Homomorphic sensing

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

Abstract

A recent line of research termed *unlabeled sensing* and *shuffled linear regression* has been exploring under great generality the recovery of signals from subsampled and permuted measurements; a challenging problem in diverse fields of data science and machine learning. In this paper we introduce an abstraction of this problem which we call *homomorphic sensing*. Given a linear subspace and a finite set of linear transformations we develop an algebraic theory which establishes conditions guaranteeing that points in the subspace are uniquely determined from their homomorphic image under some transformation in the set. As a special case, we recover known conditions for unlabeled sensing, as well as new results and extensions. On the algorithmic level we exhibit two dynamic programming based algorithms, which to the best of our knowledge are the first working solutions for the unlabeled sensing problem for small dimensions. One of them, additionally based on branch-and-bound, when applied to image registration under affine transformations, performs on par with or outperforms state-of-the-art methods on benchmark datasets.

1. Introduction

In a recent line of research termed *unlabeled sensing*, it has been established that uniquely recovering a signal from shuffled and subsampled measurements is possible as long as the number of measurements is at least twice the intrinsic dimension of the signal (Unnikrishnan et al., 2018). The special case where the signal is fully observed but subject to a permutation of its values is known as *shuffled linear regression* (Hsu et al., 2017; Pananjady et al., 2018; Song et al., 2018; Abid & Zou, 2018). In its simplest form, it consists of solving a linear system of equations, with the

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

entries of the right hand side vector permuted (Unnikrishnan et al., 2015; Tsakiris et al., 2018).

The unlabeled sensing or shuffled linear regression problems and their variations naturally arise in many applications in data science and engineering, such as 1) record linkage for data integration (Lahiri & Larsen, 2005; Shi et al., 2018), a particularly important problem in medical data analysis where publicly available health records are anonymized, 2) image registration (Lian et al., 2017), multi-target tracking (Poore & Gadaleta, 2006) and pose/correspondence estimation (David et al., 2004; Marques et al., 2009), 3) header-free communications in Internet-Of-Things networks (Pananjady et al., 2017; Song et al., 2018; Peng et al., 2019) and user de-anonymization (Narayanan & Shmatikov, 2008; Keller et al., 2009), 4) acoustic wave field reconstruction (Dokmanić, 2019), 5) system identification under asynchronous input/output samples (Balakrishnan, 1962), and many more, e.g., see Unnikrishnan et al. 2018; Pananjady et al. 2018.

1.1. Prior-Art

Theory. Suppose that $y = \Pi^* A x^* + \varepsilon \in \mathbb{R}^m$ is a noisy and shuffled version of some signal $A x^*$, where $x^* \in \mathbb{R}^n$ is some unknown regression vector, Π^* is some unknown permutation, and ε is noise. What can be said about the estimation of x^* and Π^* given y, A and the distribution of ε? This shuffled linear regression problem has been a classic subject of research in the area of record linkage, where predominant methods study maximum likelihood estimators under the working hypothesis that an accurate estimate for the probabilities of transpositions between samples is available (Fellegi & Sunter, 1969; Lahiri & Larsen, 2005). However, this is a strong hypothesis that does not extend to many applications beyond record linkage.

Very recently important theoretical advances have been made towards understanding this problem in greater generality. Specifically, Slawski & Ben-David 2019; Abid et al. 2017 and Hsu et al. 2017 have demonstrated that in the absence of any further assumptions, the maximum likelihood estimator $\widehat{x}_{\text{ML}}$ given by

$$(\widehat{\Pi}_{\text{ML}}, \widehat{x}_{\text{ML}}) = \underset{\Pi, x}{\operatorname{argmin}} \| y - \Pi A x \|_2, \qquad (1)$$

where Π ranges over all permutations, is biased. On the

other hand, if the SNR is large enough, Pananjady et al. 2016; 2018 have asserted that $\widehat{\Pi}_{\mathrm{ML}} = \Pi^*$ with high probability. If Π^* is sparse enough, i.e., only a small percentage of entries of Ax^* have been shuffled (this is the support of Π^*), Slawski & Ben-David 2019 have shown that under weaker SNR conditions the supports of $\widehat{\Pi}_{\mathrm{ML}}, \Pi^*$ coincide. Moreover, they provide well behaved error bounds for $\|\widehat{x}_{\mathrm{ML}} - x^*\|_2$ as well as for $\|\widehat{x}_{\mathrm{RR}} - x^*\|_2$, where $\widehat{x}_{\mathrm{RR}}$ is the solution to the convex ℓ_1 robust regression problem

$$\min_{x,e} \; \|y - Ax - \sqrt{m}e\|_2^2 + m\lambda\|e\|_1, \;\; \lambda > 0, \quad (2)$$

in which the support of the sparse error e is meant to capture the support of the sparse permutation Π^*.

Another interesting line of work related more to algebraic geometry rather than statistics, is that of Song et al. 2018, which for the noiseless case ($\varepsilon = 0$) has proposed the use of symmetric polynomials towards extracting permutation-invariant constraints that $x^* \in \mathbb{R}^n$ must satisfy. Such a *self-moment* estimator had already been briefly investigated by Abid et al. 2017 from a statistical point of view, where the authors noted that in the presence of noise it is unclear whether the resulting system of equations has any solutions. Perhaps surprisingly, working with n non-homogeneous polynomials of degrees $1, 2, \ldots, n$ in n variables, the work of Tsakiris et al. 2018 has established that regardless of the value of the noise ε and under the sole requirement that A is generic[1], the polynomial system always has a solution and in fact at most $n!$ of them, thus proving the existence of a purely algebraic estimator for x^*.

Much less is known for the more challenging and realistic case of unlabeled sensing, where now $y \in \mathbb{R}^k$ consists of a shuffled noisy subset of the entries of $Ax^* \in \mathbb{R}^m$, i.e., there is no longer a 1-1 correspondence between y and Ax^*. The main theoretical finding up to date comes from the seminal work of Unnikrishnan et al. 2018, according to which, in the absence of noise, x^* is uniquely recoverable from y and A as long as 1) $k \geq 2n$ and 2) A is generic. Inspired by a certain duality between compressed and unlabeled sensing, a recovery condition for noisy data has further been given by Haghighatshoar & Caire 2018 in terms of a restricted isometry property. However, this approach is valid only for the special case of y obtained by subsampling Ax^* while maintaining the relative order of the samples.

Algorithms. Towards computing a solution to the shuffled linear regression problem, which can be solved by brute force in $\mathcal{O}(m!)$, the algorithms presented by Hsu et al. 2017; Pananjady et al. 2017 are conceptually important but applicable only for noiseless data or they have a complexity of at least $\mathcal{O}(m^7)$. When the ratio of shuffled data is small, one may apply the ℓ_1 robust regression method of

(2) (Slawski & Ben-David, 2019). Other approaches use alternating minimization or multi-start gradient descent to solve (1) (Abid et al., 2017; Abid & Zou, 2018), an NP-hard problem for $n > 1$ (Pananjady et al., 2018). Due to the high non-convexity such methods are very sensitive to initialization. This is remedied by the algebraically-initialized expectation-maximization method of Tsakiris et al. 2018, which uses the solution to the polynomial system of equations mentioned above to obtain a high-quality initialization. This approach is robust to small levels of noise, efficient for $n \leq 5$, and is able to handle fully shuffled data; its main drawback is its exponential complexity in n.

