition* of the n lines $\ell_1, \dots, \ell_n$ spanned respectively by n standard basis vectors $e_1, \dots, e_n$ of $\mathbb{R}^n$, i.e., $\ell_i = \mathrm{Span}(e_i)$, via the following lens. Indeed, any k 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 the converse way allows us to understand it through the n independent lines rather than through the potentially dependent subspaces. Though independent, the n lines are *far from generic*: each line ℓ_i is a hypersurface of $\mathbb{R}^n$ defined by the polynomial $x_i - 1$. However, a generic $A \in \mathbb{R}^m$ is able to transform ℓ_i into a generic line $\tau_A(\ell_i)$. Consequently, our analysis of the intersection of $\mathcal{U}_{T_1,T_2}$ and $\tau_A(\mathcal{K})$ is based on the n independent and generic lines $\tau_A(\ell_1), \dots, \tau_A(\ell_n)$ which compose $\tau_A(\mathcal{K})$. Overall, our insight, combined with techniques in algebraic geometry and based on Theorem 1, gives:

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

$$\dim(\mathcal{U}_{T_1,T_2}) \le m - k, \ \ \forall\, T_1, T_2 \in \mathcal{T}. \tag{14}$$

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

From Theorem 1 to Theorem 2, the change is that we replaced n by k in the two conditions, and replaced hsp by sparse-hsp in the conclusions. Whenever practical applications promise $k \ll n$, the conditions of Theorem 2 are less demanding for guaranteeing unique sparse recovery.

An immediate consequence of sparse-hsp is that, whenever $\text{sparse-hsp}(\mathcal{T}, A)$ is true, (1) has a unique k-sparse solution for any k-sparse $x^* \in \mathbb{R}^n$. In particular, we have the following result in terms of optimization.

Proposition 6. *If* $\text{sparse-hsp}(\mathcal{T}, A)$ *is true, then for any* k*-sparse vector* $x^* \in \mathbb{R}^n$*, the* ℓ_0 *minimization problem* (6) *has a unique optimal solution, which is necessarily* x^*.

Proof. Note that $y = T^* A x^*$ for some $T^* \in \mathcal{T}$. Let x^+ be an optimal solution, so $\|x^+\|_0 \le \|x^*\|_0$. Then x^+ is k-sparse and $y = T^+ A x^+$ for some $T^+ \in \mathcal{T}$. By $\text{sparse-hsp}(\mathcal{T}, A)$ we know that $x^* = x^+$. □

Remark 1. *The* ℓ_0 *norm minimization problem* (6) *is NP-hard. The exploration of its convex relaxation for specific types of* $\mathcal{T}$ *(e.g.,* $\mathcal{S}_{r,m}, \mathcal{S}_m$*) is left as future work.*

Remark 2. *It is easy to extend Theorem 2 for guaranteeing* $\text{sparse-hsp}_{\pm}(\mathcal{T}, A)$*; see also Corollary 5.*

Remark 3. *The dimensions* $\dim(\mathcal{U}_{T_1,T_2})$, $\dim(\mathcal{U}^{\pm}_{T_1,T_2})$ *that appeared in the quasi-variety constraints can be computed using algebraic geometry software, e.g.,* `Macaulay2`.

2.3. Noisy homomorphic sensing

We consider the homomorphic sensing problem in the presence of additive noise $\epsilon \in \mathbb{R}^r$. This gives us the relation:

$$\overline{y} = TAx + \epsilon, \quad T \in \mathcal{T}, \quad x \in \mathbb{R}^n, \tag{15}$$

Compared to (1), the change is that we are instead given noisy measurements $\overline{y}$. The questions here are i) how we can estimate x^* and ii) how good the estimation is.

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

Question ii), or more specifically whether $\hat{x}$ is close to x, is our focus in the paper. We first notice that the problem is naturally *discrete* for the following reason. For arbitrary noise ϵ, the optimal $\hat{T}$ can be any matrix of the discrete set $\mathcal{T}$. Since $\mathcal{T}$ is also arbitrary, the corresponding least-squares solution $\hat{x}$ could be arbitrarily far from x^*.

