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

Anonymous Authors[1]

Abstract

Liangzu Peng, December 17, 2020.

1. Introduction

Many tasks in machine learning belong to the category of 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.

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. Let it be recalled that a solution x^* is unique if and only if A is of full column rank ($m \geq n$). This characterization gives the following *mental decision boundary*. On the one side of the boundary, where A has full column rank, we can choose Gaussian elimination or else to find the unique solution. 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. Of the infinitely many, practitioners have assumed x^* as a sparsest solution, thus entering the realm of compressed sensing. Interestingly, the fundamental problem recurs: is a sparsest solution 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.

Recent years have also seen growing interests in the uniqueness for modern inverse problems, e.g., matrix completion (Tsakiris, 2020), deep networks (Puthawala et al., 2020). This paper focuses on the uniqueness for a class of generalizations of (sparse) linear regression, where the linear measurements Ax^* are corrupted by an *unknown* $r \times m$ matrix T^* with $r \leq m$; we observe $y = T^*Ax^*$. Without any further information about T^*, uniquely recovering x^* appears difficult, if not impossible. The case of our interest is when T is 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 T and x. This is the problem of *homomorphic sensing*, recently posed by (Tsakiris & Peng, 2019). Note that x^* (with T^*) is clearly a solution to (1), the fundamental question is whether it is unique.[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. On the other hand, real-world data applications with missing values, missing correspondences, or missing signs have supplied several 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) 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, 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, those applications have given several choices of $\mathcal{T}$. The special case where $\mathcal{T}$ contains only one matrix is that of linear regression. We next discuss more cases, from a unified perspective and thus as part of our contributions.

[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]Unique recovery of T^* is not considered here.

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 & Peng, 2019) as an example of homomorphic sensing. Here we motivate it from *record linkage* (Fellegi & Sunter, 1969; Lahiri & Larsen, 2005), 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 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 $r \le m$. 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 now 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, e.g., the above S_1 is order-preserving but S_2 is not.

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

We found the *missing data recovery* problem (Zhang, 2006; Liu et al., 2017; Liu et al., 2019) or as *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.

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 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 (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

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

1.2. Contributions of this paper

2. Theory

2.1. Homomorphic sensing

The uniqueness of the solution to (1) involves the regression coefficients $x^* \in \mathbb{R}^n$, the design matrix $A \in \mathbb{R}^{m\times n}$, and the finite set $\mathcal{T} \subset \mathbb{R}^{r\times m}$ of matrices. We first determine what kinds of x^*, A, and $\mathcal{T}$ to work with, respectively in §2.1.1, §2.1.2, and §2.1.3. This will shed light on the unique recovery conditions for homomorphic sensing (§2.1.4).

2.1.1. Which regression coefficients?

Since x^* might be arbitrary, and we want to guarantee the unique recovery of all possible $x^* \in \mathbb{R}^n$ (a sparse x^* is considered in §2.2), we need 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*

$$T_1Ax_1 = T_2Ax_2 \;\Rightarrow\; x_1 = x_2, \tag{6}$$

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

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

Definition 1 (hsp) encodes our desire: if $\operatorname{hsp}(\mathcal{T}, A)$ holds, then for any $x^* \in \mathbb{R}^n$, the solution to (1) is unique.

2.1.2. Which design matrices?

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, $\operatorname{hsp}(\mathcal{T}, A)$ can not hold for any A. The second best to hope is that $\operatorname{hsp}(\mathcal{T}, A)$ 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*. It is a subset $\mathcal{Q}$ of $\mathbb{C}^m$ (*resp.* $\mathbb{R}^m$), consisting of the common roots of polynomials $p_1, \dots, p_s$ in m variables with coefficients in $\mathbb{C}$ (*resp.* $\mathbb{R}$), i.e.,