In the unlabeled sensing case, which may be thought of as shuffled linear regression with outliers, the above methods in principle break down. Instead, we are aware of only two relevant algorithms, which nevertheless are suitable under strong structural assumptions on the data. The $\mathcal{O}(nm^{n+1})$ method of Elhami et al. 2017 applies a brute-force solution, which explicitly relies on the data being noiseless and whose theoretical guarantees require a particular *exponentially spaced* structure on A. On the other hand, Haghighatshoar & Caire 2018 attempt to solve

$$\min_{S,x} \|y - SAx\|_2 \,, \quad (3)$$

via alternating minimization, with S in (3) being a selection matrix[2]. Their main algorithmic insight is to solve for S given x via dynamic programming. However, their algorithm works only for order-preserving selection matrices S, a rather strong limitation, and seems to fail otherwise. It is thus fair to conclude that, to the best of our knowledge, there does not seem to exist a satisfactory algorithm for unlabeled sensing, even for small values of n.

1.2. Contributions

Theory. In this work we adopt an abstract view of the shuffled linear regression and unlabeled sensing problems, which naturally leads us to a more general formulation that we refer to as *homomorphic sensing*. In homomorphic sensing one is given a finite set $\mathcal{T}$ of linear transformations $\mathbb{R}^m \to \mathbb{R}^m$ (to be called *endomorphisms*) and a linear subspace $\mathcal{V} \subset \mathbb{R}^m$ of dimension n, and asks under what conditions the image $\tau(v)$ of some unknown $v \in \mathcal{V}$ under some unknown $\tau \in \mathcal{T}$ is enough to uniquely determine what v is. This is equivalent to asking under what conditions the relation $\tau_1(v_1) = \tau_2(v_2)$ implies $v_1 = v_2$ whenever $\tau_1, \tau_2 \in \mathcal{T}$ and $v_1, v_2 \in \mathcal{V}$. E.g., in shuffled linear regression these endomorphisms are permutations, while in unlabeled sensing they are compositions of permutations with coordinate pro-

---

[1] A rigorous definition of *generic* will be given in §2.1

[2] An *order-preserving selection matrix* is a row-submatrix of the identity matrix. A *selection matrix* is a row-permutation of an order-preserving selection matrix. This is equivalent to a permutation matrix composed by a coordinate projection.

jections, and the unlabeled sensing theorem of Unnikrishnan et al. 2018 asserts that a sufficient condition for unique recovery is that 1) the coordinate projections preserve at least $2n$ coordinates and 2) $\mathcal{V}$ is generic.

The first theoretical contribution of this paper is a general homomorphic sensing result (Theorem 1) applicable to arbitrary endomorphisms, and thus of potential interest in a broad spectrum of applications. For generic $\mathcal{V}$, the key condition asks that the dimension n of $\mathcal{V}$ does not exceed the codimension of a certain algebraic variety associated with pairs of endomorphisms from $\mathcal{T}$. This in turn can be used to obtain within a principled framework the unlabeled sensing result of Unnikrishnan et al. 2018. Our second theoretical contribution is a recovery result (Theorem 2) for generic points in $\mathcal{V}$ (as opposed to all points), which for the unlabeled sensing case says that the coordinate projections need to preserve at least $n + 1$ coordinates (as opposed to $2n$).

Algorithms. Inspired by Haghighatshoar & Caire 2018; Elhami et al. 2017 and Yang et al. 2016 we make three algorithmic contributions. First, we introduce a branch-and-bound algorithm for the unlabeled sensing problem by globally minimizing (3). Instead of branching over the space of selection matrices, which is known to be intractable (Li & Hartley, 2007), our algorithm only branches over the space of $x \in \mathbb{R}^n$, relying on a locally optimal computation of the selection matrix via dynamic programming. Second, it is this dynamic programming feature that also allows us to modify the purely theoretical algorithm of Elhami et al. 2017 into a robust and efficient method for small dimensions n. These two algorithms constitute to the best of our knowledge the first working solutions for the unlabeled sensing problem. Third, when customized for image registration under an affine transformation, our branch-and-bound algorithm is on par with or outperforms state-of-the-art methods (Lian et al., 2017; Jian & Vemuri, 2011; Myronenko & Song, 2010) on benchmark datasets.

2. Homomorphic Sensing: Algebraic Theory

The main results of this section are Theorems 1 and 2. We sketch the proofs wherever space allows and refer the reader to Tsakiris 2019 for the complete arguments. All results in this section refer to the noiseless case; an analysis for corrupted data is left to future research.

2.1. Preliminaries

For an integer k, $[k] = \{1, \dots, k\}$. For non-negative real number α, $\lfloor \alpha \rfloor$ is the greatest integer k such that $k \le \alpha$.

2.1.1. $\mathbb{R}$ VERSUS $\mathbb{C}$

We work over the complex numbers $\mathbb{C}$. This does not contradict the fact that in this paper we are primarily inter-

ested in $\mathbb{R}^m$, rather it facilitates the analysis. E.g., a matrix $T \in \mathbb{R}^{m \times m}$ may be diagonalizable over $\mathbb{C}$ but not over $\mathbb{R}$. This is the case with permutations, whose eigenvalues are associated with the complex roots of unity. Hence our philosophy is "check the conditions over $\mathbb{C}$, then draw a conclusion over $\mathbb{R}$"; see Remark 1.

2.1.2. ABSTRACT LINEAR ALGEBRA

We adopt the terminology of abstract linear algebra (Roman, 2008), since the ideas we discuss in this paper are best delivered in a coordinate-free way. The reader who insists on thinking in terms of matrices may safely replace linear transformations, kernels and images by matrices, nullspaces and rangespaces, respectively.

We work in $\mathbb{C}^m$. For a subspace $\mathcal{V}$ we denote by $\dim(\mathcal{V})$ its dimension. For subspaces $\mathcal{V}, \mathcal{W}$ we say that "$\mathcal{V}, \mathcal{W}$ do not intersect" if $\mathcal{V} \cap \mathcal{W} = 0$. An endomorphism is a linear transformation $\tau : \mathbb{C}^m \to \mathbb{C}^m$; an automorphism is an invertible endomorphism. We denote by i the identity map $i(w) = w, \forall w \in \mathbb{C}^m$. If τ is an endomorphism, its kernel $\ker(\tau)$ is the set of all $v \in \mathbb{C}^m$ such that $\tau(v) = 0$, and its image $\operatorname{im}(\tau)$ is the set of all $\tau(w)$ for $w \in \mathbb{C}^m$. By $\operatorname{rank}(\tau)$ we mean $\dim(\operatorname{im}(\tau))$. The preimage of $\tau(v)$ is the set of all $w \in \mathbb{C}^m$ such that $\tau(v) = \tau(w)$, i.e., the set of all $v + \xi$, for all $\xi \in \ker(\tau)$. If $\mathcal{V}$ is a linear subspace, by $\tau(\mathcal{V})$ we mean the set of all vectors $\tau(v)$ for all $v \in \mathcal{V}$. We denote by $\mathscr{E}_{\tau, \lambda}$ the eigenspace of τ associated to eigenvalue λ, i.e., the set of all $v \in \mathbb{C}^m$ such that $\tau(v) = \lambda v$. For τ_1, τ_2 endomorphisms the generalized eigenspace of the pair (τ_1, τ_2) of eigenvalue λ is the set of all $w \in \mathbb{C}^m$ for which $\tau_1(w) = \lambda \tau_2(w)$. By a projection ρ of $\mathbb{C}^m$ we mean an idempotent ($\rho^2 = \rho$) endomorphism. By a coordinate projection we mean a projection that sets to zero certain coordinates of $\mathbb{C}^m$ while preserving the rest.

