

An Algebraic-Geometric Approach for Linear Regression Without Correspondences

Manolis C. Tsakiris¹, Liangzu Peng, Aldo Conca, Laurent Kneip²,
Yuanming Shi³, *Member, IEEE*, and Hayoung Choi

Abstract—Linear regression without correspondences is the problem of performing a linear regression fit to a dataset for which the correspondences between the independent samples and the observations are unknown. Such a problem naturally arises in diverse domains such as computer vision, data mining, communications and biology. In its simplest form, it is tantamount to solving a linear system of equations, for which the entries of the right hand side vector have been permuted. This type of data corruption renders the linear regression task considerably harder, even in the absence of other corruptions, such as noise, outliers or missing entries. Existing methods are either applicable only to noiseless data or they are very sensitive to initialization or they work only for partially shuffled data. In this paper we address these issues via an algebraic geometric approach, which uses symmetric polynomials to extract permutation-invariant constraints that the parameters $\xi^* \in \mathbb{R}^n$ of the linear regression model must satisfy. This naturally leads to a polynomial system of n equations in n unknowns, which contains ξ^* in its root locus. Using the machinery of algebraic geometry we prove that as long as the independent samples are generic, this polynomial system is always consistent with at most $n!$ complex roots, regardless of any type of corruption inflicted on the observations. The algorithmic implication of this fact is that one can always solve this polynomial system and use its most suitable root as initialization to the Expectation Maximization algorithm. To the best of our knowledge, the resulting method is the first working solution for small values of n able to handle thousands of fully shuffled noisy observations in milliseconds.

Index Terms—Linear regression without correspondences, linear regression with shuffled data, shuffled linear regression,

Manuscript received November 1, 2018; revised September 25, 2019; accepted February 10, 2020. Date of publication February 28, 2020; date of current version July 14, 2020. This work was supported in part by INDAM through the Membership of the Research Group GNSAGA, in part by the Korea Research Fellowship Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science and ICT under Grant 2019H1D3A1A01102728, in part by the Natural Science Foundation of Shanghai under Grant 19ZR1434000, and in part by the National Nature Science Foundation of China under Grant 61950410612 and Grant 61601290. This article was presented in part at ICASSP'19. (*Corresponding author: Manolis C. Tsakiris.*)

Manolis C. Tsakiris, Liangzu Peng, Laurent Kneip, and Yuanming Shi are with the School of Information Science and Technology, ShanghaiTech University, Shanghai 201210, China (e-mail: mtsakiris@shanghaitech.edu.cn).

Aldo Conca is with the Department of Mathematics, University of Genova, 16146 Genova, Italy.

Hayoung Choi was with the School of Information Science and Technology, ShanghaiTech University, Shanghai 201210, China. He is now with the Research Institute of Mathematics, Seoul National University, Seoul 08826, South Korea.

Communicated by N. Kiyavash, Associate Editor for Statistical Learning.

Color versions of one or more of the figures in this article are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TIT.2020.2977166

unlabeled sensing, homomorphic sensing, expectation maximization, algebraic geometry.

I. INTRODUCTION

IN THE span of more than 200 years since the work of Legendre [1] and Gauss [2], linear regression has grown to be a cornerstone of statistics, with applications in almost every branch of science and engineering that involves computing with data. In its simplest form, classical linear regression considers a data model whose output y is a linear function of known functions of the input data u . More precisely, given correspondences $\{u_j, y_j\}_{j=1}^m$ with $u_j \in \mathbb{R}^s$, $y_j \in \mathbb{R}$ and known functions $a_i : \mathbb{R}^s \rightarrow \mathbb{R}$, one seeks to find real numbers $\xi = [\xi_1, \dots, \xi_n]^\top \in \mathbb{R}^n$ such that¹

$$y_j \approx a_j^\top \xi, \quad \forall j = 1, \dots, m, \quad (1)$$

$$a_j := [a_1(u_j), \dots, a_n(u_j)]^\top. \quad (2)$$

Between the least-squares solution of Gauss and modern approaches designed to deal with highly corrupted data [3]–[6], a literature too vast to enumerate has been devoted to progressively more complicated versions of the linear regression problem [7].

A. Linear Regression Without Correspondences

In this paper we are interested in a particular type of data corruption, which is *lack of correspondences*. In such a case, one is given the input samples, or more precisely functions of the input samples as in (2), i.e.,

$$A = [a_1, \dots, a_m]^\top \in \mathbb{R}^{m \times n}, \quad (3)$$

and a *shuffled* version $y = [y_{\pi(1)}, \dots, y_{\pi(m)}]^\top \in \mathbb{R}^m$, of the observations y_1, \dots, y_m , where π is an unknown permutation of $[m] = \{1, \dots, m\}$. This problem of *linear regression without correspondences* [8] also known as *linear regression with shuffled data* [9], *shuffled linear regression* [28], [29], *unlabeled sensing* [10] or *permuted linear model* [11], can be stated in the absence of any other data corruptions as follows:

Problem 1: Suppose we are given a matrix $A \in \mathbb{R}^{m \times n}$ with $m > n$, and a vector $y = (\Pi^*)^\top A \xi^* \in \mathbb{R}^m$, where $\xi^* \in \mathbb{R}^n$ is some vector and Π^* an $m \times m$ permutation matrix.

¹A classical example is when y is a polynomial function of u , in which case the a_i represent the monomials that appear in the polynomial and ξ_i are their coefficients to be fitted.

We wish to efficiently compute ξ^* when Π^* is unknown and $m \gg n$. In other words, without knowing Π^* , we want to solve the linear system

$$\Pi^* y = Ax. \quad (4)$$

Problem 1 arises in a wide variety of applications, such as 1) computer vision, e.g., multi-target tracking [15] and pose/correspondence estimation [16], [17], 2) record linkage [18]–[20], 3) biology, e.g., for cell tracking [21], genome-assembly [22], and identical tokens in signaling pathways [23], 4) communication networks, e.g., for data de-anonymization [24], [25], and low-latency communications in Internet-Of-Things networks [11], 5) signal processing, e.g., when dealing with signals sampled in the presence of time jitter [26], or in applications where a spatial field is being inaccurately sampled by a moving sensor [13], [14].

B. Prior Art

Over the past years there has been a considerable amount of work on instances of Problem 1 that come with additional structure in diverse contexts, e.g., see the excellent literature reviews in [9], [10], [18]. Nevertheless, it has only been until very recently that the problem of shuffled linear regression has been considered in its full generality. In fact, the main achievements so far have been concentrating on a theoretical understanding of the conditions that allow unique recovery of ξ^* or Π^* ; see [8]–[11], [18], [19], [28], [29], [31]–[36].

Letting A be drawn at random from any continuous probability distribution, [10] proved that any such ξ^* can be uniquely recovered with probability 1 as long as² $m \geq 2n$. If on the other hand $m < 2n$, then ξ^* is not unique with probability 1. Further considering additive random noise on y , the authors in [8] established lower bounds on the SNR, below which for any estimator there is a ξ^* whose estimation error is *large*. With A drawn from a normal distribution, ξ^* unknown but fixed and y corrupted by additive random noise, [9] showed that, as long as the SNR exceeds a threshold, Π^* coincides with high probability with the *Maximum Likelihood Estimator* (MLE), which they defined as (also considered in [28], [29])

$$(\hat{\Pi}_{\text{ML}}, \hat{x}_{\text{ML}}) = \underset{\Pi, x}{\operatorname{argmin}} \|\Pi y - Ax\|_2, \quad (5)$$

where Π in (5) is constrained to be a permutation matrix. If on the other hand the SNR is not large enough, $\hat{\Pi}_{\text{ML}}$ differs from Π^* with high probability, in agreement with the results of [8]. This was further complemented by [10], which showed that \hat{x}_{ML} is locally stable under noise, in the sense that as the SNR tends to infinity \hat{x}_{ML} tends to ξ^* . However, according to [28], for SNR fixed, \hat{x}_{ML} is asymptotically inconsistent. Interestingly, if the data are only sparsely shuffled, the work of [18] proved that ξ^* coincides with the optimal solution of a robust ℓ_1 regression problem.

On the algorithmic front of solving Problem 1 much less has been achieved. In [10] the authors write, *...although we*

showed that recovery of the unknown ξ^ is possible from unlabeled measurements, we do not study the problem of designing an efficient algorithm to recover ξ^* . Our solution is to consider all possible permutations of the unlabeled observations which might be prohibitively complex in large dimensional problems.* Indeed, this involves checking whether the linear system $\Pi y = Ax$ is consistent for each permutation Π among the $m!$ permutations of the m entries of y , yielding a complexity $\mathcal{O}((m!)mn^2)$. A more efficient algorithm is that of [8], which is able to reduce the complexity as a function of m to a factor of at least m^7 . However, as the authors of [8] write, their algorithm *strongly exploits the assumption of noiseless measurements and is also very brittle and very likely fails in the presence of noise*; the same is true for the $\mathcal{O}(m^n)$ complexity algorithm of [38]. Finally, the authors in [8] write *we are not aware of previous algorithms for the average-case problem in general dimension n .* In the same paper a $(1 + \epsilon)$ approximation algorithm with theoretical guarantees and of complexity $\mathcal{O}((m/\epsilon)^n)$ is proposed, which however, *is not meant for practical deployment, but instead is intended to shed light on the computational difficulty of the least squares problem* (5). Indeed, as per [9] for³ $n > 1$ this is an NP-hard problem.⁴ On the other hand, the approach that seems to be the predominant one in terms of practical deployment is that of solving (5) via alternating minimization [29]: given an estimate for ξ^* one computes an estimate for Π^* via sorting; given an estimate for Π^* one computes an estimate for ξ^* via ordinary least-squares. However, this approach is very sensitive to initialization and generally works only for partially shuffled data; see [29] for a *soft* variation of this alternating scheme. In conclusion and to the best of our knowledge, there does not yet exist an algorithm for solving Problem 1 that is i) theoretically justifiable, ii) efficient and iii) robust to even mild levels of noise.

C. Contributions

In this work, we contribute to the study of Problem 1 on both theory (§II) and algorithms (§III). On the theoretical level, we show that for generic noiseless data y, A , there is a unique solution Π^*, ξ^* , as soon as $m > n$. We show that ξ^* is contained in the root locus of a system of n polynomial equations in n unknowns. Using tools from algebraic geometry, we show that this polynomial system is always consistent with at most $n!$ complex solutions, regardless of any noise that may further corrupt the observations y . Furthermore, we show that the euclidean distance of ξ^* from the root locus of the noisy system is bounded by a polynomial function of the noise which vanishes for zero noise. The algorithmic implication of these results is that (under the genericity assumption) one can always solve the noisy polynomial system and use a simple criterion to identify its most appropriate root to be used as initialization for computing the MLE estimator (5) via alternating minimization. Even though solving the polynomial system

³The case $n = 1$ is well understood and solved at a complexity of $\mathcal{O}(m \log(m))$ by sorting ([9], [28]).

⁴While the present paper was under review, [35] proposed an empirical algorithm based on branch & bound and dynamic programming.

²While the present paper was under review, this result was greatly generalized in [34]–[36].

entails in principle exponential complexity in n , its complexity in m is linear for any n . Furthermore, we use methods from automatic algebraic-geometric solver generation to obtain highly efficient solvers for $n = 3, 4$. Deriving the SNR rate of the resulting estimator is a challenging problem that we do not pursue in this paper. Even so, we provide empirical evidence according to which for $n \leq 5$ our approach is the first working solution to linear regression without correspondences that remains stable under noise and has manageable complexity. As an example, for $n = 4$, $m = 10,000$, and 1% additive noise, our method computes in 313 milliseconds a solution that is within 0.6% error from the ground truth.

II. THEORETICAL CONTRIBUTIONS

The main contribution of this paper is to develop the theory to an algebraic geometric approach for solving Problem 1. The key idea, described in detail in §II-B, uses symmetric power-sum polynomials to eliminate the unknown permutation, thus resulting in a polynomial system \mathcal{P} of n equations in n unknowns. These polynomials were considered implicitly in the statistical approach of [28], towards constructing a *self-moment* estimator. The authors of that paper wrote, *...in fact there may not be a solution to the system*, which led them to compute their estimator via gradient descent on a highly non-convex objective function, a procedure lacking theoretical guarantees and very sensitive to initialization. The geometric significance of the polynomial system \mathcal{P} was also recognized by the last two authors of the present paper in the short conference paper [11], but important questions such as

- 1) “does \mathcal{P} have finitely many solutions?” or
- 2) “does \mathcal{P} have any solutions in the presence of noise?”,

were left as open problems.