$$\mathcal{Q} = \{z : p_i(z) = 0, \ \forall i = 1, \dots, s\}. \tag{7}$$

For example, $\mathbb{R}^{m\times n}$ is a real algebraic variety defined by the zero polynomial. Since each polynomial defines a hypersurface[4], an algebraic variety can be thought of as the intersection of finitely many hypersurfaces. A subvariety of an algebraic variety $\mathcal{Q}$ is a variety contained in $\mathcal{Q}$. By a *generic* matrix of $\mathbb{R}^{m\times n}$ having some property of interest, 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. To be more specific, that a generic $A \in \mathbb{R}^{m\times n}$ satisfies $\operatorname{hsp}(\mathcal{T}, A)$ implies that $\operatorname{hsp}(\mathcal{T}, A)$ holds with probability 1 if the entries of A are sampled independently at random according to some continuous probability distribution. Finally, observe that a generic $A \in \mathbb{R}^{m\times n}$ is of full column rank whenever $m \geq n$.

2.1.3. 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 $\operatorname{hsp}(\mathcal{T}, A)$ for a generic A? A very easy case is when $\mathcal{T} = \{T\}$ contains only one matrix (linear regression); then hsp (6) would require to consider only the null space of TA. The difficulty is to generalize this consideration to where $\mathcal{T}$ contains two or more matrices.

[4] In this paper, hypersurfaces are not necessarily irreducible.

[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.

We approach the answer by first investigating into how $\mathcal{T}$ violates $\operatorname{hsp}(\mathcal{T}, A)$. Since hsp of Definition 1 involves two matrices of $\mathcal{T}$ at a time, let us first focus on $T_1, T_2 \in \mathcal{T}$. By Definition 1, $\operatorname{hsp}(\mathcal{T}, A)$ is related to how the column spaces of T_1A and T_2A interact. It is thus natural to look first at how T_1 and T_2 interact. Define the set

$$\mathcal{Z}_{T_1,T_2} = \{w \in \mathbb{C}^m : T_1w = T_2w\}. \tag{8}$$

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 $\dim_{\mathbb{C}}(\mathcal{Z}_{T_1,T_2}) = m$ then 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, $T_1w = T_2w$. Finally, we remark that, even if T_1, 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* $\operatorname{rank}[T_1\ T_2] < 2n$ *and* $\dim_{\mathbb{R}}(\mathcal{Z}_{T_1,T_2} \cap \mathbb{R}^m) \leq m - n$. *Then a generic $A \in \mathbb{R}^{m\times n}$ violates* $\operatorname{hsp}(\mathcal{T}, A)$.

Proof. If $\operatorname{rank}(T_1A) < n$ then any $A \in \mathbb{R}^{m\times n}$ violates $\operatorname{hsp}(\{T_1\}, A)$. So assume $m \geq n$, $r \geq n$, $\operatorname{rank}(T_1A) = n$, and similarly $\operatorname{rank}(T_2A) = n$. 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 $x_1, x_2 \in \mathbb{R}^n$ such that $T_1Ax_1 = T_2Ax_2$. Assume for 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 $A \in \mathbb{R}^{m\times n}$ we have that $R(A) \cap \mathcal{Z}_{T_1,T_2} = \{0\}$. This gives $Ax_1 = 0$ and $x_1 = 0$, a contradiction. Hence $x_1 \neq x_2$. $\square$

To understand why the condition $\operatorname{rank}[T_1\ T_2] < 2n$ of Proposition 1 is the potential source of the violating $\operatorname{hsp}(\mathcal{T}, A)$, consider the $r \times 2n$ matrix $[T_1A\ T_2A]$. If it has full column rank $2n$, then $T_1Ax_1 = T_2Ax_2$ with any $x_1, x_2 \in \mathbb{R}^n$ implies $x_1 = x_2 = 0$; there is no chance for violation. Such chance only emerges when $\operatorname{rank}[T_1A\ T_2A] < 2n$, or simply when $\operatorname{rank}[T_1\ T_2] < 2n$, and moreover the chance becomes a truth, indeed, if T_1 and T_2 are not similar in the sense that $\dim_{\mathbb{C}}(\mathcal{Z}_{T_1,T_2}) \leq m - n$.