2.1.3. ALGEBRAIC GEOMETRY

By an algebraic variety (or variety) of $\mathbb{C}^m$ we mean the zero locus of a set of polynomials in m variables. The study of such varieties is facilitated by the use of the Zariski topology, in which every variety is a closed set. In particular, there is a well-developed theory in which topological and algebraic notions of dimension coincide (Hartshorne, 1977). This allows us to assign dimensions to sets such as the intersection of a variety with the complement of another, called quasi-variety. A linear subspace $\mathcal{S}$ is an algebraic variety and its linear-algebra dimension coincides with its algebraic-geometric dimension. The union $\mathcal{A} = \cup_{i \in [\ell]} \mathcal{S}_i$ of ℓ linear subspaces is also an algebraic variety and $\dim(\mathcal{A}) = \max_{i \in [\ell]} \dim(\mathcal{S}_i)$. For a variety $\mathcal{Y}$ which is defined by homogeneous polynomials $\dim(\mathcal{Y})$ can be characterized as the smallest number of hyperplanes through the origin 0 that one needs to intersect $\mathcal{Y}$ with to obtain

the origin. For $\mathcal{Y}$ a variety of $\mathbb{C}^m$, we set $\mathrm{codim}(\mathcal{Y}) = m - \dim(\mathcal{Y})$. The set of all n-dimensional linear subspaces of $\mathbb{C}^m$ is itself an algebraic variety of $\mathbb{C}^{\binom{m}{n}}$ called Grassmannian and denoted by $\mathrm{Gr}(n, m)$. By a generic subspace $\mathcal{V}$ of dimension n we mean a non-empty open subset $\mathbb{U} \subset \mathrm{Gr}(n, m)$ in the Zariski topology of $\mathrm{Gr}(n, m)$. Such a $\mathbb{U}$ is dense in $\mathrm{Gr}(n, m)$ and if one endows $\mathrm{Gr}(n, m)$ with a continuous probability measure, then $\mathbb{U}$ has measure 1 (Bürgisser & Cucker, 2013). Hence the reader may safely think of a *generic subspace* as a *random subspace*. When we say "for a generic $\mathcal{V}$ property $\mathscr{P}$ is true", we mean that the set of all $\mathcal{V}$ for which property $\mathscr{P}$ is true contains a non-empty Zariski open subset of $\mathrm{Gr}(n, m)$. Hence for a randomly drawn $\mathcal{V}$ property $\mathscr{P}$ will be true with probability 1. We will make repeated use of the following fact:

Lemma 1. *Let $\mathcal{Y}$ be a variety defined by homogeneous polynomials and $\mathcal{V}$ a generic linear subspace. Then*

$$\dim(\mathcal{Y} \cap \mathcal{V}) = \max\{\dim(\mathcal{V}) - \mathrm{codim}(\mathcal{Y}), 0\}. \quad (4)$$

A property of the Zariski topology that we need is that the intersection of finitely many non-empty open sets of an irreducible space (such as $\mathrm{Gr}(n, m)$) is open and non-empty.

2.2. Formulation and first insights

Let $\mathcal{V} \subset \mathbb{C}^m$ be a linear subspace of dimension n and let $\mathcal{T}$ be a finite set of endomorphisms of $\mathbb{C}^m$. Let v be some point (vector) in $\mathcal{V}$. Suppose that we know the image $\tau(v) \in \mathbb{C}^m$ of v for some unspecified $\tau \in \mathcal{T}$. Can we uniquely recover v from knowledge of $\mathcal{V}, \mathcal{T}$ and $\tau(v)$?

Example 1. *In shuffled linear regression $\mathcal{T}$ consists of the $m!$ permutations on the m coordinates of $\mathbb{C}^m$. In unlabeled sensing $\mathcal{T}$ consists of the set of all possible combinations of permutations composed with coordinate projections. In both cases $\mathcal{V}$ is the range-space of some matrix $A \in \mathbb{C}^{m \times n}$ and $v = Ax$ for some $x \in \mathbb{C}^n$. The meaning of A and x may vary depending on the application. E.g., in signal processing/control systems x may be the impulse response of some linear filter, while in image registration x may represent the parameters of some affine transformation.*

The above question motivates the following definition.

Definition 1. *Let $\mathcal{V} \subset \mathbb{C}^m$ be a subspace and $\mathcal{T}$ a finite set of endomorphisms of $\mathbb{C}^m$. We say that "$v_1 \in \mathcal{V}$ is uniquely recoverable in $\mathcal{V}$ under $\mathcal{T}$" if whenever $\tau_1(v_1) = \tau_2(v_2)$ for some $\tau_1, \tau_2 \in \mathcal{T}$ and $v_2 \in \mathcal{V}$, then $v_1 = v_2$. If this holds for every $v \in \mathcal{V}$, we have "unique recovery in $\mathcal{V}$ under $\mathcal{T}$".*

Remark 1. *Let $\mathcal{T}$ be a set of endomorphisms of $\mathbb{R}^m$. These can also be viewed as endomorphisms of $\mathbb{C}^m$ (Theorem 2.29 of Roman 2008). Let $\mathcal{V}$ be a subspace of $\mathbb{R}^m$ with basis $v_1, \ldots, v_n$ and $\mathcal{V}_\mathbb{C} = \mathrm{Span}_\mathbb{C}(v_1, \ldots, v_n)$ the subspace of $\mathbb{C}^m$ generated by the basis of $\mathcal{V}$. Then unique recovery in $\mathcal{V}_\mathbb{C}$ under $\mathcal{T}$ implies unique recovery in $\mathcal{V}$ under $\mathcal{T}$.*

To build some intuition about the notion of unique recovery in $\mathcal{V}$ under $\mathcal{T}$, consider first the case where $\mathcal{T} = \{i, \tau\}$ with i the identity map and τ some automorphism. As a first step, we characterize the purely combinatorial condition of Definition 1 given in terms of points by the geometric condition of the next proposition given in terms of subspaces.

Lemma 2. *Let τ be any automorphism of $\mathbb{C}^m$ and let $\mathcal{V}$ be any linear subspace. Then we have unique recovery in $\mathcal{V}$ under $\{i, \tau\}$ if and only if $\mathcal{V} \cap \tau(\mathcal{V}) \subset \mathscr{E}_{\tau, 1}$.*

Proof. (Necessity) Suppose that $\mathcal{V} \cap \tau(\mathcal{V}) \not\subset \mathscr{E}_{\tau, 1}$. Then there exists some $v_1 \in \mathcal{V} \cap \tau(\mathcal{V})$ not inside $\mathscr{E}_{\tau, 1}$. Note that this also implies that $v_1 \notin \mathscr{E}_{\tau^{-1}, 1}$. Since $v_1 \in \tau(\mathcal{V})$ there exists some $v_2 \in \mathcal{V}$ such that $v_1 = \tau(v_2)$. Since $v_1 \notin \mathscr{E}_{\tau^{-1}, 1}$ we must have that $v_1 \neq \tau^{-1}(v_1) = v_2$. Hence $v_1 = \tau(v_2)$ with $v_1 \neq v_2$. *(Sufficiency)* Suppose that $v_1 = \tau(v_2)$ for some $v_1, v_2 \in \mathcal{V}$, whence $v_1 \in \mathcal{V} \cap \tau(\mathcal{V})$. By hypothesis $v_1 \in \mathscr{E}_{\tau, 1}$. Hence $\tau(v_1) = v_1$. Hence $\tau(v_1) = \tau(v_2)$. Since τ is invertible, $v_1 = v_2$. $\square$