In this paper we answer these two questions in the affirmative in §II-C, via Theorems 2 and 3-4 respectively. The main message is that if the input data $A \in \mathbb{R}^{m \times n}$ are generic (to be made precise in §II-A), then using n power-sum polynomials of degrees $1, 2, \dots, n$ as constraints, defines an algebraic variety that consists of at most $n!$ points, among which lies the solution ξ^* of the shuffled linear regression Problem 1. In addition, the same conclusion holds true in the case where the observation vector y has undergone any type of corruption: the variety defined by the *noisy* polynomials is non-empty and consisting of at most $n!$ complex points. This guarantees that the equations are almost always consistent even if the data are imperfect, which enables algorithmic development as discussed in §III. Even though we do provide a bound on the effect of the noise on the solutions of the algebraic equations \mathcal{P} , deriving the SNR rate for the resulting estimator involves further challenges. Hence we leave such a statistical analysis as an open problem, which we hope that the present algebraic developments will facilitate in solving.

The proofs of Theorems 2-4, given in §II-D, require a thorough understanding of the notion of dimension of polynomial systems of equations. We describe the necessary notions in the series of self-contained appendices A-E in an expository style for the benefit of the reader who is not familiar with algebraic geometry.

A. Genericity and Well-Posedness

Before we are in a position to state our main results, i.e., Theorems 2-4 described in §II-C, we need to clarify what we mean when we refer to A, y as being generic (§II-A.1), and also settle the well-posedness of Problem 1 (§II-A.2).

1) *The Notion of Generic A, y* : We start with an example.

Example 1: Consider the 2×2 matrix

$$B = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}, \quad (6)$$

where the entries of B are real numbers. Then B is invertible if and only if $b_{11}b_{22} - b_{12}b_{21} \neq 0$. Now consider the polynomial ring $\mathbb{R}[x_{11}, x_{12}, x_{21}, x_{22}]$ in four variables, with each of them corresponding to an entry of B . The equation

$$x_{11}x_{22} - x_{12}x_{21} = 0 \quad (7)$$

defines a hypersurface $\mathcal{V}(x_{11}x_{22} - x_{12}x_{21})$ of $\mathbb{R}^4 \cong \mathbb{R}^{2 \times 2}$, and $B \in \mathcal{V}(x_{11}x_{22} - x_{12}x_{21})$ if and only if B is non-invertible. This hypersurface has measure zero, say, under the standard normal distribution of \mathbb{R}^4 , and hence if one samples B at random from this distribution, B will lie outside of $\mathcal{V}(x_{11}x_{22} - x_{12}x_{21})$ with probability 1. We express this by saying that “if B is a generic 2×2 matrix, then B is invertible”.

As Example 1 suggests, we usually attach the attribute *generic* to an object θ (matrix B in Example 1) with respect to some property \mathcal{P} of θ (invertibility of B in Example 1). We say that “if θ is generic then \mathcal{P} is true”, and mean that the set of objects for which \mathcal{P} is not true forms a zero-measure set of the underlying space that parametrizes that object under some continuous probability distribution. Hence sampling θ at random from that probability distribution results in θ having the property \mathcal{P} with probability 1. Finally, if there are finitely many properties $\mathcal{P}_1, \dots, \mathcal{P}_t$ of interest with regard to the object θ , and if θ is generic with respect to each of the \mathcal{P}_i , then θ is generic with respect to all of them; this follows from the fact that the union of finitely many zero-measure sets is a zero-measure set. The connection between algebraic geometry and measure theory that the reader should keep in mind for our purpose here, is that algebraic subvarieties of irreducible varieties (see Definition A1) have zero measure and that the union of finitely many algebraic subvarieties is an algebraic subvariety.

2) *Uniqueness of Π^*, ξ^** : Our first result, Theorem 1, guarantees that Problem 1 is well-posed for generic data, in which case it makes sense to talk about a unique permutation Π^* and a unique solution ξ^* .

Theorem 1: Suppose that $m > n$. Then as long as $A \in \mathbb{R}^{m \times n}$ is a generic matrix and y the permutation of a generic vector in $\mathcal{R}(A)$, Π^* and ξ^* in Problem 1 are unique.

It is interesting to compare Theorem 1 with a simplified version of the main result of Unnikrishnan-Haghighatshoar-Vetterli ([10]), stated next for convenience.

Theorem (U-H-V): Suppose that $m \geq 2n$. Then as long as $A \in \mathbb{R}^{m \times n}$ is a generic matrix and y the permutation of any vector in $\mathcal{R}(A)$, ξ^* in Problem 1 is unique.

Both Theorems 1 and U-H-V are concerned with a generic matrix A and the permutation y of some vector v in $\mathcal{R}(A)$. In Theorem 1 the vector v is taken to be a generic vector in

$\mathcal{R}(A)$, and as it turns out it is enough that $m > n$ for v to be uniquely defined from the data A, y . On the other hand, in Theorem U-H-V the vector v is allowed to be *any* vector in $\mathcal{R}(A)$. This is a considerably more difficult setting, and the remarkable proof of [10] reveals that v is still uniquely defined from the data, as long as now⁵ $m \geq 2n$.

Finally, we note that in the setting of Theorem U-H-V unique recovery of the permutation Π^* is in principle not possible, as per Theorem 10 in [10]. Instead, one has to either allow for the observed vector y to be generic (i.e., the permutation of a generic vector of $\mathcal{R}(A)$) in which case Π^* is uniquely recoverable by Theorem 1, or consider unique recovery with high probability, which is indeed possible even when y is corrupted by noise [9].

B. Eliminating Π^* via Symmetric Polynomials

To describe the main idea of the algebraic-geometric approach to solving Problem 1, let $\mathbb{R}[z] := \mathbb{R}[z_1, \dots, z_m]$ be the ring of polynomials with real coefficients over variables $z := [z_1, \dots, z_m]^\top$. A polynomial⁶ $p \in \mathbb{R}[z]$ is called *symmetric* if it is invariant to any permutation of the variables z , i.e.,

$$p(z) := p(z_1, \dots, z_m) = p(z_{\pi(1)}, \dots, z_{\pi(m)}) =: p(\Pi z), \quad (8)$$

where π is a permutation on $\{1, \dots, m\} =: [m]$ (i.e., a bijective function from $[m]$ to itself) and Π is an $m \times m$ matrix representing this permutation, with i th row the canonical vector $e_{\pi(i)}^\top$ of all zeros, except a 1 at position $\pi(i)$.

Now let A, y be as in Problem 1 and let (Π^*, ξ^*) , $\xi^* = [\xi_1^*, \dots, \xi_n^*]^\top$ be a solution. Let $p \in \mathbb{R}[z]$ be a symmetric polynomial. Then the key observation is that the equality $\Pi^* y = A \xi^*$ implies the equality $p(\Pi^* y) = p(A \xi^*)$, and since p is symmetric, this in turn implies

$$\Pi^* y = A \xi^* \xrightarrow{p: \text{symmetric}} p(y) = p(\Pi^* y) = p(A \xi^*). \quad (9)$$

That is, the symmetric polynomial p serves in eliminating the unknown permutation Π^* and providing a constraint equation that depends only on the known data A, y ,

$$\hat{p}(x) := p(Ax) - p(y) = 0, \quad (10)$$

and which the solution ξ^* that we seek needs to satisfy.

Example 2: Consider the data

$$A = \begin{bmatrix} -1 & -2 \\ 2 & -3 \\ 0 & 4 \end{bmatrix}, \quad y = \begin{bmatrix} 8 \\ -5 \\ -4 \end{bmatrix}. \quad (11)$$

It is simple to check that there is only one permutation

$$\Pi^* = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} \quad (12)$$

that results in a consistent linear system of equations

$$\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 8 \\ -5 \\ -4 \end{bmatrix} = \begin{bmatrix} -1 & -2 \\ 2 & -3 \\ 0 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \quad (13)$$

⁵See [35], [36] for a generalization of this result to arbitrary linear transformations instead of just permutations.

⁶We do not distinguish between p and $p(z)$.

with solution $\xi_1^* = 1, \xi_2^* = 2$; every other permutation results in inconsistent equations. Now consider the symmetric polynomial

$$p_1(z_1, z_2, z_3) = z_1 + z_2 + z_3, \quad (14)$$

which we may use as in (10) to generate the constraint

$$(-x_1 - 2x_2) + (2x_1 - 3x_2) + 4x_2 = 8 - 5 - 4, \Leftrightarrow \quad (15)$$

$$x_1 - x_2 = -1, \quad (16)$$

$$\Leftrightarrow \hat{p}_1(x) := p_1(Ax) - p_1(y) = 0. \quad (17)$$

Indeed, we see that the solution $\xi^* = [1, 2]^\top$ satisfies (16).

The polynomial \hat{p} in (10) is an element of the polynomial ring $\mathbb{R}[x]$ in n variables $x := [x_1, \dots, x_n]^\top$, and the set of its roots, denoted as $\mathcal{V}(\hat{p}) := \{\xi \in \mathbb{R}^n : \hat{p}(\xi) = 0\}$ and called an *algebraic variety*, in principle defines a hypersurface of \mathbb{R}^n . Since the solution ξ^* to (4) is an element of the n -dimensional space \mathbb{R}^n , and $\xi^* \in \mathcal{V}(\hat{p})$ for any such \hat{p} , one expects that using n *sufficiently independent* such polynomials, will yield a system of n equations in n unknowns,

$$\hat{p}_1(x) = \dots = \hat{p}_n(x) = 0, \quad (18)$$

that has a finite number of solutions. Geometrically, these solutions are the intersection points of the corresponding n hypersurfaces $\mathcal{V}(\hat{p}_1), \dots, \mathcal{V}(\hat{p}_n)$, which contain all solutions to Problem (1), as well as potentially other irrelevant points.

Example 3: Continuing with Example 2, suppose we further use the symmetric polynomial

$$p_2(z_1, z_2, z_3) = z_1^2 + z_2^2 + z_3^2 \quad (19)$$

in (10) to obtain the additional constraint

$$\begin{aligned} (-x_1 - 2x_2)^2 + (2x_1 - 3x_2)^2 + (4x_2)^2 &= 8^2 + (-5)^2 + (-4)^2, \\ \Leftrightarrow \hat{p}_2(x) := p_2(Ax) - p_2(y) &= 0. \end{aligned} \quad (20)$$

Solving (16) with respect to x_1 and substituting to (20), gives a quadratic equation in x_2 with solutions $\xi_2 = 2$ and $\xi_2 = -25/13$. Solving (16) for x_1 gives

$$\mathcal{V}(\hat{p}_1, \hat{p}_2) = \left\{ \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} -38/13 \\ -25/13 \end{bmatrix} \right\}. \quad (21)$$

We see that $\mathcal{V}(\hat{p}_1, \hat{p}_2)$ contains the solution of the linear system (13) but also an additional irrelevant point.

We note here that one may use $n + 1$ polynomials in order to remove the irrelevant points, e.g., as was done in [11]. However, such an approach is of theoretical interest only, since a system of $n + 1$ (*sufficiently independent*) equations in n unknowns is bound to be inconsistent even in the slightest presence of noise. Instead, here we study a system of n equations in n unknowns and later show (see §III) how one can filter its roots of interest.

C. Main Results

We present our main results in §II-C.1 (Theorem 2) and §II-C.2 (Theorems 3-4) for exact and corrupted data, respectively.

1) *Exact Data*: As Examples 2-3 suggest, a natural choice for our n symmetric polynomials are the first n power sums $p_k(z) \in \mathbb{R}[z] := \mathbb{R}[z_1, \dots, z_m]$, $k \in [n] := \{1, \dots, n\}$, defined as

$$p_k(z) := z_1^k + \dots + z_m^k. \quad (22)$$

The above discussion has already established that any solution ξ^* of (4) must satisfy the polynomial constraints

$$\hat{p}_k(x) = 0, \quad k \in [n], \quad \text{where} \quad (23)$$

$$\hat{p}_k(x) := p_k(Ax) - p_k(y) = \sum_{i=1}^m (a_i^\top x)^k - \sum_{j=1}^m y_j^k, \quad (24)$$

and a_i^\top denotes the i th row of A . The next major result guarantees that there can only be a finite number of other irrelevant solutions.