We handle this discreteness by identifying "good" matrices in $\mathcal{T}$; by "good" 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, a concrete definition of $\mathcal{T}_1$ is given as:

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

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

Proposition 7. *Assume that* $\mathrm{hsp}(\mathcal{T}, A)$ *holds for some* $A \in \mathbb{R}^{m\times n}$ *and that* $\hat{T} \in \mathcal{T}_1$. *Then* $\hat{x} - x^* = (\hat{T}A)^\dagger \epsilon$. *In particular we have* $\|\hat{x} - x^*\|_2 \le \sigma((\hat{T}A)^\dagger)\|\epsilon\|_2$.

Proof. Since $\hat{T} \in \mathcal{T}_1$, we get $y = \hat{T}Ax_0$ for some $x_0 \in \mathbb{R}^n$. But $y = T^*Ax^*$ and $\mathrm{hsp}(\mathcal{T}, A)$ holds, so it must be that $x_0 = x^*$. This implies $y = \hat{T}Ax^*$. Noticing

$$\hat{x} = (\hat{T}A)^\dagger(y + \epsilon) = (\hat{T}A)^\dagger \hat{T}Ax^* + (\hat{T}A)^\dagger \epsilon,$$

which implies $\hat{x} - x^* = (\hat{T}A)^\dagger \epsilon$, we finished the proof. □

Proposition 7 provides in particular an upper bound of $\|\hat{x} - x^*\|_2$ in terms of noise and data, under two conditions. It first requires hsp to hold; a natural condition for studying the (local) stability, given the uniqueness. It also demands $\hat{T} \in \mathcal{T}_1$. 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 that the event $\hat{T} \in \mathcal{T}$ is always true in presence of sufficiently small noise.

Proposition 8. *We have* $\hat{T} \in \mathcal{T}_1$ *whenever* $\mathcal{T}_1 = \mathcal{T}$ *or*

$$\|\epsilon\|_2 \le \|y\|_2 \left(1 - \max_{T\in\mathcal{T}\setminus\mathcal{T}_1, x\in\mathbb{R}^n} \frac{y^\top TAx}{\|y\|_2\|TAx\|_2}\right). \quad (16)$$

Since for every $T \in \mathcal{T}\setminus\mathcal{T}_1$, the range space of TA does not contain y, the maximization term of (16) is strictly smaller than 1. Hence, the right-hand side of (16) is positive. Proposition 8 implies that $\hat{T}$ is "good" under small noise.

From Theorem 1 and Propositions 7,8, we are ready to draw a local stability result for noisy homomorphic sensing.

Theorem 3. *Suppose i)* $A \in \mathbb{R}^{m\times n}$ *satisfies* $\mathrm{hsp}(\mathcal{T}, A)$, *ii)* $\mathcal{T}_1 = \mathcal{T}$ *or* (16) *holds, then* $\hat{x} - x^* = (\hat{T}A)^\dagger \epsilon$, *and we have in particular* $\|\hat{x} - x^*\|_2 \le \sigma((\hat{T}A)^\dagger)\|\epsilon\|_2$.

Remark 4. *Theorem 3 reflects our belief of (local) stability in presence of the uniqueness (the 1st paragraph of §1).*

3. Applications and related works

We now consider applying Theorems 1-3 to the problems mentioned in §1.1, namely linear regression without correspondences ($\mathcal{S}_m$), unlabeled sensing ($\mathcal{S}_{r,m}$), real phase retrieval ($\mathcal{B}_m$), and unsigned unlabeled sensing ($\mathcal{S}_{r,m}\mathcal{B}_m$).

Since $\mathcal{T}$ now has more structure (e.g., when set to $\mathcal{S}_m$), the rank and quasi-variety constraint might be simplified. Indeed, every permutation matrix of $\mathcal{S}_m$ and every sign matrix of $\mathcal{B}_m$ have rank m, and every selection matrix of $\mathcal{S}_{r,m}$ and every matrix of $\mathcal{S}_{r,m}\mathcal{B}_m$ have rank r. As a result, the rank constraint (10) becomes $m \ge 2n$ or $r \ge 2n$, a requirement on the number of samples. Moreover, inspired by (Tsakiris, 2018b; 2020a; Tsakiris & Peng, 2019), an interesting discovery is that, whenever the rank constraint is fulfilled, the quasi-variety constraint (12), (13) is automatically satisfied:

Proposition 9. *Let* $\Pi_1, \Pi_2 \in \mathcal{S}_m$, $S_1, S_2 \in \mathcal{S}_{r,m}$, *and* $B_1, B_2 \in \mathcal{B}_m$ *be permutation matrices, selection matrices, and sign matrices, respectively.*

- $m \ge 2n \Rightarrow \dim(\mathcal{U}_{\Pi_1,\Pi_2}) \le m - n$.
- $r \ge 2n \Rightarrow \dim(\mathcal{U}_{S_1,S_2}) \le m - n$.
- $m \ge 2n \Rightarrow \dim(\mathcal{U}^{\pm}_{B_1,B_2}) \le m - n$.
- $r \ge 2n \Rightarrow \dim(\mathcal{U}^{\pm}_{S_1B_1,S_2B_2}) \le m - n$.

Based on Proposition 9, we are ready to apply Theorem 1-3 and discuss related works in §3.1-§3.3, respectively.

3.1. The number of samples for hsp

Setting $\mathcal{T}$ to be $\mathcal{S}_m, \mathcal{S}_{r,m}, \mathcal{B}_m$, or $\mathcal{S}_{r,m}\mathcal{B}_m$, combining Proposition 9 with Theorem 1, we get the following series of corollaries, which hold for a generic $A \in \mathbb{R}^{m\times n}$.

Corollary 1. $m \ge 2n \Rightarrow \mathrm{hsp}(\mathcal{S}_m, A)$ *(Unnikrishnan et al., 2018; Han et al., 2018; Dokmanic, 2019; Tsakiris & Peng, 2019).*

Corollary 2. $r \ge 2n \Rightarrow \mathrm{hsp}(\mathcal{S}_{r,m}, A)$ *(Unnikrishnan et al., 2018; Han et al., 2018; Tsakiris & Peng, 2019).*

Corollary 3. $m \ge 2n \Rightarrow \mathrm{hsp}_{\pm}(\mathcal{B}_m, A)$ *(Balan et al., 2006; Dokmanic, 2019).*

Corollary 4. $r \geq 2n \Rightarrow \text{hsp}_{\pm}(\mathcal{S}_{r,m}\mathcal{B}_m, A)$ *(Lv & Sun, 2018; Tsakiris, 2018a).*

Corollary 1 is a special case of Corollary 2; the latter was proved by (Unnikrishnan et al., 2018) using a combinatorial argument in the context of signal processing. Since then a series of algorithms that operate (explicitly or implicitly) in the well-posed regime $r \geq 2n$ have followed; see, e.g., (Zhang et al., 2019a; Slawski et al., 2019; Tsakiris & Peng, 2019; Zhang & Li, 2020; Tsakiris et al., 2020; Peng & Tsakiris, 2020; Slawski et al., 2020; Wang et al., 2020).

Corollary 2 was also independently proved by (Han et al., 2018) in the context of harmonic analysis, using a different algebraic-combinatorial approach. The work of (Han et al., 2018) motivated (Lv & Sun, 2018) to prove Corollary 4.

Corollary 3 is a consequence of a result of (Balan et al., 2006), whose proof was stated in a frame-theoretical language. Note that the matrices of $\mathcal{S}_m$ and $\mathcal{B}_m$ are invertible and diagonalizable, so the result of (Dokmanic, 2019) which assumed invertibility and diagonalizability of matrices of $\mathcal{T}$ can be applied to yield Corollaries 1 and 3.

Finally, the result of (Tsakiris, 2018b; 2020a; Tsakiris & Peng, 2019) can be applied to obtain Corollaries 1-4. However, our quasi-variety condition (12) is tighter and simpler than theirs, as mentioned before.

3.2. The number of samples for sparse-hsp

With the $m \times m$ identity matrix I_m, we recall the following result in compressed sensing (see (Wright & Ma, 2020)).

Proposition 10. *If* $m \geq 2k$ *then* $\text{sparse-hsp}(\{I_m\}, A)$ *holds for* $A \in \mathbb{R}^{m\times n}$ *generic*[10]. *Conversely, if* $m < 2k \leq n$ *then* $\text{sparse-hsp}(\{I_m\}, A)$ *is violated for any* $A \in \mathbb{R}^{m\times n}$.