Notice that condition $\mathcal{V} \cap \tau(\mathcal{V}) \subset \mathscr{E}_{\tau, 1}$ of Lemma 2 prevents $\mathcal{V} \cap \tau(\mathcal{V})$ from intersecting $\mathscr{E}_{\tau, \lambda}$ for any $\lambda \neq 1$. Hence, a necessary condition for unique recovery is $\mathcal{V} \cap \tau(\mathcal{V})$ to not intersect $\mathscr{E}_{\tau, \lambda \neq 1}$. Notice that $\mathcal{V} \cap \tau(\mathcal{V}) \cap \mathscr{E}_{\tau, \lambda} = 0$ if and only if $\mathcal{V} \cap \mathscr{E}_{\tau, \lambda} = 0$. This places a restriction on the dimension of $\mathcal{V}$, for if $\dim(\mathcal{V}) > \mathrm{codim}(\mathscr{E}_{\tau, \lambda})$ then $\mathcal{V}, \mathscr{E}_{\tau, \lambda}$ will necessarily intersect. Hence, a necessary condition for unique recovery in $\mathcal{V}$ under $\{i, \tau\}$ is

$$\dim(\mathcal{V}) \leq \min_{\lambda \neq 1} \mathrm{codim} \, \mathscr{E}_{\tau, \lambda}. \quad (5)$$

Since the algebraic-geometric dimension of a finite union of linear subspaces is the maximum among the subspace dimensions (§2.1.3), condition (5) is equivalent to

$$\dim(\mathcal{V}) \leq \mathrm{codim}\left(\cup_{\lambda \neq 1} \mathscr{E}_{\tau, \lambda}\right). \quad (6)$$

Then for a generic $\mathcal{V}$ satisfying condition (6), Lemma 1 guarantees that $\mathcal{V} \cap \left(\cup_{\lambda \neq 1} \mathscr{E}_{\tau, \lambda}\right) = 0$.

2.3. Recovery under diagonalizable automorphisms

It is not hard to show that when τ is diagonalizable and $n = \dim(\mathcal{V})$ is small enough compared to m, condition (6) is also sufficient for unique recovery in $\mathcal{V}$ under $\{i, \tau\}$[3]:

Lemma 3. *Let τ be a diagonalizable automorphism of $\mathbb{C}^m$ and $\mathcal{V}$ generic subspace, $\dim(\mathcal{V}) \leq \lfloor m/2 \rfloor$. We have unique recovery in $\mathcal{V}$ under $\{i, \tau\}$ if and only if (6) is true.*

---

[3]Lemma 3 (with a different proof) is the main insight of the parallel and independent work of Dokmanić 2019. That work studies the same problem as the present paper, but focuses only on diagonalizable automorphisms. On the other hand, it has the advantage that it considers countably many automorphisms, while here we only consider finitely many.

Proof. ($\Rightarrow$) By Lemma 2 $\mathcal{V} \cap \tau(\mathcal{V}) \subset \mathscr{E}_{\tau,1}$. Hence $\mathcal{V}$ does not intersect $\mathscr{E}_{\tau,\lambda}$ for every $\lambda \neq 1$. This implies (6). ($\Leftarrow$) Suppose first that $\dim(\mathscr{E}_{\tau,1}) \leq m - n$. The subset $\mathbb{X}$ of $\mathrm{Gr}(n, m)$ on which $\mathcal{V} \cap \tau(\mathcal{V}) = 0$ contains the open subset $\mathbb{U}_\tau$ on which $\dim(\mathcal{V} + \tau(\mathcal{V})) = 2n$. Then $\mathbb{U}_\tau$ is non-empty. To see this, note that there exist $2n$ linearly independent eigenvectors $w_1, \ldots, w_{2n}$ of τ such that no more than n of them correspond to the same eigenvalue. Ordering the w_i such that eigenvectors corresponding to the same eigenvalue are placed consecutively, we then define $v_i = w_i + w_{i+n}, \forall i \in [n]$. Then $\mathcal{V} = \mathrm{Span}(v_1, \ldots, v_n) \in \mathbb{U}_\tau$.

Next, suppose that $\dim(\mathscr{E}_{\tau,1}) > m - n$. Since $n \leq \lfloor m/2 \rfloor$ we have $\dim(\mathscr{E}_{\tau,1}) \geq n$. Suppose that $v_1 = \tau(v_2)$ for $v_1, v_2 \in \mathcal{V}$. Let $\mathfrak{B}$ be a basis of $\mathbb{C}^m$ on which τ is represented by a diagonal matrix $T \in \mathbb{C}^{m \times m}$, and let $V \in \mathbb{C}^{m \times n}$ and $V\xi_1, V\xi_2 \in \mathbb{C}^m$ be the corresponding representations of a basis of $\mathcal{V}$, and of v_1, v_2 respectively, with $\xi_1, \xi_2 \in \mathbb{C}^n$. Then the equation $v_1 = \tau(v_2)$ is equivalent to $V\xi_1 = TV\xi_2$. Since $\dim(\mathscr{E}_{\tau,1}) \geq n$ we may assume without loss of generality that the first n diagonal elements of T are equal to 1. Letting $V_1 \in \mathbb{C}^{n \times n}$ be the top $n \times n$ submatrix of V, this implies that $V_1\xi_1 = V_1\xi_2$. Then V_1 is invertible on a non-empty open subset $\mathbb{U}'_\tau$ of $\mathrm{Gr}(n, m)$, on which $v_1 = v_2$. Thus $\mathcal{V} \cap \tau(\mathcal{V}) \subset \mathscr{E}_{\tau,1}, \forall \mathcal{V} \in \mathbb{U}'_\tau$. In conclusion, there is an open set $\mathbb{U} = \mathbb{U}_\tau$ or $\mathbb{U} = \mathbb{U}'_\tau$, such that for any $\mathcal{V} \in \mathbb{U}$ we have $\mathcal{V} \cap \tau(\mathcal{V}) \subset \mathscr{E}_{\tau,1}$, and so we are done by Lemma 2. $\qquad \square$

The extension to multiple automorphisms follows from Lemma 3 and the fact that the intersection of finitely many non-empty open sets of $\mathrm{Gr}(n, m)$ is non-empty and open:

Proposition 1. *Let $\mathcal{T}$ be a finite set of automorphisms of $\mathbb{C}^m$ such that for any $\tau_1, \tau_2 \in \mathcal{T}$ we have that $\tau_1^{-1}\tau_2$ is diagonalizable. Let $\mathcal{V}$ be a generic subspace with $\dim(\mathcal{V}) \leq \lfloor m/2 \rfloor$. We have unique recovery in $\mathcal{V}$ under $\mathcal{T}$ if and only if (6) is true for every $\tau = \tau_1^{-1}\tau_2$ with $\tau_1, \tau_2 \in \mathcal{T}$.*

Even though the invertibility and diagonalizability requirement of Proposition 1 may seem too strong, it is satisfied by our canonical example where $\mathcal{T}$ is the set of $m!$ permutations on the m coordinates of $\mathbb{C}^m$. In fact, more is true:

Proposition 2. *Let $\mathcal{T}$ be the permutations on the m coordinates of $\mathbb{C}^m$. Then $\dim(\mathscr{E}_{\pi,\lambda}) \leq m - \lfloor m/2 \rfloor, \forall \pi \in \mathcal{T}, \forall \lambda \neq 1$. Hence, for generic subspace $\mathcal{V}$ with $\dim(\mathcal{V}) \leq \lfloor m/2 \rfloor$ we have unique recovery in $\mathcal{V}$ under $\mathcal{T}$.*