Theorem 2: If A is generic and y is some permutation of some vector in $\mathcal{R}(A)$, then the algebraic variety $\mathcal{V}(\hat{p}_1, \dots, \hat{p}_n)$ contains all $\xi_1^*, \dots, \xi_\ell^* \in \mathbb{R}^n$ such that there exist permutations $\Pi_1^*, \dots, \Pi_\ell^*$ with $\Pi_i^* y = A \xi_i^*$, $\forall i \in [\ell]$, while it may contain at most $n! - \ell$ other points of \mathbb{C}^n . If in addition y is some permutation of a generic vector in $\mathcal{R}(A)$, then $\ell = 1$.

Theorem 2 guarantees that the system of polynomial equations

$$\hat{p}_1(x) = \dots = \hat{p}_n(x) = 0, \quad (25)$$

always has a finite number of solutions in \mathbb{C}^n (at most $n!$), among which lie all possible solutions $\xi_1^*, \dots, \xi_\ell^* \in \mathbb{R}^n$ of Problem 1. The importance of the solutions being finite in \mathbb{C}^n is computational: even if one is interested only in the real roots, knowing that the system has finitely many complex roots allows one to use much more efficient solvers. On the other hand, there exist many pathological cases where a system of polynomial equations has finitely many real roots but an infinity of complex roots, as the next example demonstrates; Theorem 2 guarantees that such a pathological case can not occur.

Example 4: The polynomial equation $x_1^2 + x_2^2 = 0$ has only one real root $[0, 0]^\top$, while over the complex numbers it defines a union of two lines in \mathbb{C}^2 .

2) *Corrupted Data*: We next consider corrupted versions \tilde{A}, \tilde{y} of A, y respectively, the case of interest in practical applications. Define the *corrupted* power-sum polynomials as

$$\tilde{p}_k(x) := p_k(\tilde{A}x) - p_k(\tilde{y}), \quad k \in [n], \quad (26)$$

and consider the polynomial system $\tilde{\mathcal{P}}$ of n equations of degrees $1, 2, \dots, n$ in n unknowns, given by

$$\tilde{\mathcal{P}}: \tilde{p}_1 = \dots = \tilde{p}_n = 0. \quad (27)$$

A natural question is: how do the roots of $\tilde{\mathcal{P}}$ behave as functions of the corruptions $\tilde{A} - A, \tilde{y} - y$? This is a challenging question to answer analytically. But in fact, there is an even more fundamental question lurking in our development: does $\tilde{\mathcal{P}}$ have any roots at all in \mathbb{C}^n ? As the next example shows, the answer is not necessarily affirmative.

Example 5: Let A, y be as in Example 2, and let

$$\tilde{A} = \begin{bmatrix} -2 & -1 \\ 2 & -3 \\ 0 & 4 \end{bmatrix} \quad (28)$$

be a corrupted version of A resulting from swapping the two elements in the first row of A . Moreover, consider no corruption on y , i.e., $\tilde{y} = y$. Then $\tilde{\mathcal{P}}$ has no solutions because $\tilde{p}_1 = -1$, i.e., the polynomial constraint corresponding to \tilde{p}_1 becomes infeasible.

A second fundamental question is: Suppose $\tilde{\mathcal{P}}$ has a solution, then does it have finitely many solutions? Once again, this need not be true, as the following example illustrates.

Example 6: Let \tilde{A} be as in Example 5 and

$$\tilde{y} = \begin{bmatrix} 9 \\ -5 \\ -4 \end{bmatrix} = y + \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}. \quad (29)$$

Then $\tilde{\mathcal{P}}$ has infinitely many solutions because \tilde{p}_1 is identically the zero polynomial so that the root locus of $\tilde{\mathcal{P}}$ is the curve $\mathcal{V}(\tilde{p}_2) = \{\xi \in \mathbb{C}^n : 8\xi_1^2 - 2\xi_1\xi_2 + 26\xi_2^2 - 105 = 0\}$.

Theorem 3 below essentially states that the pathological situations of Examples 5-6 can only occur for \tilde{A} taking values on a subset of $\mathbb{R}^{m \times n}$ of measure zero, regardless of what \tilde{y} is.

Theorem 3: If \tilde{A} is generic and $\tilde{y} \in \mathbb{R}^m$ is any vector, then $\mathcal{V}(\tilde{p}_1, \dots, \tilde{p}_n)$ is non-empty containing at most $n!$ points of \mathbb{C}^n .

Due to the linearity of the linear regression model it is customary to restrict attention to the case where the corruptions affect only the observations. Then a consequence of Theorem 3 and [30] is:

Theorem 4: Suppose the corruptions affect only the observations, i.e. $\tilde{A} = A$ and $\tilde{y} = y + \varepsilon$, for $\varepsilon \in \mathbb{R}^m$. If A is generic then $\mathcal{V}(\tilde{p}_1, \dots, \tilde{p}_n) = \{\tilde{\xi}_1, \dots, \tilde{\xi}_L\} \subset \mathbb{C}^n$, $1 \leq L \leq n!$. Moreover, there exist positive constants τ, γ and non-negative constant γ' such that for every $\xi^* \in \mathbb{R}^n$ for which there exists a permutation Π^* with $\Pi^* y = A \xi^*$, there exists $i^* \in [L]$ such that

$$\|\xi^* - (\tilde{\xi}_{i^*})_{\mathbb{R}}\|_2 \leq \tau (1 + \|\xi^*\|_2)^{\gamma'} \|e\|_2^{\gamma}, \quad (30)$$

where $(\tilde{\xi}_{i^*})_{\mathbb{R}}$ denotes the real part of $\tilde{\xi}_{i^*}$ and $e \in \mathbb{R}^n$ with $e_k = \sum_{j=1}^m \sum_{\ell=1}^k \binom{k}{\ell} y_j^{k-\ell} \varepsilon_j^\ell$ for every $k \in [n]$.

Theorems 3-4 are important for several reasons. First, they guarantee that the system of polynomial equations (27) is almost always consistent, i.e., there exists at least one solution. In the absence of noise this property is immediate simply because ξ^* is a root to any of the noiseless polynomials \hat{p}_k . However, for noisy data the consistency of (27) is far from obvious; for example, the authors of [28] write *It is generally impossible to solve these equations analytically; in fact there may not be a solution to the system*. Theorems 3-4 guarantee that such an issue is of no concern. Secondly, they guarantee finiteness of solutions in \mathbb{C}^n : this is important because it allows the deployment of efficient polynomial solvers specifically designed for *zero-dimensional* polynomial systems; moreover, it gives us the option of seeking an estimate for the parameters ξ^* of the regression problem among a finite set of points. Indeed, the bound of Theorem 4 implies that the system is

Hölder continuous [30], so that for a *small* level of observation noise ε , some root of the noisy system (27) is expected to be *close* to ξ^* (the entries of e in (30) are polynomials in ε which evaluate to zero for $\varepsilon = 0$). Admittedly, characterizing the SNR rate of such an estimator requires tight bounds on the constants τ, γ, γ' . We leave this problem to future research with the notes that 1) as per [30] *the exponent γ is typically a small (and unknown) positive number in contrast to the exponent $\gamma = 1$* , and 2) the encouraging simulations of §III suggest that τ, γ, γ' are well-behaved.

D. Proofs

1) *Proof of Theorem 1:* Let A be a generic $m \times n$ matrix with $m > n$ and ξ an $n \times 1$ generic vector. Since A is generic and $m > n$, the rank of A is equal to n . Let Π^* be a permutation as in Problem 1. We want to show that the only permutation Π for which $\Pi(\Pi^*)^\top A\xi$ is in the range-space of A is $\Pi = \Pi^*$. This is equivalent to proving that for any permutation Π different than the identity we have

$$\text{rank}[A \ \Pi A\xi] = n + 1. \quad (31)$$

Since A is generic, it can be written as the product $A = LU$ of an $m \times m$ generic lower triangular matrix L and an $m \times n$ matrix U , whose top $n \times n$ block is a generic upper triangular matrix and its $(m-n) \times n$ bottom part is the zero matrix. Then

$$L^{-1}[A \ \Pi A\xi] = [U \ L^{-1}\Pi LU\xi]. \quad (32)$$

Because of the structure of U it is enough to show that one of the last $m-n$ entries of the vector $L^{-1}\Pi LU\xi$ is non-zero. But because ξ is generic, it is enough to show that one of the last $m-n$ rows of the matrix $L^{-1}\Pi LU$ is non-zero. Towards that end, we will show that the $(m, 1)$ entry of this matrix is non-zero. This entry is zero if and only if

$$L_{m,:}^{-1}\Pi L_{:,1} = 0, \quad (33)$$

where $L_{m,:}^{-1}$ denotes the last row of L^{-1} and $L_{:,1}$ denotes the first column of L . This last equation says that $L_{m,:}^{-1}$ must be orthogonal to $\Pi L_{:,1}$. But by definition, $L_{m,:}^{-1}$ is orthogonal to all the columns $L_{:,1}, \dots, L_{:,m-1}$ of L except the last one. Put together, we have that $L_{m,:}^{-1}$ must be orthogonal to $L_{:,1}, \dots, L_{:,m-1}, \Pi L_{:,1}$, which is possible only if

$$\text{rank}[L_{:,1} \ \dots \ L_{:,m-1} \ \Pi L_{:,1}] = m - 1. \quad (34)$$

Hence, it is enough to show that

$$\det[L_{:,1} \ \dots \ L_{:,m-1} \ \Pi L_{:,1}] \neq 0. \quad (35)$$

Towards that end, we view the entries ℓ_{ij} of L as variables of a polynomial ring $\mathbb{C}[\ell_{11}, \dots, \ell_{mm}]$ in $m(m+1)/2$ variables. Then the determinant in (35) is a polynomial of $\mathbb{C}[\ell_{11}, \dots, \ell_{mm}]$. It is enough to show that this polynomial is non-zero. For in that case, it defines a hypersurface of $\mathbb{C}^{m(m+1)/2}$, which will not contain L , since L is generic. To show that the determinant is indeed a non-zero polynomial we consider an ordering of the variables

$$\ell_{11} > \ell_{21} > \ell_{22} > \ell_{31} > \ell_{32} > \dots > \ell_{m-1,m} > \ell_{mm}, \quad (36)$$

and then consider the induced lexicographic order on all the monomials of $\mathbb{C}[\ell_{11}, \dots, \ell_{mm}]$.

First, suppose that ℓ_{11} occurs in the k th entry of $\Pi L_{:,1}$ where $k > 1$. We will show that the largest monomial appearing in the definition of $\det[L_{:,1} \ \dots \ L_{:,m-1} \ \Pi L_{:,1}]$ only occurs in one way; this guarantees that the determinant is non-zero. Indeed, the largest monomial occurs in the following unique way: choose the largest element from each row, starting from the rows that contain the largest elements. That is, pick element ℓ_{11} from row 1, element ℓ_{11} from row k , element ℓ_{ii} from row i for $i = 2, \dots, k-1$, and elements $\ell_{i+1,i}$ from row $i+1$ for $i = k+1, \dots, m-1$.

Next, suppose that $k \geq 1$ is the largest index such that the i th entry of $\Pi L_{:,1}$ is equal to ℓ_{i1} for $i = 1, \dots, k$. In that case, we apply an elementary column operation by subtracting the first column of $[L_{:,1} \ \dots \ L_{:,m-1} \ \Pi L_{:,1}]$ from its last, to obtain a new matrix M of the same rank. Then the first k entries of the last column of M are zero, while its (i, m) entry for $i > k$ is of the form $\ell_{s_i,1} - \ell_{i,1}$, with $s_i > k, \forall i = k+1, \dots, m$, and with $s_{k+1} \neq k+1$. Let $t > 1$ be such that $s_{k+t} = k+1$. Then the largest monomial in $\det(M)$ occurs in a unique way as the first term of the expanded product of elements ℓ_{ii} from row i for $i = 1, \dots, k+t-1$, element $(\ell_{k+1,1} - \ell_{k+t,1})$ from row $k+t$, and elements $\ell_{i+1,i}$ from row $i+1$ for $i = k+t, \dots, m-1$. That is, $\det(M)$ is a non-zero polynomial.