Proposition 10 presents a threshold $2k$ for compressed sensing ($\text{sparse-hsp}(\{I_m\}, A)$). Note that, the sets of matrices of our interests, say $\mathcal{S}_m$, are of exponential cardinality, and that $\text{sparse-hsp}(\mathcal{S}_m, A)$ generalizes $\text{sparse-hsp}(\{I_m\}, A)$. Naturally, do we need exponential many samples to guarantee $\text{sparse-hsp}(\mathcal{S}_m, A)$? To our surprise, from Theorem 2 and Proposition 9 we see that $2k$ samples still suffice for sparse unique recovery (up to sign), as summarized in the following corollaries, which hold for a generic $A \in \mathbb{R}^{m\times n}$.

Corollary 5. $m \geq 2k \Rightarrow \text{sparse-hsp}(\mathcal{S}_m, A)$

Corollary 6. $r \geq 2k \Rightarrow \text{sparse-hsp}(\mathcal{S}_{r,m}, A)$

Corollary 7. $m \geq 2k \Rightarrow \text{sparse-hsp}_{\pm}(\mathcal{B}_m, A)$
(Akçakaya & Tarokh, 2013; Wang & Xu, 2014)

Corollary 8. $r \geq 2k \Rightarrow \text{sparse-hsp}_{\pm}(\mathcal{S}_{r,m}\mathcal{B}_m, A)$

Corollary 7 is for sparse phase retrieval, and was proved independently by (Wang & Xu, 2014) and (Akçakaya & Tarokh, 2013). On the other hand, Corollaries 5, 6, and 8 are for sparse versions of linear regression without correspondences, unlabeled sensing, unsigned unlabeled sensing, respectively, and they are novel, to the best our knowledge. In fact, unlabeled sensing variants with sparsity assumptions on x^* are mostly unexplored in prior works: we have not known any other related theoretical or algorithmic results.

Finally, Proposition 6, with Corollaries 5-8, implies that, whenever there are more than $2k$ samples ($r \geq 2n$ or $m \geq 2n$), the ℓ_0 minimization problem (6) with $\mathcal{T}$ replaced by any one of the four sets of matrices admits a unique (or up to sign) solution x^* for a generic $A \in \mathbb{R}^{m\times n}$.

3.3. The local stability for unlabeled sensing

4. Discussion and future work

References

Abid, A. and Zou, J. A stochastic expectation-maximization approach to shuffled linear regression. In *2018 56th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pp. 470–477, 2018.

Akçakaya, M. and Tarokh, V. New conditions for sparse phase retrieval. *arXiv preprint arXiv:1310.1351*, 2013.

Arridge, S., Maass, P., Öktem, O., and Schönlieb, C.-B. Solving inverse problems using data-driven models. *Acta Numerica*, 28:1–174, 2019. doi: 10.1017/S0962492919000059.

Balakrishnan, S., Wainwright, M. J., and Yu, B. Statistical guarantees for the em algorithm: From population to sample-based analysis. *Ann. Statist.*, 45(1):77–120, 02 2017. doi: 10.1214/16-AOS1435. URL https://doi.org/10.1214/16-AOS1435.

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

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. doi: 10.1007/s10107-019-01363-6. URL https://doi.org/10.1007/s10107-019-01363-6.

Dokmanic, I. Permutations unlabeled beyond sampling unknown. *IEEE Signal Process. Lett.*, 26(6):823–827, June 2019. doi: 10.1109/LSP.2019.2908505.

Donoho, D. L. and Elad, M. Optimally sparse representation in general (nonorthogonal) dictionaries via l1 minimization. *Proceedings of the National Academy*

[10] If $m \geq 2k$ then a generic $A \in \mathbb{R}^{m\times n}$ is of *kruskal rank* at least $\min\{n, 2k\}$, which implies $\text{sparse-hsp}(\{I_m\}, A)$.

of Sciences, 100(5):2197–2202, 2003. doi: 10.1073/pnas.0437847100. URL https://www.pnas.org/content/100/5/2197.