Proof. This follows from basic structural facts about permutations. Let $\pi \in \mathcal{T}$ be a permutation. Then π is the product of $c \geq 1$ disjoint cycles, say $\pi = \pi_1 \cdots \pi_c$. Suppose that cycle π_i cycles m_i coordinates, i.e., it has length m_i. Since the cycles are disjoint we have $m = \sum_{i=1}^c m_i$. Now, each cycle is diagonalizable with m_i eigenvalues equal to the m_i complex roots of unity, i.e., the roots of the equation $x^{m_i} = 1$. Since the cycles are disjoint, the dimensions of

the eigenspaces of π are counted additively across cycles. Hence for $\lambda \neq 1$ the dimension of $\mathscr{E}_{\pi,\lambda}$ is at most equal to the number of cycles of length at least 2. But the number of such cycles is at most $\lfloor m/2 \rfloor$. Hence $\dim(\mathscr{E}_{\pi,\lambda}) \leq \lfloor m/2 \rfloor$. But $\lfloor m/2 \rfloor \leq m - \lfloor m/2 \rfloor$, i.e., $\dim(\mathscr{E}_{\pi,\lambda}) \leq m - \lfloor m/2 \rfloor$. The rest of the statement is a corollary of Proposition 1. $\quad \square$

2.4. Recovery under arbitrary endomorphisms

2.4.1. UNIQUE RECOVERY FOR ALL POINTS

The arguments that led to Proposition 1 relied heavily on the invertibility of the endomorphisms in $\mathcal{T}$. This is because in that case unique recovery in $\mathcal{V}$ under $\{\tau_1, \tau_2\}$ is equivalent to unique recovery in $\mathcal{V}$ under $\{i, \tau_1^{-1}\tau_2\}$, where i is the identity map. It was this feature that helped us understand the homomorphic sensing property of Definition 1 in terms of $\mathcal{V}$ intersecting its image $\tau_1^{-1}\tau_2(\mathcal{V})$. In turn, the key objects controlling this intersection turned out to be the eigenspaces of $\tau = \tau_1^{-1}\tau_2$ corresponding to eigenvalues different than 1, as per Lemma 3, whose proof however made explicit use of the diagonalizability of τ. As a consequence, generalizing Proposition 1 to arbitrary endomorphisms for which τ_1 might not even be invertible, let alone $\tau_1^{-1}\tau_2$ diagonalizable, is not straightforward.

Example 2. *In unlabeled sensing, a permutation composed with a coordinate projection is in general neither invertible nor diagonalizable. For example, consider a cycle π of length 3, a coordinate projection ρ onto the first two coordinates, and their composition $\rho\pi$:*

$$\pi = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \rho = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \rho\pi = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

First, $\mathrm{rank}(\rho\pi) = 2$ so that $\rho\pi$ is not invertible. Secondly, $\rho\pi$ is nilpotent, i.e., $(\rho\pi)^3 = 0$, and so the only eigenvalue of $\rho\pi$ is zero. This means that $\rho\pi$ is similar to a 3×3 Jordan block of eigenvalue 0, i.e., $\rho\pi$ is far from diagonalizable.

We proceed by developing two devices. The first one is a generalization of Lemma 3 and overcomes the challenge of the potential non-diagonalizability of the endomorphisms.

Lemma 4. *Let $\mathcal{V}$ be a generic subspace with $\dim(\mathcal{V}) \leq \lfloor m/2 \rfloor$, and τ any endomorphism of $\mathbb{C}^m$ for which (6) is true. Then we have unique recovery in $\mathcal{V}$ under $\{i, \tau\}$.*

Proof. (Sketch) The arguments are similar in spirit with those in the proof of Lemma 3 but technically more involved. Let $n = \dim(\mathcal{V})$. The difficult part is when $\dim(\mathscr{E}_{\tau,1}) \leq m - n$, where we prove the existence a non-empty open set $\mathbb{U}$ of $\mathrm{Gr}(n, m)$, such that for every $\mathcal{V} \in \mathbb{U}$ we have $\dim(\mathcal{V} + \tau(\mathcal{V})) = n + \min\{n, \mathrm{rank}(\tau)\}$, which is the maximal dimension that the subspace $\mathcal{V} + \tau(\mathcal{V})$ can have. In analogy with the diagonalizable case, this can be

done by working with the Jordan canonical form of τ and constructing a $\mathcal{V} = \mathrm{Span}(v_1, \ldots, v_n) \in \mathbb{U}$ for which the v_i are suitably paired (generalized) eigenvectors of τ. $\qquad\square$

Our second device overcomes the challenge of potential lack of invertibility. We need some notation. Let τ_1, τ_2 be endomorphisms of $\mathbb{C}^m$ and let ρ be a projection onto $\mathrm{im}(\tau_2)$. Define the variety $\mathcal{Y}_{\rho\tau_1,\tau_2}$ as the set of $w \in \mathbb{C}^m$ for which $\rho\tau_1(w)$ and $\tau_2(w)$ are linearly dependent, i.e.,

$$\mathcal{Y}_{\rho\tau_1,\tau_2} = \{ w : \dim \left(\mathrm{Span}(\rho\tau_1(w), \tau_2(w)) \right) \leq 1 \}. \quad (7)$$

This is indeed a variety because if τ_1, τ_2, ρ are represented by matrices T_1, T_2, P, then $\mathcal{Y}_{\rho\tau_1,\tau_2}$ is defined by the vanishing of all 2×2 minors of the matrix $[PT_1 w \ \ T_2 w]$, which are quadratic polynomials in w. If $w \in \mathcal{Y}_{\rho\tau_1,\tau_2}$ then there exists some $\lambda_w \in \mathbb{C}$ such that either $\tau_2(w) = \lambda_w \rho\tau_1(w)$ or $\rho\tau_1(w) = \lambda_w \tau_2(w)$. Hence $\mathcal{Y}_{\rho\tau_1,\tau_2}$ is the union of all generalized eigenspaces of the endomorphism pairs $(\rho\tau_1, \tau_2)$ and $(\tau_2, \rho\tau_1)$. Note that $\ker(\rho\tau_1 - \tau_2)$ is the generalized eigenspace corresponding to eigenvalue 1, while $\ker(\rho\tau_1), \ker(\tau_2)$ are the generalized eigenspaces of $(\rho\tau_1, \tau_2), (\tau_2, \rho\tau_1)$ respectively, corresponding to eigenvalue 0. In analogy with the automorphism case of Lemma 2 where the eigenspace of eigenvalue 1 was irrelevant for unique recovery, it turns out that in the general case the same is true for the generalized eigenspaces of eigenvalue 1 and 0. Removing their union

$$\mathcal{Z}_{\rho\tau_1,\tau_2} = \ker(\tau_2) \cup \ker(\rho\tau_1) \cup \ker(\rho\tau_1 - \tau_2), \quad (8)$$

from $\mathcal{Y}_{\rho\tau_1,\tau_2}$ yields the quasi-variety

$$\mathcal{U}_{\rho\tau_1,\tau_2} = \mathcal{Y}_{\rho\tau_1,\tau_2} \setminus \mathcal{Z}_{\rho\tau_1,\tau_2}. \quad (9)$$

$\mathcal{U}_{\rho\tau_1,\tau_2}$ plays an analogous role with $\cup_{\lambda\neq1}\mathscr{E}_{\tau,\lambda}$ when τ is an automorphism. More precisely, in analogy with (6), the next theorem shows that the condition that controls homomorphic sensing in general is

$$\dim(\mathcal{V}) \leq \mathrm{codim}\left(\mathcal{U}_{\rho\tau_1,\tau_2} \right). \quad (10)$$

Theorem 1. *Let $\mathcal{T}$ be a finite set of endomorphisms of $\mathbb{C}^m$ such that for every $\tau \in \mathcal{T}$ we have $\mathrm{rank}(\tau) \geq 2n$, for some $n \leq \lfloor m/2 \rfloor$. Then for a general subspace $\mathcal{V}$ of dimension n, we have unique recovery in $\mathcal{V}$ under $\mathcal{T}$ as long as for every $\tau_1, \tau_2 \in \mathcal{T}$ there is a projection ρ onto $\mathrm{im}(\tau_2)$ such that for $\mathcal{U}_{\rho\tau_1,\tau_2}$ defined in (9) condition (10) is true.*