2) *Power-Sums and Regular Sequences:* The goal of this section is to prove Lemma 4, which is the main tool needed for the proof of Theorems 2 and 3; the reader is encouraged to consult Appendix D before proceeding. To begin with, recall the power-sums polynomials

$$p_k = z_1^k + \dots + z_m^k, \quad k: \text{positive integer}, \quad (37)$$

that were used in §II-C as the base symmetric polynomials towards eliminating the unknown permutation Π^* . These are polynomials in m variables z_1, \dots, z_m , i.e., $p_k \in \mathbb{C}[z] := \mathbb{C}[z_1, \dots, z_m]$. The next two lemmas establish that p_1, \dots, p_m form a regular sequence.⁷

Lemma 1: Let $\sigma_1, \dots, \sigma_m$ be the elementary symmetric polynomials in m variables z , defined as

$$\sigma_k := \sum_{1 \leq i_1 < i_2 < \dots < i_k \leq m} z_{i_1} z_{i_2} \dots z_{i_k}. \quad (38)$$

Then $\sigma_1, \dots, \sigma_m$ form a regular sequence of $\mathbb{C}[z]$.

Proof: We will show that $\mathcal{V}_{\mathbb{C}^m}(\sigma_1, \dots, \sigma_m) = \{0\}$, in which case the statement will follow from Proposition D1 together with the fact that each σ_k is homogeneous (of degree k). We proceed by induction on m . If $m = 1$, $\mathbb{C}[z] = \mathbb{C}[z_1]$ and the only elementary symmetric polynomial is z_1 . Clearly, z_1 vanishes only on $0 \in \mathbb{C}$. Suppose now that $m > 1$. Let $\zeta = [\zeta_1, \dots, \zeta_m]^\top \in \mathcal{V}(\sigma_1, \dots, \sigma_m)$. Then $\sigma_m(\zeta) = \zeta_1 \dots \zeta_m = 0$, and without loss of generality we can assume that $\zeta_m = 0$. Then $[\zeta_1, \dots, \zeta_{m-1}]^\top$ is in the variety generated by the $m-1$ elementary symmetric polynomials in $m-1$ variables and the induction on m gives $\zeta_1 = \dots = \zeta_{m-1} = 0$. \square

Lemma 2: The first m power-sum polynomials p_1, \dots, p_m defined in (37) form a regular sequence of $\mathbb{C}[z]$.

⁷This also follows from a more general theorem in [47], which states that any m consecutive such polynomials $p_\ell, p_{\ell+1}, \dots, p_{\ell+m-1}$, where ℓ is any positive integer, form a regular sequence of $\mathbb{C}[z]$. To make the paper more accessible and self-contained, we have taken the liberty of giving the rather simple case $\ell = 1$ its own proof.

Proof: Newton's identities

$$k\sigma_k = \sum_{i=1}^k (-1)^{i-1} \sigma_{k-i} p_i, \quad \forall k \in [m], \quad (39)$$

show that $\sigma_1, \dots, \sigma_m$ can be obtained inductively in terms of p_1, \dots, p_m . On the other hand, the fundamental theorem of symmetric polynomials states that every symmetric polynomial can be written as a polynomial in $\sigma_1, \dots, \sigma_m$. This implies that we have an equality of ideals $(p_1, \dots, p_m) = (\sigma_1, \dots, \sigma_m)$, hence an equality of algebraic varieties $\mathcal{V}_{\mathbb{C}^m}(p_1, \dots, p_m) = \mathcal{V}_{\mathbb{C}^m}(\sigma_1, \dots, \sigma_m)$. Thus Lemma 1 gives that $\mathcal{V}_{\mathbb{C}^m}(p_1, \dots, p_m)$ consists of a single point and Proposition D1 together with the fact that each p_k is homogeneous (of degree k) establishes that p_1, \dots, p_m is a regular sequence. \square

Now, recall that in §II-C each *base* polynomial p_k was used to furnish a polynomial equation $\hat{p}_k(x) = p_k(Ax) - p_k(y) = 0$ that the unique solution ξ^* to Problem 1 should satisfy. Notice here how we are passing from polynomials $p_k \in \mathbb{C}[z]$ in m variables to polynomials $\hat{p}_k \in \mathbb{C}[x]$ in n variables. Notice further that even though the p_k are symmetric and homogeneous, the \hat{p}_k are in principle neither symmetric nor homogeneous. In fact,

$$\hat{p}_k = \bar{p}_k - p_k(y), \quad \text{where} \quad (40)$$

$$\bar{p}_k := p_k(Ax) = \sum_{i=1}^m (a_i^\top x)^k, \quad k = 1, \dots, n. \quad (41)$$

Here \bar{p}_k is a homogeneous polynomial of degree k and $p_k(y) \in \mathbb{R}$ is a constant. As it turns out, the homogeneous parts \bar{p}_k of the \hat{p}_k are the bridge for passing properties of the base polynomials p_k to the polynomials of interest \hat{p}_k . The next two lemmas are the two required steps towards building that bridge.

Lemma 3: Let $\ell_1^\top, \dots, \ell_{m-n}^\top \in \mathbb{R}^{1 \times m}$ be a basis for the left nullspace of A . Then the algebraic varieties $\mathcal{V}_{\mathbb{C}^n}(\bar{p}_1, \dots, \bar{p}_n)$ and $\mathcal{V}_{\mathbb{C}^m}(\ell_1^\top z, \dots, \ell_{m-n}^\top z, p_1, \dots, p_n)$ are isomorphic. In particular, the two varieties have the same dimension.

Proof: Let $\mathcal{X} := \mathcal{V}_{\mathbb{C}^m}(\ell_1^\top z, \dots, \ell_{m-n}^\top z, p_1, \dots, p_n)$ and $\mathcal{Y} := \mathcal{V}_{\mathbb{C}^n}(\bar{p}_1, \dots, \bar{p}_n)$. To show that \mathcal{X} and \mathcal{Y} are isomorphic as algebraic varieties, it is enough to find a bijective map $f: \mathcal{Y} \rightarrow \mathcal{X}$, such that both f and f^{-1} are given by polynomials. First, we define f by specifying its image on an arbitrary point ξ of \mathcal{Y} :

$$\xi \in \mathcal{Y} \subset \mathbb{C}^n \xrightarrow{f} A\xi \in \mathbb{C}^m. \quad (42)$$

We need to show that $f(\xi) \in \mathcal{X}$. To do that, we need to check that $f(\xi)$ satisfies the defining equations of \mathcal{X} . Towards that end, note that $\ell_j^\top f(\xi) = \ell_j^\top A\xi = 0, \forall j \in [m-n]$, by definition of the ℓ_j . Moreover, $p_k(f(\xi)) = p_k(A\xi) = \bar{p}_k(\xi) = 0, \forall k \in [n]$, where the last equality is true because $\xi \in \mathcal{Y}$. Hence f is indeed a map of the form $f: \mathcal{Y} \rightarrow \mathcal{X}$. Moreover, f is given by linear polynomials, i.e., the coordinates of $f(\xi)$ are linear polynomials of the coordinates of ξ .

Next, we define a map $g: \mathcal{X} \rightarrow \mathcal{Y}$ as follows. Let $\zeta \in \mathcal{X}$. Then $\ell_j^\top \zeta = 0, \forall j \in [m-n]$. Since the ℓ_j form a basis for $\mathcal{N}(A^\top)$, this means that $\zeta \in \mathcal{N}(A^\top)^\perp$. But from basic linear

algebra $\mathcal{N}(A^\top)^\perp = \mathcal{R}(A)$. Hence there exists some $\xi \in \mathbb{C}^n$ such that $\zeta = A\xi$. Moreover, this ξ is unique because A has full rank by the assumption of Problem 1. This allows us to define the map g as $g(\zeta) = \xi := (A^\top A)^{-1} A^\top \zeta$. Then

$$\bar{p}_k(g(\zeta)) = p_k(Ag(\zeta)) = p_k(A(A^\top A)^{-1} A^\top \zeta) \quad (43)$$

$$= p_k(A(A^\top A)^{-1} A^\top A\xi) \quad (44)$$

$$= p_k(A\xi) \quad (45)$$

$$= p_k(\zeta) = 0, \quad (46)$$

where the last equality is true because $\zeta \in \mathcal{X}$. Hence $g(\zeta) \in \mathcal{Y}$. Moreover, g is given by polynomials, since each coordinate of $g(\zeta)$ is a linear polynomial in the coordinates of ξ . Finally, it is simple to check that f and g are inverses of each other. \square

Lemma 4: If A is generic, then the polynomials $\bar{p}_1, \dots, \bar{p}_n$ defined in (41) form a regular sequence of $\mathbb{C}[x]$.

Proof: The space of all linear subspaces of dimension n of \mathbb{R}^m is an algebraic variety of \mathbb{R}^M , $M = \binom{m}{n}$, called *Grassmannian* and denoted $\mathbb{G}(n, m)$. The space $\mathbb{G}(m-n, m)$ of all linear subspaces of dimension $m-n$ of \mathbb{R}^m is also an algebraic variety of \mathbb{R}^M . These two varieties are isomorphic under a mapping that takes a linear subspace \mathcal{S} of dimension n to its orthogonal complement \mathcal{S}^\perp . Hence, \mathcal{S} is generic if and only if \mathcal{S}^\perp is generic. Choosing A generic is the same as choosing n generic vectors in \mathbb{R}^m , which is the same as choosing a generic subspace $\mathcal{S} = \mathcal{R}(A)$ of \mathbb{R}^m . This is the same as choosing $\mathcal{S}^\perp = \mathcal{N}(A^\top)$ generic, which is the same as choosing $m-n$ generic vectors of \mathbb{R}^m . In other words, if A is generic, then a generic set of vectors $\ell_1, \dots, \ell_{m-n}$ inside $\mathcal{N}(A^\top)$, is a set of $m-n$ generic vectors of a generic $(m-n)$ -dimensional linear subspace of \mathbb{R}^m . Hence, with respect to any property that does not depend on A , the vectors $\ell_1, \dots, \ell_{m-n}$ behave as generic vectors of \mathbb{R}^m .

So let $L = [\ell_1 \cdots \ell_{m-n}]$ be a matrix containing in its columns a set of generic vectors of $\mathcal{N}(A^\top)$. By Lemma 2 p_1, \dots, p_m is a regular sequence of $\mathbb{C}[z]$. By the definition of regular sequence, the subsequence p_1, \dots, p_n is also regular. By inductive application of Proposition D2,

$$p_1, \dots, p_n, \ell_1^\top z, \dots, \ell_{m-n}^\top z \quad (47)$$

is also a regular sequence of $\mathbb{C}[z]$. By Proposition D1, we have that

$$\dim \mathcal{V}_{\mathbb{C}^m}(\ell_1^\top z, \dots, \ell_{m-n}^\top z, p_1, \dots, p_n) = 0. \quad (48)$$

By Lemma 3 we have that $\dim \mathcal{V}_{\mathbb{C}^n}(\bar{p}_1, \dots, \bar{p}_n) = 0$. Then by Proposition D1 $\bar{p}_1, \dots, \bar{p}_n$ is a regular sequence of $\mathbb{C}[x]$. \square

3) *Proof of Theorem 2:* Having developed the machinery that led to Lemma 4, the task of proving Theorem 2 is a matter of putting this machinery to work; at this stage the reader is encouraged to consult appendices A-E before proceeding.

Let $\mathcal{I} = (\hat{p}_1, \dots, \hat{p}_n)$ be the ideal generated by our polynomials $\hat{p}_1, \dots, \hat{p}_n$. Consider the weight order of $\mathbb{C}[x]$ defined by the vector $w = [1, \dots, 1]^\top \in \mathbb{Z}^n$ (see Appendix E). Since $\bar{p}_1, \dots, \bar{p}_n$ is a regular sequence of homogeneous polynomials (Lemma 4) and $\hat{p}_k = \bar{p}_k + \text{constant}$, we can obtain $\text{in}_w(\mathcal{I})$ just from the leading homogeneous terms of the generators of \mathcal{I} , i.e., $\text{in}_w(\mathcal{I}) = (\bar{p}_1, \dots, \bar{p}_n)$ (Proposition E2). Since $\bar{p}_1, \dots, \bar{p}_n$ is a regular sequence of length n in the polynomial

ring $\mathbb{C}[x]$ of n variables, we have $\dim \mathcal{V}_{\mathbb{C}^n}(\text{in}_w(\mathcal{I})) = 0$ (Proposition D1). Since $\mathcal{V}_{\mathbb{C}^n}(\mathcal{I})$ and $\mathcal{V}_{\mathbb{C}^n}(\text{in}_w(\mathcal{I}))$ have the same dimension (Proposition E1), we have $\dim \mathcal{V}_{\mathbb{C}^n}(\mathcal{I}) = 0$. Since zero-dimensional varieties have a finite number of points (Proposition A2), we have that $\mathcal{V}_{\mathbb{C}^n}(\hat{p}_1, \dots, \hat{p}_n)$ consists of finitely many points.