Fellegi, I. P. and Sunter, A. B. A theory for record linkage. *Journal of the American Statistical Association*, 64 (328):1183–1210, 1969. doi: 10.1080/01621459.1969.10501049.

Grohs, P., Koppensteiner, S., and Rathmair, M. Phase retrieval: Uniqueness and stability. *SIAM Review*, 62 (2):301–350, 2020. doi: 10.1137/19M1256865. URL https://doi.org/10.1137/19M1256865.

Hadamard, J. Sur les problèmes aux dérivées partielles et leur signification physique. *Princeton university bulletin*, pp. 49–52, 1902.

Hadamard, J. *Lectures on Cauchy's Problem in Linear Partial Differential Equations*, volume 37. Yale University Press, 1923.

Haghighatshoar, S. and Caire, G. Signal recovery from unlabeled samples. *IEEE Trans. Signal Process.*, 66(5): 1242–1257, Mar. 2018. doi: 10.1109/TSP.2017.2786276.

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, July 2014. doi: 10.1109/TIT.2014.2320937.

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. doi: https://doi.org/10.1016/j.acha.2016.04.002. URL http://www.sciencedirect.com/science/article/pii/S1063520316300021.

Hsu, D., Shi, K., and Sun, X. Linear regression without correspondence. In *Advances in Neural Inf. Process. Syst. 30 (NeurIPS)*, 2017.

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. doi: 10.1109/ACCESS.2019.2933650.

Klibanov, M. V., Sacks, P. E., and Tikhonravov, A. V. The phase retrieval problem. *Inverse Problems*, 11(1):1–28, feb 1995. doi: 10.1088/0266-5611/11/1/001. URL https://doi.org/10.1088%2F0266-5611%2F11%2F1%2F001.

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, June 2019. doi: 10.1109/TIT.2019.2891628.

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. doi: https://doi.org/10.1016/0024-3795(77)90069-6. URL http://www.sciencedirect.com/science/article/pii/0024379577900696.

Lahiri, P. and Larsen, M. D. Regression analysis with linked data. *J. Am. Stat. Assoc.*, 100(469):222–230, 2005. doi: 10.1198/016214504000001277.

Liu, G., Liu, Q., and Yuan, X. A new theory for matrix completion. In Guyon, I., Luxburg, U. V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., and Garnett, R. (eds.), *Advances in Neural Information Processing Systems*, volume 30, pp. 785–794. Curran Associates, Inc., 2017. URL https://proceedings.neurips.cc/paper/2017/file/07563a3fe3bbe7e3ba84431ad9d055af-Paper.pdf.

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. doi: 10.1109/TPAMI.2019.2937869.

Lv, F. and Sun, W. Real phase retrieval from unordered partial frame coefficients. *Advances in Computational Mathematics*, 44(3):879–896, Jun 2018. doi: 10.1007/s10444-017-9566-4. URL https://doi.org/10.1007/s10444-017-9566-4.

Nejatbakhsh, A. and Varol, E. Robust approximate linear regression without correspondence. *arXiv e-prints*, art. arXiv:1906.00273, Jun 2019.

Nguyen, J. P., Linder, A. N., Plummer, G. S., Shaevitz, J. W., and Leifer, A. M. Automatically tracking neurons in a moving and deforming brain. *PLOS Computational Biology*, 13(5):1–19, 05 2017. doi: 10.1371/journal.pcbi.1005517. URL https://doi.org/10.1371/journal.pcbi.1005517.

Pananjady, A., Wainwright, M. J., and Courtade, T. A. Denoising linear models with permuted data. In *2017 IEEE International Symposium on Information Theory (ISIT)*, pp. 446–450, June 2017. doi: 10.1109/ISIT.2017.8006567.

Pananjady, A., Wainwright, M. J., and Courtade, T. A. Linear regression with shuffled data: Statistical and computational limits of permutation recovery. *IEEE Trans. Inf. Theory*, 64(5):3286–3300, May 2018. doi: 10.1109/TIT.2017.2776217.

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

Peng, L., Song, X., Tsakiris, M. C., Choi, H., Kneip, L., and Shi, Y. Algebraically-initialized expectation maximization for header-free communication. In *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, pp. 5182–5186, May 2019. doi: 10.1109/ICASSP.2019.8682683.