How to prevent the violation presented in Proposition 1? It is the insight of (Tsakiris & Peng, 2019) that considered:

The Rank Constraint

$$\mathrm{rank}(T) \geq 2n, \ \ \forall\, T \in \mathcal{T}. \tag{9}$$

The rank constraint (9) 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 (9), however, does not exclude all possible violations of hsp. We next introduce an algebraic-geometric object that also accounts for the similarity between T_1 and T_2. For a column vector w of m variables, consider all 2×2 determinants of the $r\times 2$ matrix $[T_1 w \ \ T_2 w]$. Since each determinant is a quadratic polynomial in entries of w, we obtain $\binom{r}{2}$ polynomials in total. Recalling (7), 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 solutions of $w_1 w_2 - w_3^2 = 0$.

Alternatively, we might describe $\mathcal{Y}_{T_1,T_2}$ as the set of vectors w's of $\mathbb{C}^m$ such that $[T_1 w \ \ T_2 w]$ has rank at most 1, or equivalently such that $T_1 w$ and $T_2 w$ are linearly dependent. Observe that $\mathcal{Z}_{T_1,T_2}$ of (8) is a subvariety of $\mathcal{Y}_{T_1,T_2}$. Define

$$\mathcal{U}_{T_1,T_2} := \mathcal{Y}_{T_1,T_2} \backslash \mathcal{Z}_{T_1,T_2} \tag{10}$$

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 $w \in \mathcal{U}_{T_1,T_2}$, we have $T_1 w = \lambda T_2 w$ or $\lambda T_1 w = T_2 w$, for some $\lambda \in \mathbb{C}$ with $\lambda \neq 1$. In words, T_1 and T_2 send w to the same destination, e.g., $\lambda T_1 w = T_2 w$, 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 $\mathrm{hsp}(\mathcal{T}, A)$ is violated.*

Proof. Let $w \in \mathbb{R}^m$ in the intersection. Then $T_1 w$ and $T_2 w$ are linearly dependent. So there is some $\lambda \in \mathbb{R}$ such that $T_1 w = \lambda T_2 w$ or $\lambda T_1 w = T_2 w$. Since $w \in \mathcal{U}_{T_1,T_2}$, definition (10) implies that $w \neq 0$ and $\lambda \neq 1$. Since $w \in R(A)$ we have for some $x \in \mathbb{R}^n$ with $Ax = w$ that $T_1 A x = T_2 A(\lambda x)$ or $T_1 A(\lambda x) = T_2 A x$. 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[6], 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[7], 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., $\mathcal{U}_{T_1,T_2}$)? Noting that $0 \notin \mathcal{U}_{T_1,T_2}$, we answer it 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 motivates:

The Quasi-variety Constraint

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

[6] 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.

[7] 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.

The quasi-variety constraint (11) and the rank constraint (9), once combined together, are able to sidestep all possible violations of hsp, as we will soon see (§2.1.4).

2.1.4. Recovery guarantees

Theorem 1. *If a finite set $\mathcal{T} \subset \mathbb{R}^{r\times m}$ of matrices with $r \leq m$ satisfies the rank constraint (9) and quasi-variety constraint (11), we have $\mathrm{hsp}(\mathcal{T}, A)$ 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 (9) and (11) provide a mathematical awareness about the uniqueness of (1). It encourages pragmatic practitioners to make their algorithmic choices that respect (9) and (11).[8]

We next compare Theorem 1 to that of (Tsakiris & Peng, 2019). In their result (also mentioned in §1.2), they used the same rank constraint (9) and a different quasi-variety constraint. More specifically, given $T_1, T_2 \in \mathcal{T}$, they considered a projection $P \in \mathbb{R}^{r\times r}$ onto the column space of T_2, defined $\mathcal{Z}_{PT_1,T_2}$ as the set of w's of $\mathbb{C}^m$ satisfying $PT_1w = T_2w$, defined $\mathcal{Y}_{PT_1,T_2}$ as the set of w's of $\mathbb{C}^m$ for which PT_1w and T_2w is linearly dependent, and defined $\mathcal{U}_{PT_1,T_2} := \mathcal{Y}_{PT_1,T_2}\backslash\mathcal{Z}_{PT_1,T_2}$. Their quasi-variety constraint is that $\dim(\mathcal{U}_{PT_1,T_2}) \leq m-n$ for every $T_1, T_2 \in \mathcal{T}$. Their constraint implies (11) for the following reason.