We now bound from above the number of points of $\mathcal{Y} := \mathcal{V}_{\mathbb{C}^n}(\hat{p}_1, \dots, \hat{p}_n)$. Consider the affine cone $\mathcal{Y}^{(h)} = \mathcal{V}_{\mathbb{C}^{n+1}}(\hat{p}_1^{(h)}, \dots, \hat{p}_n^{(h)})$ of \mathcal{Y} (see Appendix C). Since $\dim \mathcal{Y} = 0$, we have $\dim \mathcal{Y}^{(h)} = 1$ (Proposition C1). Let $\mathcal{Y}^{(h)} = \mathcal{W}_1 \cup \dots \cup \mathcal{W}_\ell$ be the irreducible decomposition of $\mathcal{Y}^{(h)}$ (Proposition A1). Since $\mathcal{Y}^{(h)}$ is generated by homogeneous polynomials, it is the union of lines through the origin. Hence, each irreducible component \mathcal{W}_i is the union of lines through the origin. We argue that each \mathcal{W}_i is a single line through the origin. For if some \mathcal{W}_j is not a single line, let $\xi \in \mathcal{W}_j$ be a point different than the origin 0. Letting \mathcal{L} be the line through the origin and ξ , we have a chain $\mathcal{W}_j \supseteq \mathcal{L} \supseteq \{0\}$ of irreducible subsets of $\mathcal{Y}^{(h)}$ of length 2. But this contradicts the fact that $\dim \mathcal{Y}^{(h)} = 1$ (Definition A1). This shows that $\mathcal{Y}^{(h)}$ is the union of ℓ lines through the origin. Then $\ell \leq \deg(\hat{p}_1^{(h)}) \deg(\hat{p}_2^{(h)}) \dots \deg(\hat{p}_n^{(h)}) = n!$ (Proposition C2). Finally, the points of \mathcal{Y} are in 1-1 correspondence with the intersection points of $\mathcal{Y}^{(h)}$ with the hyperplane $t = 1$ of \mathbb{C}^{n+1} (see Appendix C). But the number of these intersection points can not exceed the number of lines of $\mathcal{Y}^{(h)}$.

4) *Proof of Theorem 3:* The proof is almost the same as the proof of Theorem 2 with an additional twist: we need to show that the variety $\tilde{\mathcal{Y}} := \mathcal{V}_{\mathbb{C}^n}(\tilde{p}_1, \dots, \tilde{p}_n)$, where the \tilde{p}_k are defined in (26), is non-empty. So suppose that $\tilde{\mathcal{Y}} = \emptyset$. Then $1 \in \tilde{I} := (\tilde{p}_1, \dots, \tilde{p}_n)$ (Proposition B2). Hence $1 = \text{in}_w(1) \in \text{in}_w(\tilde{I})$. Since \tilde{A} is generic, the polynomials $\tilde{p}_k := p_k(\tilde{A}x)$, $k \in [n]$, form a regular sequence (Lemma 4). Since $\tilde{p}_k := \tilde{p}_k - p_k(\tilde{y})$, we must have that $\text{in}_w(\tilde{I}) = (\tilde{p}_1, \dots, \tilde{p}_n)$ (Proposition E2). But now $1 \in (\tilde{p}_1, \dots, \tilde{p}_n)$ is impossible, because all \tilde{p}_k are homogeneous of positive degree, i.e., the equation $1 = q_1\tilde{p}_1 + \dots + q_n\tilde{p}_n$ can not be true for any $q_k \in \mathbb{C}[x]$. This contradiction means that $\tilde{\mathcal{Y}}$ is non-empty. The rest of the proof is the same as that of Theorem 2.

5) *Proof of Theorem 4:* In view of Theorem 3 we only need to prove (30). For $\xi \in \mathbb{C}^n$ write $\xi = (\xi)_{\mathbb{R}} + \sqrt{-1}(\xi)_{\mathbb{I}}$, where $(\xi)_{\mathbb{R}}, (\xi)_{\mathbb{I}}$ are the real and imaginary parts of ξ respectively. Let $\mathbb{R}[u, v] = \mathbb{R}[u_1, \dots, u_n, v_1, \dots, v_n]$ be a polynomial ring in $2n$ variables. Then for each \tilde{p}_k (defined in (41)) there exist unique polynomials $\bar{g}_k, \bar{h}_k \in \mathbb{R}[u, v]$, such that

$$\begin{aligned} \tilde{p}_k(\xi) &= \bar{g}_k((\xi)_{\mathbb{R}}, (\xi)_{\mathbb{I}}) + \sqrt{-1}\bar{h}_k((\xi)_{\mathbb{R}}, (\xi)_{\mathbb{I}}), \\ \forall \xi &= (\xi)_{\mathbb{R}} + \sqrt{-1}(\xi)_{\mathbb{I}} \in \mathbb{C}^n. \end{aligned} \quad (49)$$

Define $\hat{g}_k = \bar{g}_k - \sum_{j=1}^m y_j^k$ and $\tilde{g}_k = \bar{g}_k - \sum_{j=1}^m (y_j + \varepsilon_j)^k$. Then

$$\hat{p}_k(\xi) = \hat{g}_k((\xi)_{\mathbb{R}}, (\xi)_{\mathbb{I}}) + \sqrt{-1}\bar{h}_k((\xi)_{\mathbb{R}}, (\xi)_{\mathbb{I}}), \quad (50)$$

$$\begin{aligned} \tilde{p}_k(\xi) &= \tilde{g}_k((\xi)_{\mathbb{R}}, (\xi)_{\mathbb{I}}) + \sqrt{-1}\bar{h}_k((\xi)_{\mathbb{R}}, (\xi)_{\mathbb{I}}), \\ \forall \xi &= (\xi)_{\mathbb{R}} + \sqrt{-1}(\xi)_{\mathbb{I}} \in \mathbb{C}^n. \end{aligned} \quad (51)$$

We thus have the following set-theoretic equalities, where we identify $\xi = (\xi)_{\mathbb{R}} + \sqrt{-1}(\xi)_{\mathbb{I}} \in \mathbb{C}^n$ with $((\xi)_{\mathbb{R}}, (\xi)_{\mathbb{I}}) \in \mathbb{R}^{2n}$:

$$\mathcal{V}_{\mathbb{C}^n}(\hat{p}_1, \dots, \hat{p}_n) = \mathcal{V}_{\mathbb{R}^{2n}}(\hat{g}_1, \dots, \hat{g}_n, \bar{h}_1, \dots, \bar{h}_n), \quad (52)$$

$$\mathcal{V}_{\mathbb{C}^n}(\tilde{p}_1, \dots, \tilde{p}_n) = \mathcal{V}_{\mathbb{R}^{2n}}(\tilde{g}_1, \dots, \tilde{g}_n, \bar{h}_1, \dots, \bar{h}_n). \quad (53)$$

Let $\xi^* \in \mathbb{R}^n$ be such that there exists a permutation Π^* for which $\Pi^*y = A\xi^*$. Then $\hat{p}_k(\xi^*) = 0$ for every $k \in [n]$. Equivalently, for $(\xi^*, 0) \in \mathbb{R}^{2n}$ we have $\hat{g}_k(\xi^*, 0) = \bar{h}_k(\xi^*, 0) = 0$. Moreover,

$$\tilde{g}_k(\xi^*, 0) = \bar{g}_k(\xi^*, 0) - \sum_{j=1}^m (y_j + \varepsilon_j)^k \quad (54)$$

$$= \hat{g}_k(\xi^*, 0) - \sum_{j=1}^m \sum_{\ell=1}^k \binom{k}{\ell} y_j^{k-\ell} \varepsilon_j^\ell \quad (55)$$

$$= - \sum_{j=1}^m \sum_{\ell=1}^k \binom{k}{\ell} y_j^{k-\ell} \varepsilon_j^\ell. \quad (56)$$

Now, Theorem 2.2 in [30] guarantees the existence of constants τ, γ, γ' as claimed for which there exists $\xi \in \mathcal{V}_{\mathbb{C}^n}(\tilde{p}_1, \dots, \tilde{p}_n)$ such that

$$\|\alpha - (\xi)_{\mathbb{R}}\|_2 \leq \|(\alpha, \beta) - ((\xi)_{\mathbb{R}}, (\xi)_{\mathbb{I}})\|_2 \leq \quad (57)$$

$$\tau(1 + \|(\alpha, \beta)\|_2)^\gamma \|e(\alpha, \beta)\|_2^{\gamma'}, \quad \forall \alpha, \beta \in \mathbb{R}^n, \quad (58)$$

where

$$e(\alpha, \beta) = (\tilde{g}_1(\alpha, \beta), \dots, \tilde{g}_n(\alpha, \beta), \bar{h}_1(\alpha, \beta), \dots, \bar{h}_n(\alpha, \beta)). \quad (59)$$

Setting $(\alpha, \beta) = (\xi^*, 0)$ concludes the proof.

III. ALGORITHM: ALGEBRAICALLY INITIALIZED EXPECTATION MAXIMIZATION

A. Algorithm Design

Assuming for simplicity (and without much loss of generality as per Theorem 1) that there is a unique solution ξ^* to Problem 1, Theorem 2 guarantees that ξ^* is one of the finitely many complex roots of the noiseless polynomial system \mathcal{P} of n equations in n unknowns given in (23). Even if the data \tilde{y}, \tilde{A} are corrupted, Theorem 3 further guarantees that the noisy system $\tilde{\mathcal{P}}$ given in (27) remains consistent with $L \leq n!$ complex roots. If in addition the corruption level is mild, as per Theorem 4 we expect one of the roots of $\tilde{\mathcal{P}}$ to be a good approximation to ξ^* . Our goal is to isolate that root and refine it.

The polynomial system $\tilde{\mathcal{P}}$ ((27)) is solvable in closed form for $n = 1, 2$, while for $n \geq 3$ it can be solved using any state-of-the-art solver [48]. Nevertheless, the complexity of solving such a system is known to be exponential in n , even for well-behaved cases [49]. Practically speaking, this currently limits us to the regime $n \leq 6$, since for $n = 6$ a standard homotopy-based solver such as Bertini [50] takes about 37 minutes on an Intel(R) i7-8650U, 1.9GHz, 16GB machine. On the other hand, for $n = 3, 4$ using an automatic solver generator [51] along the lines of [52] for the specific structure of our system, we are able to obtain very efficient linear algebra based solvers that run in milliseconds.

Having obtained the roots $\tilde{\xi}_1, \dots, \tilde{\xi}_L \in \mathbb{C}^n$, $L \leq n!$, of $\tilde{\mathcal{P}}$, we retain only their real parts $(\tilde{\xi}_1)_{\mathbb{R}}, \dots, (\tilde{\xi}_L)_{\mathbb{R}}$, and identify the root that can serve as a first approximation to ξ^* . We do this by selecting the root that yields the smallest ℓ_2 error among all possible permutations Π :

$$\xi_{\text{AI}} := \operatorname{argmin}_{i \in [L]} \left\{ \min_{\Pi} \left\| \Pi \tilde{y} - \tilde{A}(\tilde{\xi}_i)_{\mathbb{R}} \right\|_2 \right\}. \quad (60)$$

We note that each inner minimization $\min_{\Pi} \left\| \Pi \tilde{y} - \tilde{A}(\tilde{\xi}_i)_{\mathbb{R}} \right\|_2$ in (60) can be solved by sorting (see [29]) and so the computation of ξ_{AI} is of complexity $\mathcal{O}(Lm \log(m))$. Finally, we use the *algebraic initialization* ξ_{AI} as an initialization to the Expectation Maximization algorithm of [29], which as noted in §I-B consists of solving (5) via alternating minimization. The complete algorithmic listing, which we refer to as *Algebraically-Initialized Expectation-Maximization (AI-EM)*⁸ is given in Algorithm 1.