Proof. (Sketch) The key idea for the case $\mathcal{T} = \{\tau_1, \tau_2\}$ is to view $\mathcal{V}$ as a generic n-dimensional subspace of a generic k-dimensional subspace $\mathcal{H}$, where $k = \mathrm{rank}(\tau_2)$. Then $\tau_2|_{\mathcal{H}}$ is an isomorphism from $\mathcal{H}$ onto $\mathrm{im}(\tau_2)$, and so unique recovery in $\mathcal{V}$ under $\{\tau_1, \tau_2\}$ follows from unique recovery in $\mathcal{V}$ under $\{i|_{\mathcal{H}}, \tau_{\mathcal{H}}\}$, with $\tau_{\mathcal{H}} = \left(\tau_2|_{\mathcal{H}}\right)^{-1}\rho\tau_1|_{\mathcal{H}}$ endomorphism of $\mathcal{H}$. By Lemma 4 we are done if

$\dim\left(\mathscr{E}_{\tau_{\mathcal{H}},\lambda}\right) \leq k - n, \ \forall\lambda \neq 1$. Let $\tau_{\mathcal{H}}(w) = \lambda w$, then $\tau_2\left(\tau_2|_{\mathcal{H}}\right)^{-1}\rho\tau_1(w) = \lambda\tau_2(w)$. Now, $\tau_2\left(\tau_2|_{\mathcal{H}}\right)^{-1}\rho = \rho$, thus $\rho\tau_1(w) = \lambda\tau_2(w)$. Hence, $\mathscr{E}_{\tau_{\mathcal{H}},\lambda} \subset \left(\mathcal{U}_{\rho\tau_1,\tau_2} \cap \mathcal{H}\right)$ and the rest follows from dimension considerations. $\qquad\square$

In unlabeled sensing the endomorphisms in $\mathcal{T}$ have the form $\rho\pi$, where π is a permutation and ρ is a coordinate projection. Then as per Theorem 1 if $\dim(\mathcal{V}) \leq \mathrm{codim}\left(\mathcal{U}_{\rho_2\rho_1\pi_1,\rho_2\pi_2}\right)$ one has unique recovery in $\mathcal{V}$ under $\{\rho_1\pi_1, \rho_2\pi_2\}$. Furthermore, via a combinatorial algebraic-geometric argument we obtain a convenient lower bound on $\mathrm{codim}\left(\mathcal{U}_{\rho_2\rho_1\pi_1,\rho_2\pi_2}\right)$:

Proposition 3. *Let π_1, π_2 be permutations on the m coordinates of $\mathbb{C}^m$ and ρ_1, ρ_2 coordinate projections. For $\mathcal{U}_{\rho_2\rho_1\pi_1,\rho_2\pi_2}$ defined in (9) we have*

$$\lfloor \mathrm{rank}(\rho_2)/2 \rfloor \leq \mathrm{codim}\left(\mathcal{U}_{\rho_2\rho_1\pi_1,\rho_2\pi_2}\right). \quad (11)$$

As a consequence of Theorem 1 and Proposition 3 one has unique recovery in the unlabeled sensing case as long as the dimension of $\mathcal{V}$ does not exceed half the number of the coordinates preserved by each coordinate projection. This is precisely the result of Unnikrishnan et al. 2018 which they obtained by attacking the problem directly via ingenious yet complicated combinatorial arguments. Even though our proof is not necessarily less complicated, it has the advantage of using a framework that generalizes relatively easily. For example, one may consider entry-wise sign corruptions on top of coordinate projections and permutations. In such a case, it is not hard to show that for the same condition as for unlabeled sensing one has unique recovery up to a sign. In general, even though analytically computing $\mathrm{codim}\left(\mathcal{U}_{\rho\tau_1,\tau_2}\right)$ may be challenging, performing this computation in an algebraic geometry software environment such as `Macaulay2` is in principle straightforward.

2.4.2. UNIQUE RECOVERY FOR GENERIC POINTS

Often the requirement that *every* $v \in \mathcal{V}$ is uniquely recoverable is unnecessarily strict. Instead, it may be of interest to ask whether unique recovery holds true for a generic $v \in \mathcal{V}$. In such a situation a less demanding technical analysis gives unique recovery under much weaker conditions:

Theorem 2. *Let $\mathcal{T}$ be a finite set of endomorphisms of $\mathbb{C}^m$. Then a generic point v inside a generic subspace $\mathcal{V}$ of dimension n is uniquely recoverable in $\mathcal{V}$ under $\mathcal{T}$ as long as 1) $\mathrm{rank}(\tau) \geq n + 1$ for every $\tau \in \mathcal{T}$, and 2) no two endomorphisms in $\mathcal{T}$ are a scalar multiple of each other.*

Proof. (Sketch) Let $V \in \mathbb{C}^{m \times n}$ be a basis of $\mathcal{V}$. If $\tau_1(v_1) = \tau_2(v_2)$ then $\tau_2(v_2) \in \tau_1(\mathcal{V})$ and so $\mathrm{rank}([T_1 V \ \ T_2 V \xi]) \leq n$ for $\xi \in \mathbb{C}^n$ with $v_2 = V\xi$. The proof then proceeds by exhibiting, for $\tau_1 \neq \tau_2$, a $\mathcal{V}$ and a ξ for which $\mathrm{rank}([T_1 V \ \ T_2 V \xi]) = n + 1$. This implies that

for generic $\mathcal{V}$ and $v_2 \in \mathcal{V}$, $\tau_1(v_1) = \tau_2(v_2)$ for $v_1 \in \mathcal{V}$ only when $\tau_1 = \tau_2$. In that case $v_1 - v_2 \in \ker(\tau_1)$, and Lemma 1 implies that $v_1 - v_2 = 0$. $\qquad\square$

A consequence of Theorem 2 is the unique recovery of a generic vector in the unlabeled sensing case as soon as the coordinate projections preserve at least $n + 1$ entries:

Corollary 1. *Let $\mathcal{T}$ be the set of endomorphisms of $\mathbb{C}^m$ such that every $\tau \in \mathcal{T}$ has the form $\tau = \rho\pi$, where π is a permutation and ρ a coordinate projection. Then for a generic subspace $\mathcal{V}$ of dimension n, and a generic $v \in \mathcal{V}$, we have unique recovery of v in $\mathcal{V}$ under $\mathcal{T}$, as long as $\mathrm{rank}(\rho) \geq n + 1$ for every $\rho\pi \in \mathcal{T}$.*