Puthawala, M., Kothari, K., Lassas, M., Dokmanić, I., and de Hoop, M. Globally Injective ReLU Networks. *arXiv e-prints*, art. arXiv:2006.08464, June 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. doi: 10.1080/01621459.2020.1752219. URL https://doi.org/10.1080/01621459.2020.1752219.

Slawski, M. and Ben-David, E. Linear regression with sparsely permuted data. *Electron. J. Statist.*, 13(1):1–36, 2019. doi: 10.1214/18-EJS1498.

Slawski, M., Diao, G., and Ben-David, E. A Pseudo-Likelihood Approach to Linear Regression with Partially Shuffled Data. 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. URL http://jmlr.org/papers/v21/19-645.html.

Song, X., Choi, H., and Shi, Y. Permuted linear model for header-free communication via symmetric polynomials. In *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, pp. 661–665, June 2018. doi: 10.1109/ISIT.2018.8437552.

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

Tsakiris, M. C. Determinantal conditions for homomorphic sensing. *arXiv e-prints*, art. arXiv:1812.07966, December 2018a.

Tsakiris, M. C. Eigenspace conditions for homomorphic sensing. *ArXiv e-prints*, 2018b.

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

Tsakiris, M. C. An exposition to the finiteness of fibers in matrix completion via Plucker coordinates. *arXiv e-prints*, art. arXiv:2004.12430, April 2020b.

Tsakiris, M. C. and Peng, L. Homomorphic sensing. In *Proc. the 36th Int. Conf. Mach. Learning*, Long Beach, California, USA, June 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*, pp. 1–1, 2020. doi: 10.1109/TIT.2020.2977166.

Unnikrishnan, J., Haghighatshoar, S., and Vetterli, M. Unlabeled sensing: Solving a linear system with unordered measurements. In *2015 53rd Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pp. 786–793, 2015.

Unnikrishnan, J., Haghighatshoar, S., and Vetterli, M. Unlabeled sensing with random linear measurements. *IEEE Trans. Inf. Theory*, 64(5):3237–3253, May 2018. doi: 10.1109/TIT.2018.2809002.

Wang, G., Marano, S., Zhu, J., and Xu, Z. Target localization by unlabeled range measurements. *IEEE Transactions on Signal Processing*, 68:6607–6620, 2020. doi: 10.1109/TSP.2020.3038230.

Wang, Y. and Xu, Z. Phase retrieval for sparse signals. *Applied and Computational Harmonic Analysis*, 37(3):531 – 544, 2014. doi: https://doi.org/10.1016/j.acha.2014.04.001. URL http://www.sciencedirect.com/science/article/pii/S1063520314000517.

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. *arXiv e-prints*, art. arXiv:2012.00123, November 2020.

Xu, Z. The minimal measurement number for low-rank matrix recovery. *Applied and Computational Harmonic Analysis*, 44(2):497 – 508, 2018. doi: https://doi.org/10.1016/j.acha.2017.01.005. URL http://www.sciencedirect.com/science/article/pii/S1063520317300088.

Zhang, H. and Li, P. Optimal estimator for unlabeled linear regression. In III, H. D. and Singh, A. (eds.), *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pp. 11153–11162, Virtual, 13–18 Jul 2020. PMLR. URL http://proceedings.mlr.press/v119/zhang20n.html.

Zhang, H., Slawski, M., and Li, P. Permutation recovery from multiple measurement vectors in unlabeled sensing. In *2019 IEEE International Symposium on Information Theory (ISIT)*, pp. 1857–1861, 2019a. doi: 10.1109/ISIT.2019.8849447.

Zhang, H., Slawski, M., and Li, P. The Benefits of Diversity: Permutation Recovery in Unlabeled Sensing from Multiple Measurement Vectors. *arXiv e-prints*, art. arXiv:1909.02496, September 2019b.

Zhang, Y. When is missing data recoverable? Technical report, 2006. URL https://hdl.handle.net/1911/102061.

Zhu, J., Cao, H., Song, C., and Xu, Z. Parameter estimation via unlabeled sensing using distributed sensors. *IEEE Communications Letters*, 21(10):2130–2133, Oct 2017. doi: 10.1109/LCOMM.2017.2695998.