Algorithm 1 Algebraically-Initialized Expectation-Maximization

```

1: procedure AI-EM( $\tilde{y} \in \mathbb{R}^m$ ,  $\tilde{A} \in \mathbb{R}^{m \times n}$ ,  $T \in \mathbb{N}$ ,  $\epsilon \in \mathbb{R}_+$ )
2:    $p_k(z) := \sum_{j=1}^m z_j^k$ ,  $\tilde{p}_k := p_k(\tilde{A}x) - p_k(\tilde{y})$ ,  $k \in [n]$ ;
3:   Compute roots  $\{\xi_i\}_{i=1}^L \subset \mathbb{C}^n$  of  $\{\tilde{p}_k = 0, k \in [n]\}$ ;
4:   Extract the real parts  $\{(\tilde{\xi}_i)_{\mathbb{R}}\}_{i=1}^L$  of  $\{\tilde{\xi}_i\}_{i=1}^L \subset \mathbb{C}^n$ ;
5:    $\{\xi_0, \Pi_0\} \leftarrow \operatorname{argmin}_{\xi \in \{(\tilde{\xi}_i)_{\mathbb{R}}\}_{i=1}^L, \Pi} \left\| \Pi \tilde{y} - \tilde{A}\xi \right\|_2$ ;
6:    $t \leftarrow 0$ ,  $\Delta \mathcal{J} \leftarrow \infty$ ,  $\mathcal{J} \leftarrow \left\| \Pi_0 \tilde{y} - \tilde{A}\xi_0 \right\|_2$ ;
7:   while  $t < T$  and  $\Delta \mathcal{J} > \epsilon \mathcal{J}$  do
8:      $t \leftarrow t + 1$ ;
9:      $\xi_t \leftarrow \operatorname{argmin}_{\xi \in \mathbb{R}^n} \left\| \Pi_{t-1} \tilde{y} - \tilde{A}\xi \right\|_2$ ;
10:     $\Pi_t \leftarrow \operatorname{argmin}_{\Pi} \left\| \Pi \tilde{y} - \tilde{A}\xi_t \right\|_2$ ;
11:     $\Delta \mathcal{J} \leftarrow \mathcal{J} - \left\| \Pi_t \tilde{y} - \tilde{A}\xi_t \right\|_2$ ;
12:     $\mathcal{J} \leftarrow \left\| \Pi_t \tilde{y} - \tilde{A}\xi_t \right\|_2$ ;
13:   end while
14:   Return  $\xi_t, \Pi_t$ .
15: end procedure

```

B. Numerical Evaluation

Algorithms. We perform a numerical evaluation of our proposed AI-EM Algorithm 1, which relies on solving the polynomial system by homotopy continuation in Bertini [50] for $n \geq 5$, and on our C++ elimination-template-based custom solvers for $n = 3, 4$ (see [51], [52] for more details). We also compare with the robust ℓ_1 regression method of [18], as well as two variations of EM algorithms that were proposed in [29] towards computing the Maximum Likelihood Estimator of (5).⁹ The first, referred to as [29]¹, computes the MLE via alternating minimization exactly as in Algorithm 1, except that it uses as initialization the vector that best fits the data \tilde{y} , \tilde{A} in the least-squares sense. The second variation, referred

⁸A preliminary version of this algorithm has been published in the short conference paper [37].

⁹We do not compare with the algorithm of [33], since this algorithm assumes a strong hypothesis rendering it not applicable to the case where the observation vector y has not been subsampled; this is the case where unlabeled sensing becomes equivalent to linear regression without correspondences. For a comparison of [33] with other unlabeled sensing methods see [35].

to as [29]², uses the same initialization as [29]¹, but it replaces the brute search over all possible permutations by a dynamic empirical average of permutation matrices drawn from a suitable Markov chain. For all algorithms we use a maximal number of iterations $T = 100$ and for AI-EM and [29]¹ a convergence accuracy parameter of $\epsilon = 0.01$.

Data. We use the following generative model with additive noise. We randomly sample $A \in \mathbb{R}^{m \times n}$ and $\xi^* \in \mathbb{R}^n$ from the corresponding standard normal distributions and perform a random permutation on the entries of $A\xi^*$ to obtain a vector $y \in \mathbb{R}^m$. We further corrupt y by adding to it a vector $\varepsilon \in \mathbb{R}^m$ sampled from the zero-mean normal distribution with covariance matrix $\sigma^2 I_m$. The input data then consist of A and $\tilde{y} := y + \varepsilon$.

Metrics. We assess all methods by measuring the relative estimation error of ξ^* , e.g., if $\xi_{\text{AI-EM}}$ is the output of AI-EM, we report

$$100 \frac{\|\xi^* - \xi_{\text{AI-EM}}\|_2}{\|\xi^*\|_2} \%. \quad (61)$$

For AI-EM we further report the estimation error that corresponds to the best root ξ_{AI}^* of the polynomial system $\tilde{\mathcal{P}}$, defined as

$$\xi_{\text{AI}}^* := \operatorname{argmin}_{\hat{\xi}_i, i \in [L]} \|\xi^* - (\hat{\xi}_i)_{\mathbb{R}}\|_2, \quad (62)$$

as well as that of our estimated best root $\hat{\xi}_{\text{AI}}$ computed as in (60).

Results. Figures 1a-1c depict the estimation error of the compared methods for fully shuffled data, for $n = 3, 4, 5$, SNR = 0 : 10 : 100 dB and $m = 500$ fixed, averaged over 100 independent trials. Evidently, both [29]¹ and [29]² fail.¹⁰ This is not surprising, since, when the data are fully shuffled, the least-squares initialization used by these methods is rather far from the ground truth ξ^* . The robust ℓ_1 regression of [18] fails as well, and again this is not a surprise, since with fully shuffled data the corruptions are no longer sparse. On the other hand, ξ_{AI}^* remains relatively close to ξ^* as σ increases and our actual initialization ξ_{AI} coincides with ξ_{AI}^* for $\sigma \leq 0.04$ and is slightly worse than ξ_{AI}^* otherwise. Regardless, the alternating minimization further refines ξ_{AI} leading to even lower errors than ξ_{AI}^* .

Fig. 2 depicts the estimation error for different percentages 0% : 10% : 100% of partially shuffled data, for $n = 4$, $m = 500$, and SNR = 40dB fixed. In such a case, only a subset of the entries of the vector $A\xi^*$ is shuffled according to a random permutation. As seen, [18], [29]¹ and [29]² perform much better than for fully shuffled data. In fact, [18], [29]¹ are comparable to AI-EM for up to 40% – 50% shuffled data, upon which percentage they start deteriorating and eventually break down. On the other hand, [29]² breaks down as soon as 20% of the data have been shuffled. A more detailed behavior is also shown for 0% : 1% : 10% shuffled data in the same setting. We see that [18], [29]¹ and AI-EM are almost perfectly accurate, while [29]² presents already an error of

¹⁰We have observed that for fully shuffled data, large m and small n , [29]² tends to drive the regression vector to zero.

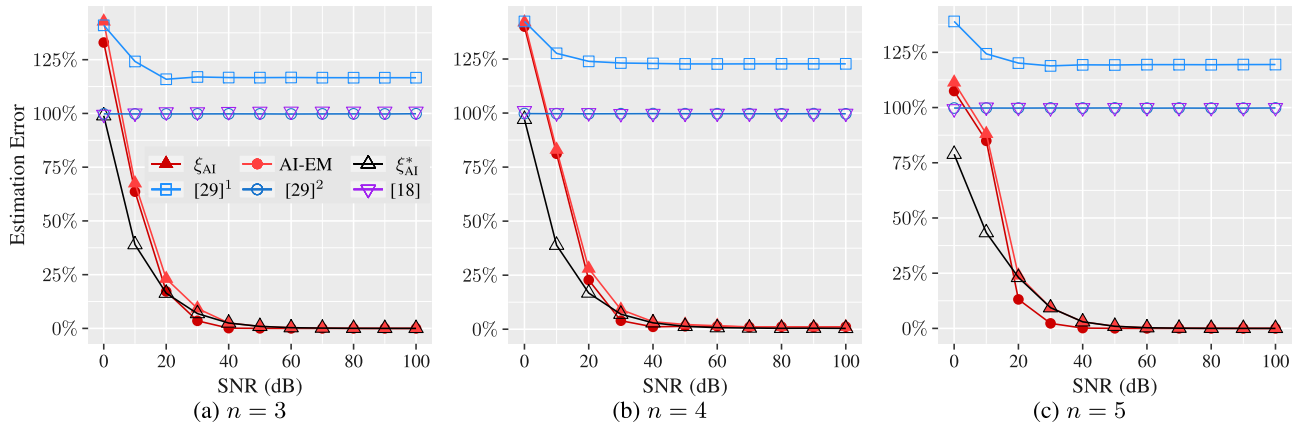


Fig. 1. Estimation error for fully shuffled data, for different values of n and SNR, with $m = 500$ fixed. Average over 100 trials.

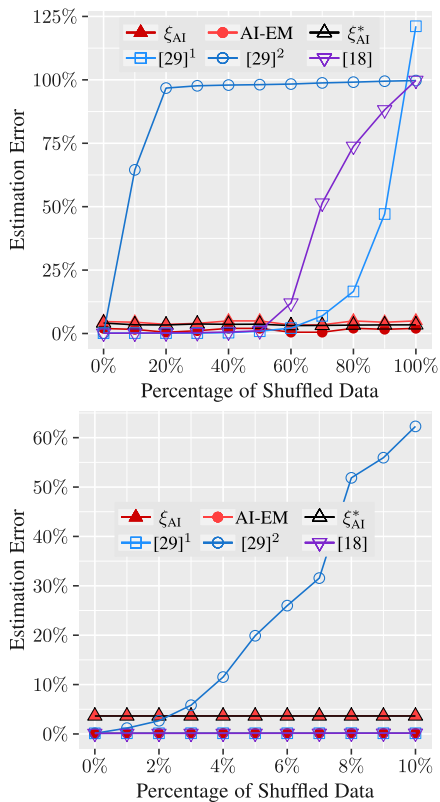


Fig. 2. Estimation error for partially shuffled data of varying percentages 0% : 10% : 100% (top) and 0% : 1% : 10% (bottom) for $m = 500$, $n = 4$, SNR = 40dB fixed (100 trials).

more than 10% for 4% shuffled data, and more than 60% for 10% shuffled data, i.e., $[29]^2$ seems to be accurate only when the percentage of shuffled data is very small.

Table I provides numerical data for the estimation error and running times for different values of $n = 3, 4, 5, 6$. We see that for $n = 3, 4$ solving the polynomial system $\tilde{\mathcal{P}}$ via our custom solvers requires only a few milliseconds. On other hand, the running time increases exponentially with 45 seconds required to solve the $n = 5$ system and about 37 minutes for $n = 6$. As expected, the running time of the Alternating Minimization (AM) remains practically unaffected

TABLE I
RUNNING TIME (IN MILLISECONDS) OF POLYNOMIAL SOLVERS AND ALTERNATING MINIMIZATION, AND ESTIMATION ERRORS FOR VARYING n . SNR = 40dB, $m = 500$, FULLY SHUFFLED DATA, 10 TRIALS

n	Running Time		Estimation Error		
	$\tilde{\mathcal{P}}$	AM	ξ_{AI}^*	ξ_{AI}	AI-EM
3	0.7	6	2.3%	2.3%	0.1%
4	9	6	3.9%	3.9%	0.1%
5	45, 157	7	1.4%	1.4%	0.1%
6	2, 243, 952	7	1.2%	1.2%	0.1%

TABLE II
RUN TIMES (MSEC) AND ESTIMATION ERRORS FOR DIFFERENT m WITH SNR = 40dB, $n = 4$ FIXED. FULLY SHUFFLED DATA, 100 TRIALS

m	Running Time		Estimation Error		
	$\tilde{\mathcal{P}}$	AM	ξ_{AI}^*	ξ_{AI}	AI-EM
1,000	10	15	3.2%	3.3%	0.4%
5,000	21	105	3.5%	4.8%	2.5%
10,000	32	281	4.1%	5.9%	4.3%
25,000	66	357	3.4%	5.2%	4.4%
50,000	126	613	3.7%	6.0%	5.5%
100,000	268	1,271	4.1%	6.8%	6.5%

by the values of n that we are considering. Interestingly, the estimation error of the AI-EM algorithm is also stable and of the order of 0.1% regardless of n .