3. Algorithms and Application

3.1. Branch-and-bound for unlabeled sensing

In this section we propose a globally optimal method for the unlabeled sensing problem, by minimizing (3) via a dynamic-programming based branch-and-bound scheme. Both ingredients are standard, and we just describe how to combine them; see Emiya et al. 2014; Yang et al. 2016 for transparent discussions of branch-and-bound in related contexts. Set $f(x, S) = \|y - SAx\|_2$, with $x \in \mathbb{R}^n$ and S a selection matrix. As branching over the space of selection/permutation matrices S is known to be inefficient (Li & Hartley, 2007), the crucial aspect of our approach is to branch only over the space of x, while relying on a local computation of the optimal S, say S_x, given x. Here is where dynamic programming comes into play: Haghighatshoar & Caire 2018 showed that if there exists an order-preserving S_x such that $f(x, S_x) = \min_S \|y - SAx\|_2$, then S_x can be computed via dynamic programming[4] at a complexity $\mathcal{O}(mk)$. At first sight this does not generalize to any y, A, x as none of the minimizers over S is expected to be order-preserving. However, if we order y, Ax in descending order to obtain say $y^\downarrow, (Ax)^\downarrow$, then 1) there is an order-preserving selection matrix S'_x such that $\|y^\downarrow - S'_x(Ax)^\downarrow\|_2 = \min_S \|y^\downarrow - S(Ax)^\downarrow\|_2$, and 2) S_x can be easily obtained from S'_x. In conclusion, given x we can compute S_x in $\mathcal{O}(mk)$; this is in sharp contrast to other linear assignment algorithms such as the Hungarian algorithm, most of which have complexity $\mathcal{O}(m^3)$ (Burkard et al., 2009). Finally, our strategy becomes that of computing an upper bound of f in a hypercube with center x_0 via alternating minimization between x and S, initialized at x_0. Computing a tight lower bound ℓ of f for a given hypercube is challenging and our choice here is a crude one: $\ell = \|y - S_{x_0}Ax_0\|_2 - \sigma_1(A)\epsilon$, where ϵ is half the hypercube diagonal and $\sigma_1(A)$ is the largest singular value of A. We refer to this as `Algorithm-A`.

---

[4]That such an assignment problem can be solved via dynamic programming was already known by Aggarwal et al. 1992.

3.2. A robust version of Elhami et al. 2017

It turns out that the dynamic programming *trick* of §3.1 is also the key to a robust version of the theoretical algorithm of Elhami et al. 2017: we randomly select a sub-vector $\bar{y}$ of y of length n, and for each A_i out of the $m!/(m-n)!$ many $n \times n$ matrices that can be made by concatenating different rows of A in any order we let $x_i = A_i^\dagger \bar{y}$. We then use dynamic programming to select the x_i with the lowest assignment error $\min_S \|y - SAx_i\|_2$. This is an algorithm of complexity $\mathcal{O}(km^{n+1})$, we call it `Algorithm-B`.

3.3. Evaluation on synthetic data

We compare the proposed 1) `Algorithm-A` of §3.1 and 2) `Algorithm-B` of §3.2 to other state-of-the-art methods (§1.1), using normally distributed A, x, ε with $n = 3, m = 100$ and $\sigma = 0.01$ for the noise. For shuffled linear regression ($k = m$) we compare with 3) `Slwasky19` that solves (2), 4) `Tsakiris18` which is the algebraic-geometric method of Tsakiris et al. 2018 and 3) `Abid18` which performs alternating minimization on (1) via least-squares and sorting (Abid & Zou, 2018)[5]. For unlabeled sensing ($k \leq m$) we compare with 4) `Haghighatshoar18` (Haghighatshoar & Caire, 2018). As seen from Fig. 1 the proposed methods perform uniformly better and often by a large margin than the other methods, when tested in their robustness against the percentage of shuffled data, outlier ratio and noise level. In particular, we see that for the unlabeled sensing problem of Figs. 1b-1d `Algorithm-A` and `Algorithm-B` are the only working solutions. Encouraging as these results may be, we do note that an important weakness of these methods is their scalability: for `Algorithm-B` this is more of an inherent issue due to its brute-force nature, while for `Algorithm-A` it is due to its naive lower bounding scheme: the consequence of it being far from tight manifests itself at higher dimensions ($n \geq 4$) or large outlier ratios ($k \ll m$), in which the method becomes too slow: it runs[6] in 1sec for $k = m = 100$, 30sec for $k = 80$ and 5min for $k = 50$. In contrast, `Algorithm-B` is immune to k and runs in about 40sec.

3.4. Application to image registration

Registering point sets P, Q between two images is a classical problem in computer vision. Assuming that P, Q are related by an affine transformation T and that each point in P (*model set*) has a counterpart in Q (*scene set*) (Lian et al., 2017), jointly searching for the affine transformation and the registration can be done by minimizing the func-

---

[5]The *soft-EM* algorithm of Abid & Zou 2018 consistently fails in our experiment, thus we do not include it in the figure.

[6]In this experiment `Algorithm-A` stops splitting a hypercube when a *depth* 6 for that hypercube has been reached. Run on an Intel(R) i7-8650U, 1.9GHz, 16GB machine.
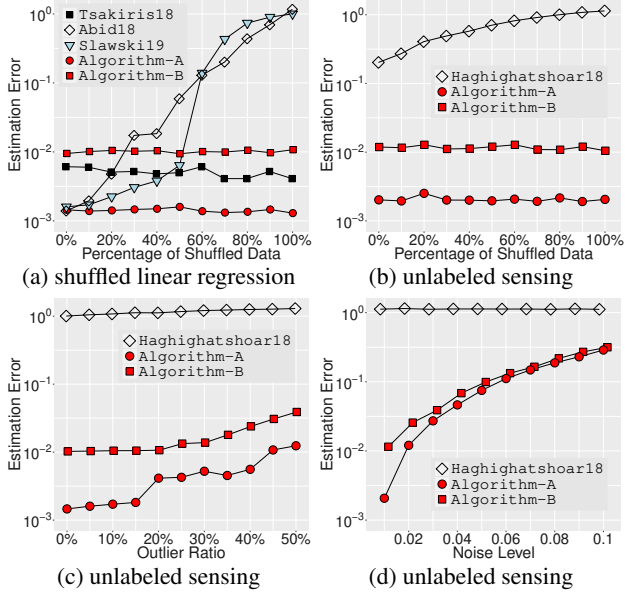
Figure 1: Relative error vs. % of shuffled data, outlier ratio $(1 - k/m)$ and noise level (σ). $n = 3, m = 100, k = 80, \sigma = 0.01$, 1000 trials. Proposed algorithms in red.

tion $F(T, S) = \|PT - SQ\|_{\mathrm{F}}$, where $Q \in \mathbb{R}^{m \times 2}, P \in \mathbb{R}^{k \times 3}, T \in \mathbb{R}^{3 \times 2}$, with homogeneous coordinates used for P, and S is a selection matrix. This is a matrix version of the unlabeled sensing objective function (3). Our contribution here is to adjust algorithm of §3.1 to solve the image registration problem. This involves branching over a 6-dimensional space to compute T, i.e., $n = 6$. This does not contradict the remark of the previous section regarding scalability: the key here is that each point correspondence imposes two constraints on T (as opposed to one constraint in the general case), so that, loosely speaking, the *effective* outlier ratio is $1 - 2k/m$ (as opposed to $1 - k/m$). As we will soon see, this has a crucial effect on performance. Finally, as dynamic programming is not applicable to obtain S given T, we employ a standard linear assignment algorithm of complexity $\mathcal{O}(m^3)$ (Jonker & Volgenant, 1987). We refer to this algorithm as Algorithm-C.