Proposition 5. *$\mathcal{U}_{T_1,T_2}$ is a subset of $\mathcal{U}_{PT_1,T_2}$, and as a result, we have $\dim(\mathcal{U}_{T_1,T_2}) \leq \dim(\mathcal{U}_{PT_1,T_2})$.*

In view of Proposition 5, constraint (11) is tighter. Moreover, since we dispensed with P, our condition becomes much simpler because it has been unknown whether such P that satisfies $\dim(\mathcal{U}_{PT_1,T_2}) \leq m-n$ exists, or even if so how to search for it. Last but not least, the technique that we used for dispensing with P is non-trivial; see the supplementary.

Finally, we extend Theorem 1 for $\mathrm{hsp}_\pm(\mathcal{T}, A)$. Defining $\mathcal{Z}^\pm_{T_1,T_2} := \{w \in \mathbb{C}^m : T_1w = T_2w \text{ or } T_1w = -T_2w\}$ as the union of two linear subspaces, it is the quasi-variety

$$\mathcal{U}^\pm_{T_1,T_2} := \mathcal{Y}_{T_1,T_2}\backslash\mathcal{Z}^\pm_{T_1,T_2} \tag{12}$$

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

Corollary 1. *Suppose $r \leq m$ and that (9) holds. Then we have $\mathrm{hsp}_\pm(\mathcal{T}, A)$ for a generic $A \in \mathbb{R}^{m\times n}$ whenever*

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

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 in the sense that 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 $\mathbb{R}^n$. This further indicates a possibility of the existence of less-demanding conditions for the uniqueness of k-sparse solutions, as explored next.

To start with, we hereby 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 (6) 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 $T_1Ax_1 = T_2Ax_2$ only implies $x_1 = \pm x_2$, then we write $\text{sparse-hsp}_\pm(\mathcal{T}, A)$.*

Similar to Definition 1, if $\text{sparse-hsp}(\mathcal{T}, A)$ holds, then for any k-sparse $x^* \in \mathbb{R}^n$, (1) has a unique k-sparse solution. In particular, we have the following immediate result.

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 (**??**) has a unique optimal solution, which is necessarily x^*.*

Proof. Note that $y = T^*Ax^*$ for some $T^* \in \mathcal{T}$. Let x^+ be an optimal solution, so $\|x^+\|_0 \leq \|x^*\|_0$. Then x^+ is k-sparse and $y = T^+Ax^+$ 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 (**??**) is NP-hard. We name the exploration of its convex relaxation for specific types of $\mathcal{T}$ (e.g., $\mathcal{S}_{r,m}, \mathcal{S}_m$) as future work.*

It remains to answer when sparse-hsp holds. The major hurdle towards this 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 two classic notions in compressed sensing, namely *spark* or *Kruskal rank* (see, e.g., (Kruskal, 1977)). However, this does not apply, at least in an obvious way, to the case where $\mathcal{T}$ contains two or more matrices.

Fortunately, we have known that the rank and quasi-variety constraints (9), (11) are key to approaching hsp for a finite set $\mathcal{T}$ (Theorem 1). It remains to analyze how the quasi-variety $\mathcal{U}_{T_1,T_2}$ interacts with $\tau_A(\mathcal{K})$ where $\tau_A : \mathbb{R}^n \to \mathbb{R}^m$ is a linear map which sends x to Ax (e.g., $\tau_A(\mathbb{R}^n)$ is the column space of A). 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 of $\mathbb{R}^n$. Specifically, 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 this way allows us to analyze it through the

[8] We discuss how constraint (11) behaves in applications (§3).

n independent lines rather than through the potentially dependent subspaces. This insight, combined with techniques in algebraic geometry and based on Theorem 1, gives:

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

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

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 $k \ll n$ in practice, the conditions of Theorem 2 are less demanding for guaranteeing unique sparse recovery.

Remark 2. *It is easy to extend Theorem 2 for guaranteeing $\text{sparse-hsp}_{\pm}(\mathcal{T}, A)$; see also Corollary 1.*

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

3. Applications

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.

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., 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.

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

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, 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.

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 2019.

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.