Table II attests to the scalability of AI-EM in terms of m . Fixing $n = 4$ and $\sigma = 0.01$, the table reports running times and estimation errors for different values of m , ranging from $m = 10^3$ to $m = 10^5$. Indeed, solving the polynomial system requires 10 msec for $m = 10^3$ and only 268 msec for $m = 10^5$. The increase in the running time of the alternating minimization from 15 msec to 1.3 sec is more noticeable due to the complexity $\mathcal{O}(m \log m)$ of the sorting operation required to compute the optimal permutation at each iteration. For $m = 1000$ the estimation error of AI-EM is only 0.4%,

but as m increases, the criterion (60) for selecting the best root of $\tilde{\mathcal{P}}$ becomes less accurate; e.g., for $m = 10^5$, ξ_{AI}^* is 4.1% away from ξ^* , as opposed to 6.8% for ξ_{AI} .

IV. CONCLUSION

We established the theory of an algebraic geometric solution to the problem of linear regression without correspondences. The main object of interest was a polynomial system \mathcal{P} of n equations in n unknowns constraining the parameters $x \in \mathbb{R}^n$ of the linear regression model. The main theoretical contribution was establishing that \mathcal{P} is almost always consistent even for corrupted data with at most $n!$ complex roots. Our algorithmic proposal consisted of solving \mathcal{P} and using its most appropriate root as initialization to the Expectation-Maximization algorithm. This yielded an efficient solution for small values of n , stable under mild levels of noise and outperforming existing state-of-the-art methods.

APPENDIX A

GEOMETRIC CHARACTERIZATION OF DIMENSION

Let f_1, \dots, f_s be polynomials in $\mathbb{R}[x] = \mathbb{R}[x_1, \dots, x_n]$ and consider their common root locus $\mathcal{V}_{\mathbb{R}^n}(f_1, \dots, f_s)$, called an *algebraic variety*, defined as

$$\mathcal{V}_{\mathbb{R}^n}(f_1, \dots, f_s) := \{\xi \in \mathbb{R}^n : f_k(\xi) = 0, \forall k \in [s]\}. \quad (63)$$

What is the dimension of $\mathcal{V}_{\mathbb{R}^n}(f_1, \dots, f_s)$? If $s = 1$, then we have a single equation and one intuitively expects $\mathcal{V}_{\mathbb{R}^n}(f_1)$ to be a hypersurface of \mathbb{R}^n having dimension $n - 1$; this is in analogy with linear algebra, where a single linear equation defines a linear subspace of dimension one less than the ambient dimension. However, as Example 4 shows, it may be the case that $\mathcal{V}_{\mathbb{R}^n}(f_1)$ consists of a single point (or even no points at all), in which case $\dim \mathcal{V}_{\mathbb{R}^n}(f_1)$ should be zero (or -1 if the variety is empty), in analogy with linear algebra where a linear subspace has zero dimension only if it contains a single point (the origin 0).

To resolve the above issue and have a consistent definition of dimension that generalizes the linear algebraic one, it is necessary that we consider the common root locus of the polynomials in the *algebraic closure* \mathbb{C} of \mathbb{R} :

$$\mathcal{V}_{\mathbb{C}^n}(f_1, \dots, f_s) := \{\xi \in \mathbb{C}^n : f_k(\xi) = 0, \forall k \in [s]\}. \quad (64)$$

In that case, there is a well developed theory ([39]–[41]) that leads to consistent characterizations of the dimension of the geometric object $\mathcal{V}_{\mathbb{C}^n}(f_1, \dots, f_s) \subset \mathbb{C}^n$ and that of its algebraic counterpart $\{f_1, \dots, f_s\} \subset \mathbb{C}[x]$. The next definition provides the geometric characterization of $\dim \mathcal{V}_{\mathbb{C}^n}(f_1, \dots, f_s)$.

Definition A1: Defining $\mathcal{Y} \subset \mathbb{C}^n$ to be closed if it is of the form $\mathcal{Y} = \mathcal{V}(g_1, \dots, g_r)$ for some polynomials $g_1, \dots, g_r \in \mathbb{C}[x]$, and irreducible if it is not the union of two proper closed subsets, $\dim \mathcal{V}_{\mathbb{C}^n}(f_1, \dots, f_s)$ is defined to be the largest¹¹ non-negative integer d such that there exists a chain of the form

$$\mathcal{V}_{\mathbb{C}^n}(f_1, \dots, f_s) \supset \mathcal{Y}_0 \supsetneq \mathcal{Y}_1 \supsetneq \mathcal{Y}_2 \supsetneq \dots \supsetneq \mathcal{Y}_d, \quad (65)$$

¹¹The acute reader may notice that there is a-priori no guarantee that such a maximal integer exists. However, this is true because \mathbb{C}^n is a *Noetherian topological space*, a technical notion that is beyond the scope of this paper.

where each \mathcal{Y}_i is a closed irreducible subset of $\mathcal{V}_{\mathbb{C}^n}(f_1, \dots, f_s)$.

Definition A1 is a generalization of the notion of dimension in linear algebra: if \mathcal{Y} is a linear subspace of \mathbb{C}^n , then $\dim \mathcal{Y}$ is precisely equal to the maximal length of a descending chain of linear subspaces that starts with \mathcal{Y} ; one can get such a chain by removing a single basis vector of \mathcal{Y} at each step.¹²

Example A1: With e_i the vector with zeros everywhere except a 1 at position i , and $\mathcal{Y}_i = \text{Span}(e_1, \dots, e_{n-i})$, \mathbb{C}^n admits a chain

$$\mathbb{C}^n = \mathcal{Y}_0 \supsetneq \mathcal{Y}_1 \supsetneq \mathcal{Y}_2 \supsetneq \dots \supsetneq \mathcal{Y}_{n-1} \supsetneq \mathcal{Y}_n := \{0\}. \quad (66)$$

A very important structural fact about algebraic varieties is the following decomposition theorem.

Proposition A1: Let $\mathcal{Y} = \mathcal{V}_{\mathbb{C}^n}(f_1, \dots, f_s)$ for some $f_i \in \mathbb{C}[x]$. Then \mathcal{Y} can be written uniquely as $\mathcal{Y} = \mathcal{Y}_1 \cup \dots \cup \mathcal{Y}_\ell$, for some positive integer ℓ , where the \mathcal{Y}_i are irreducible closed sets of \mathbb{C}^n (see Definition A1), and they are minimal, in the sense that if one removes one of the \mathcal{Y}_i , the resulting union is a strictly smaller set than \mathcal{Y} . The \mathcal{Y}_i are called the irreducible components of \mathcal{Y} .

Definition A1 together with Proposition A1 ensure that the only algebraic varieties $\mathcal{V}_{\mathbb{C}^n}(f_1, \dots, f_s)$ that have dimension zero are the ones that consist of a finite number of points; these are precisely the varieties of interest in this paper.

Proposition A2: Let $\mathcal{Y} = \mathcal{V}_{\mathbb{C}^n}(f_1, \dots, f_s)$. Then $\dim \mathcal{Y} = 0$ if and only if \mathcal{Y} consists of a finite number of points of \mathbb{C}^n .

APPENDIX B

ALGEBRAIC CHARACTERIZATION OF DIMENSION

Even though Definition A1 is quite intuitive, it is not as convenient to use in practice, since one is usually given polynomials f_1, \dots, f_s and wants to determine whether $\mathcal{V}_{\mathbb{C}^n}(f_1, \dots, f_s)$ has zero dimension, without having to solve the polynomial system. This is the case in this paper, where, e.g., to prove Theorem 2 we need to show that $\mathcal{V}_{\mathbb{C}^n}(\hat{p}_1, \dots, \hat{p}_n)$ has zero dimension for any suitable A, y : clearly, computing the common root locus of $\hat{p}_1, \dots, \hat{p}_n$, as a function of A, y , is extremely challenging if not impossible (except for $n = 1, 2$). This is precisely where the algebraic characterization of $\dim \mathcal{V}_{\mathbb{C}^n}(f_1, \dots, f_s)$ comes in handy, since it allows its computation solely from the algebraic structure of f_1, \dots, f_s .

To introduce this algebraic notion of dimension we first need the notion of an *ideal* of $\mathbb{C}[x]$. Given polynomials $f_1, \dots, f_s \in \mathbb{C}[x]$, the ideal generated by these polynomials, denoted by (f_1, \dots, f_s) , is the set of all linear combinations of the f_i , but in contrast to linear algebra, the coefficients of the linear combination are allowed to be polynomials themselves:

$$(f_1, \dots, f_s) := \left\{ \sum_{i=1}^s g_i f_i, \forall g_i \in \mathbb{C}[x] \right\}. \quad (67)$$

Next we need the notion of a *prime ideal*. An ideal $\mathcal{P} \subsetneq \mathbb{C}[x]$ is called prime if it satisfies the following property: whenever the product of two polynomials is inside \mathcal{P} , then at least one of these polynomials must be inside \mathcal{P} . With that we have:

¹²For more information on the algebraic geometric structure of linear subspaces the reader is referred to Appendix C in [42].

Proposition B1: $\dim \mathcal{V}_{\mathbb{C}^n}(f_1, \dots, f_s)$ is the largest non-negative integer d such that there exists a chain of the form

$$\mathcal{P}_0 \supseteq \mathcal{P}_1 \supseteq \mathcal{P}_2 \supseteq \dots \supseteq \mathcal{P}_d \supset (f_1, \dots, f_s), \quad (68)$$

where each \mathcal{P}_i is a prime ideal of $\mathbb{C}[x]$.

Example B1: Continuing with Example A1, $\mathbb{C}^n = \mathcal{V}_{\mathbb{C}^n}(0)$ and

$$\mathcal{Y}_i := \text{Span}(e_1, \dots, e_{n-i}) = \mathcal{V}_{\mathbb{C}^n}(x_{n-i+1}, \dots, x_n). \quad (69)$$

Since every ideal of the form (x_{n-i+1}, \dots, x_n) is prime,¹³ we have the ascending chain of prime ideals of length n :

$$(x_1, \dots, x_n) \supseteq (x_2, \dots, x_n) \supseteq \dots \supseteq (x_n) \supseteq (0). \quad (70)$$

We close this section by noting that the main tool behind the proof of Proposition B1 is the famous *Hilbert's Nullstellensatz*, stated next, which holds over \mathbb{C} but not over \mathbb{R} . This is why we need to work over \mathbb{C} to get consistent geometric and algebraic characterizations of the dimension of an algebraic variety.

Proposition B2: Let f_1, \dots, f_s be polynomials of $\mathbb{C}[x]$. Then

- $\mathcal{V}_{\mathbb{C}^n}(f_1, \dots, f_s) = \emptyset$ if and only if $1 \in (f_1, \dots, f_s)$.
- Suppose that $\mathcal{V}_{\mathbb{C}^n}(f_1, \dots, f_s) \neq \emptyset$ and let f be a polynomial such that $f(\xi) = 0, \forall \xi \in \mathcal{V}_{\mathbb{C}^n}(f_1, \dots, f_s)$. Then $f^\ell \in (f_1, \dots, f_s)$ for some positive integer ℓ .

APPENDIX C

DIMENSION AND HOMOGENIZATION

A monomial of degree d is a polynomial of the form $x^\alpha := x_1^{\alpha_1} \dots x_n^{\alpha_n}$, where $\alpha = [\alpha_1, \dots, \alpha_n]^\top$ is a vector of non-negative integers such that $\alpha_1 + \dots + \alpha_n = d$. Every polynomial f can be uniquely written as a linear combination of monomials

$$f = \sum_{\alpha \in \mathcal{A}} c_\alpha x^\alpha \in \mathbb{C}[x], \quad (71)$$

where \mathcal{A} is a finite set of *multi-exponents* and $c_\alpha \in \mathbb{C}$ are the corresponding coefficients. Then f is called *homogeneous* of degree d if all its monomials have the same degree d .

Let f_1, \dots, f_s be a set of polynomials of $\mathbb{C}[x]$. For rather subtle reasons beyond the scope of this paper, further characterizing the dimension of $\mathcal{V}_{\mathbb{C}^n}(f_1, \dots, f_s)$ beyond Proposition B1 is simpler when all the f_i are homogeneous (this will be discussed in the next section). When this is not the case, there is a simple procedure called *homogenization*, through which we can convert non-homogenous polynomials to homogeneous.