We compare 1) Algorithm-C with state-of-the-art image registration techniques, i.e., 2) CPD (Myronenko & Song, 2010), 3) GMMREG (Jian & Vemuri, 2011), and 4) APM (Lian et al., 2017), using a subset of the benchmark datasets used by Lian et al. 2017. Since APM is the most competitive among these last three, we let it run to convergence with a tolerance parameter of 0.1 and set its running time as a time budget for our method; CPD and GMMREG are local methods and they run very fast. When the affine transformation is a rotation (Fig. 2a) CPD, GMMREG only work for small angles, while they fail for general affine transformations (Figs. 2b-2c). On the other hand, Algorithm-C performs comparably to APM with the following twist in Fig. 2c: when the outlier ratio is small, APM converges
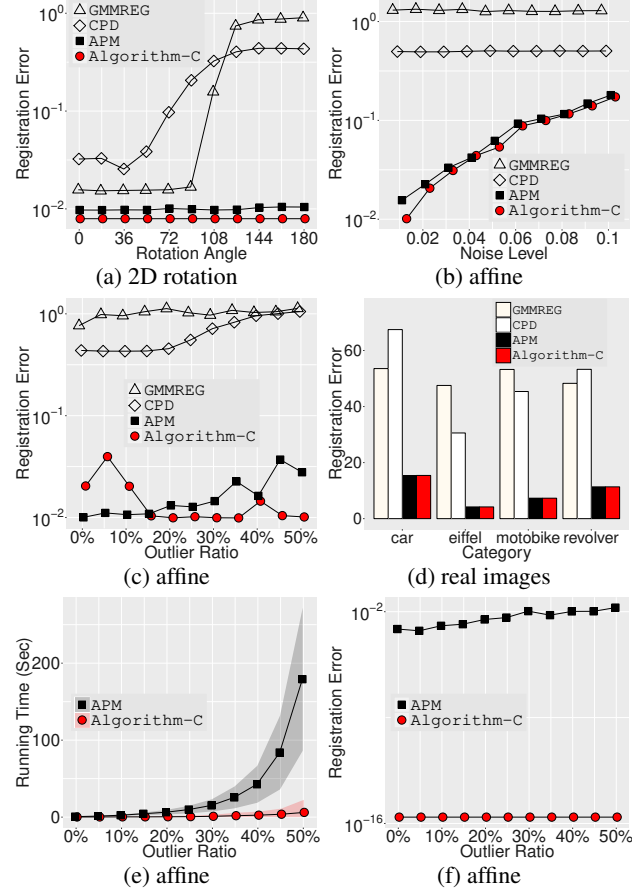

Figure 2: Image registration using the synthetic benchmark dataset *The Chinese Character* (Chui & Rangarajan, 2003) (2a-2c,2e-2f,100 trials) and the same collection of real images used in Lian et al. 2017 (2d). Proposed method in red.

very quickly resulting to an inadequate time budget for our method. Conversely, when the outlier ratio is large, APM's accuracy becomes inadequate while it is slow enough for our method to perform even better than for fewer outliers. These running times for APM are shown in Fig. 2e where the same experiment is run for noiseless data: APM still uses a tolerance of 0.1 while to make a point we set the tolerance of our method to zero and let it terminate. As seen in Figs. 2e-2f, our method terminates significantly faster than APM for large outlier ratios, suggesting that its branch-and-bound structure may have an advantage over that of APM.

Acknowledgement

References

Abid, A. and Zou, J. Stochastic em for shuffled linear regression. Technical report, arXiv:1804.00681v1 [stat.ML], 2018.

Abid, A., Poon, A., and Zou, J. Linear regression with shuffled labels. Technical report, arXiv:1705.01342v2 [stat.ML], 2017.

Aggarwal, A., Bar-Noy, A., Khuller, S., Kravets, D., and Schieber, B. Efficient minimum cost matching using quadrangle inequality. In *Annual Symposium on Foundations of Computer Science*, pp. 583–592, 1992.

Balakrishnan, A. On the problem of time jitter in sampling. *IRE Transactions on Information Theory*, 8(3):226–236, 1962.

Bürgisser, P. and Cucker, F. The geometry of numerical algorithms. *Springer Science & Business Media*, 349, 2013.

Burkard, R. E., Dell'Amico, M., and Martello, S. *Assignment Problems*. Springer, 2009.

Chui, H. and Rangarajan, A. A new point matching algorithm for non-rigid registration. *Comput. Vis. Image Underst.*, 89(2-3):114–141, 2003.

David, P., DeMenthon, D., Duraiswami, R., and Samet, H. Softposit: Simultaneous pose and correspondence determination. *International Journal of Computer Vision*, 59 (3):259–284, 2004.

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

Elhami, G., Scholefield, A., Haro, B. B., and Vetterli, M. Unlabeled sensing: Reconstruction algorithm and theoretical guarantees. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 4566–4570, 2017.

Emiya, V., Bonnefoy, A., Daudet, L., and Gribonval, R. Compressed sensing with unknown sensor permutation. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1040–1044, 2014.

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

Haghighatshoar, S. and Caire, G. Signal recovery from unlabeled samples. *IEEE Transactions on Signal Processing*, 66(5):1242–1257, 2018.

Hartshorne, R. *Algebraic Geometry*. Springer, 1977.

Hsu, D. J., Shi, K., and Sun, X. Linear regression without correspondence. In *Advances in Neural Information Processing Systems (NIPS)*, pp. 1531–1540, 2017.

Jian, B. and Vemuri, B. C. Robust point set registration using gaussian mixture models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(8):1633–1645, 2011.

Jonker, R. and Volgenant, A. A shortest augmenting path algorithm for dense and sparse linear assignment problems. *Computing*, 38(4):325–340, 1987.

Keller, L., Siavoshani, M. J., Fragouli, C., Argyraki, K., and Diggavi, S. Identity aware sensor networks. In *IEEE INFOCOM*, pp. 2177–2185, 2009.

Lahiri, P. and Larsen, M. D. Regression analysis with linked data. *Journal of the American Statistical Association*, 100(469):222–230, 2005.

Li, H. and Hartley, R. The 3d-3d registration problem revisited. In *International Conference on Computer Vision (ICCV)*, pp. 1–8, 2007.

Lian, W., Zhang, L., and Yang, M.-H. An efficient globally optimal algorithm for asymmetric point mathcing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(7):1281–1293, 2017.

Marques, M., Stosic, M., and Costeira, J. Subspace matching: Unique solution to point matching with geometric constraints. In *International Conference on Computer Vision (ICCV)*, pp. 1288–1294, 2009.

Myronenko, A. and Song, X. Point set registration: coherent point drift. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(12):2262–2275, 2010.

Narayanan, A. and Shmatikov, V. Robust de-anonymization of large sparse datasets. In *IEEE Symposium on Security and Privacy*, pp. 111–125, 2008.

Pananjady, A., Wainwright, M. J., and Courtade, T. A. Linear regression with shuffled data: Statistical and computational limits of permutation recovery. In *54th IEEE Annual Allterton Conference on Communication, Control and Computing*, pp. 417–424, 2016.

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

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

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

Poore, A. B. and Gadaleta, S. Some assignment problems arising from multiple target tracking. *Mathematical and Computer Modelling*, 43(9):1074 – 1091, 2006.

Roman, S. *Advanced Linear Algebra*. Springer, 2008.

Shi, X., Li, X., and Cai, T. Spherical regression under mismatch corruption with application to automated knowledge translation. Technical report, arXiv:1810.05679v1 [stat.ME], 2018.

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

Song, X., Choi, H., and Shi, Y. Permuted linear model for header-free communication via symmetric polynomials. In *International Symposium on Information Theory (ISIT)*, 2018.

Tsakiris, M. C. Eigenspace conditions for homomorphic sensing. Technical report, arXiv:1812.07966v3 [math.CO], 2019.

Tsakiris, M. C., Peng, L., Conca, A., Kneip, L., Shi, Y., and Choi, H. An algebraic-geometric approach to shuffled linear regression. Technical report, arXiv:1810.05440v1 [cs.LG], 2018.

Unnikrishnan, J., Haghighatshoar, S., and Vetterli, M. Unlabeled sensing: solving a linear system with unordered measurements. In *53rd IEEE Annual Allerton Conference on Communication, Control and Computing*, 2015.

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

Yang, J., Li, H., Campbell, D., and Jia, Y. Go-icp: A globally optimal solution to 3d icp point-set registration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(11):2241–2254, 2016.