Suppose that the polynomial f given in (71) is not homogeneous. Let d be the maximal degree among all the monomials of f , i.e., $d = \max\{\sum_{i \in [n]} \alpha_i : \alpha \in \mathcal{A}\}$. Then the homogenization of f is a polynomial in the extended polynomial ring $\mathbb{C}[x, t]$ with one additional variable t , defined as

$$f^{(h)} := \sum_{\alpha \in \mathcal{A}} c_\alpha x^\alpha t^{d - \sum_{i \in [n]} \alpha_i} \in \mathbb{C}[x, t]. \quad (72)$$

¹³For a justification, the reader is referred to Proposition 53 in [42].

Example C1: A non-homogeneous polynomial f of degree 6 and its homogenization $f^{(h)}$:

$$f = x_1^3 x_2^2 x_3 + x_2^2 + x_3 \in \mathbb{C}[x_1, x_2, x_3], \quad (73)$$

$$f^{(h)} = x_1^3 x_2^2 x_3 + x_2^2 t^4 + x_3 t^5 \in \mathbb{C}[x_1, x_2, x_3, t]. \quad (74)$$

Now, let \mathcal{I} be the ideal generated by some polynomials $f_1, \dots, f_s \in \mathbb{C}[x]$ and consider the homogenization of this ideal

$$\mathcal{I}^{(h)} = \{f^{(h)} : f \in \mathcal{I}\} \subset \mathbb{C}[x, t]. \quad (75)$$

We note here the subtle fact that $\mathcal{I}^{(h)}$ certainly contains $f_1^{(h)}, \dots, f_s^{(h)}$, but in principle it is larger than the ideal generated by $f_1^{(h)}, \dots, f_s^{(h)}$, as the next example illustrates.

Example C2: Let $f_1 = x_1^2 + x_2, f_2 = x_1^2 + x_3$ be polynomials of $\mathbb{C}[x_1, x_2, x_3]$. Then $f_1^{(h)} = x_1^2 + x_2 t, f_2^{(h)} = x_1^2 + x_3 t$. Now, the polynomial $x_2 - x_3 = f_1 - f_2$ is in the ideal $\mathcal{I} = (f_1, f_2)$, and it is already homogeneous so that $x_2 - x_3 \in \mathcal{I}^{(h)}$. However $x_2 - x_3$ is not inside the ideal $(f_1^{(h)}, f_2^{(h)}) = (x_1^2 + x_2 t, x_1^2 + x_3 t)$, since the latter only contains elements of degree 2 and higher.

Since the elements of $\mathcal{I}^{(h)}$ are polynomials in $n + 1$ variables, they define an algebraic variety¹⁴ $\mathcal{Y}^{(h)} = \mathcal{V}_{\mathbb{C}^{n+1}}(\mathcal{I}^{(h)})$ of \mathbb{C}^{n+1} . What is the relationship between \mathcal{Y} and $\mathcal{Y}^{(h)}$? It is actually not hard to see that if $[\xi_1, \dots, \xi_n]^\top$ is a point of \mathcal{Y} , then $\lambda[\xi_1, \dots, \xi_n, 1]^\top$ is a point of $\mathcal{Y}^{(h)}$, for any $\lambda \in \mathbb{C}$. Hence any non-zero point ξ of \mathcal{Y} gives rise to an entire line inside $\mathcal{Y}^{(h)}$; this line passes through the origin and ξ , and its intersection with the hyperplane $t = 1$ can be used to recover the original point ξ . Hence $\mathcal{Y}^{(h)}$ is called the *affine cone* over \mathcal{Y} with vertex $0 \in \mathbb{C}^{n+1}$. Moreover, the variety $\mathcal{Y} \subset \mathbb{C}^n$ is embedded inside the affine cone through a mapping that takes points to lines. In addition, $\mathcal{Y}^{(h)}$ contains so-called *points at infinity*, which are obtained by setting $t = 0$. As it turns out, there is a tight topological relationship between \mathcal{Y} and $\mathcal{Y}^{(h)}$ and the important fact for our analysis is the following dimension theorem; see [43] for a detailed discussion for the non-expert reader in the context of subspace clustering.

Proposition C1: Let \mathcal{Y} be an algebraic variety of \mathbb{C}^n and let $\mathcal{Y}^{(h)} \subset \mathbb{C}^{n+1}$ be its affine cone. Then $\dim \mathcal{Y} = \dim \mathcal{Y}^{(h)} - 1$.

Example C3: Let \mathcal{Y} be an affine line of \mathbb{C}^2 given by the equation $\alpha x_1 + \beta x_2 + \gamma = 0$. Then $\mathcal{Y}^{(h)}$ is a plane through the origin in \mathbb{C}^3 given by the equation $\alpha x_1 + \beta x_2 + \gamma t = 0$.

The next fact, known as *Bezout's Theorem*, will be used in bounding the number of points of the zero-dimensional variety of Theorem 2.

Proposition C2: Let h_1, \dots, h_n be homogeneous polynomials of $\mathbb{C}[x, t]$ of degrees $\deg(h_i) = d_i, i \in [n]$. If $\mathcal{V}_{\mathbb{C}^{n+1}}(h_1, \dots, h_n)$ is a finite union of lines through the origin, then the number of these lines is at most $d_1 d_2 \dots d_n$.

APPENDIX D

REGULAR SEQUENCES

In §II-B we argued that if the polynomials $\hat{p}_1, \dots, \hat{p}_n \in \mathbb{C}[x]$ are *sufficiently independent*, then

¹⁴Let $\mathcal{J} \subset \mathbb{C}[x]$ be an ideal. Then *Hilbert's Basis Theorem* guarantees that \mathcal{J} always has a finite set of generators, i.e., there is a positive integer ℓ and polynomials $g_1, \dots, g_\ell \in \mathbb{C}[x]$ such that $\mathcal{J} = (g_1, \dots, g_\ell)$.

$\dim \mathcal{V}_{\mathbb{C}^n}(\hat{p}_1, \dots, \hat{p}_n) = 0$, i.e., the dimension of the algebraic variety drops precisely by the number n of its defining equations. More generally, the precise notion of what *sufficiently independent* should mean for polynomials f_1, \dots, f_s , $s \leq n$, so that $\dim \mathcal{V}_{\mathbb{C}^n}(f_1, \dots, f_s) = n - s$, is easier to characterize when all the f_i are homogeneous. The right notion is that of a *regular sequence*.

Definition D1: Let f_1, \dots, f_s be polynomials of $\mathbb{C}[x]$. Then f_1, \dots, f_s is a regular sequence if $(f_1, \dots, f_s) \subsetneq \mathbb{C}[x]$, and for every $i = 2, \dots, s$ the following property is true: whenever there is a polynomial g such that $f_i g \in (f_1, \dots, f_{i-1})$, then we must have $g \in (f_1, \dots, f_{i-1})$.

The crucial fact for our analysis is the following.

Proposition D1: Let f_1, \dots, f_s , $s \leq n$, be non-constant homogeneous polynomials of $\mathbb{C}[x]$. Then $\dim \mathcal{V}_{\mathbb{C}^n}(f_1, \dots, f_s) = n - s$, if and only if f_1, \dots, f_s is a regular sequence.

Given a regular sequence of polynomials f_1, \dots, f_s in $\mathbb{C}[x]$ of length $s < n$, it is of interest to be able to augment this sequence to a regular sequence f_1, \dots, f_s, g of length $s + 1$. The simplest type of a homogeneous polynomial g that one may consider is a linear form $g = \ell^\top x$, which represents a hyperplane with normal vector $\ell \in \mathbb{C}^n$. As it turns out, almost all such hyperplanes qualify, with the exception of those with normal vector ℓ that lies inside an algebraic variety of \mathbb{C}^n determined by f_1, \dots, f_s .

Proposition D2: Let f_1, \dots, f_s , $s < n$, be a regular sequence of homogeneous polynomials of $\mathbb{C}[x]$. If $\ell \in \mathbb{C}^n$ is a generic vector, then $f_1, \dots, f_s, \ell^\top x$ is a regular sequence.

APPENDIX E INITIAL IDEALS

The notion of the *initial ideal* $\text{in}_<(\mathcal{I})$ of an ideal $\mathcal{I} \subset \mathbb{C}[x]$ with respect to a *monomial order* $<$ is a central one in computational algebraic geometry [44]. A more advanced object that is needed for our analysis in this paper is the initial ideal $\text{in}_w(\mathcal{I})$ of \mathcal{I} with respect to a *weight-order* [45], [46], which we introduce next.

Let $w = [w_1, \dots, w_n]^\top$ be a vector of positive integers. To each variable x_i of $\mathbb{C}[x]$ we assign the *weight* w_i , and to each monomial $x^\alpha = x_1^{\alpha_1} \cdots x_n^{\alpha_n}$ the *weighted degree* $d_w(x^\alpha) := w_1 \alpha_1 + \cdots + w_n \alpha_n$. A polynomial f is called *w-homogeneous*, if all its monomials have the same weighted degree.

Example E1: Let $w = [1, 2, 3]^\top$ and let $f = x_1 x_2 + x_3$. Then f is not homogeneous in the usual sense (see Appendix C), but it is *w-homogeneous* of degree 3.

Now let $f \in \mathbb{C}[x]$ be any polynomial. Then f can be uniquely written as $f = f^{(d_1)} + f^{(d_2)} + \cdots + f^{(d_s)}$, with $d_1 > d_2 > \cdots > d_s > 0$, where each $f^{(d_i)}$ is a *w-homogeneous* polynomial of degree d_i . We define the *initial form* of f with respect to w as $\text{in}_w(f) := f^{(d_1)}$. Given an ideal $\mathcal{I} = (f_1, \dots, f_s)$, we define $\text{in}_w(\mathcal{I})$ to be the ideal generated by all initial forms $\text{in}_w(f)$ for all $f \in \mathcal{I}$. That is, $h \in \text{in}_w(\mathcal{I})$ if and only if there exist polynomials $g_i \in \mathbb{C}[x]$, $i \in [s]$, such that $h = \text{in}_w(g_1 f_1 + \cdots + g_s f_s)$.

Example E2: Let $f = x_1 x_2 + x_3 + x_1^2 + x_2 + x_1$ and w as in Example E1. Then $f = f^{(3)} + f^{(2)} + f^{(1)}$ with

$$f^{(3)} = x_1 x_2 + x_3, f^{(2)} = x_1^2 + x_2, f^{(1)} = x_1. \quad (76)$$

Moreover, $\text{in}_w(f) = x_1 x_2 + x_3$.

The initial ideal $\text{in}_w(\mathcal{I})$ is certainly a significantly simpler object than the ideal \mathcal{I} itself, since it retains only the initial information about \mathcal{I} , so to speak. What is remarkable though, is that many structural properties of \mathcal{I} are inherited from those of $\text{in}_w(\mathcal{I})$. For this paper, the most important relationship is that the varieties defined by these two ideals have the same dimension:

Proposition E1: Let $\mathcal{I} \subset \mathbb{C}[x]$ be an ideal, $w \in \mathbb{Z}_{>0}^n$ a weight, and $\text{in}_w(\mathcal{I}) \subset \mathbb{C}[x]$ the initial ideal of \mathcal{I} with respect to w . Then

$$\dim \mathcal{V}_{\mathbb{C}^n}(\mathcal{I}) = \dim \mathcal{V}_{\mathbb{C}^n}(\text{in}_w(\mathcal{I})). \quad (77)$$

Hence to compute the dimension of the algebraic variety defined by an ideal $\mathcal{I} = (f_1, \dots, f_s)$, we may instead use the simpler object $\text{in}_w(\mathcal{I})$. But how can we efficiently compute a set of generators for $\text{in}_w(\mathcal{I})$ given f_1, \dots, f_s ? Note here that $(\text{in}_w(f_1), \dots, \text{in}_w(f_s)) \subset \text{in}_w(\mathcal{I})$ but equality does not hold in general, as the next example shows.

Example E3: Let $w = [1, 2, 3]^\top$ and $\mathcal{I} = (f_1, f_2) \subset \mathbb{C}[x]$ with

$$f_1 = x_1^9 + x_3 + x_2, f_2 = x_1^9 + x_2 x_3. \quad (78)$$

Then $\text{in}_w(f_1) = \text{in}_w(f_2) = x_1^9$ and so $(\text{in}_w(f_1), \text{in}_w(f_2)) = (x_1^9) = \{x_1^9 g : \forall g \in \mathbb{C}[x]\}$. On the other hand $f_2 - f_1 = x_2 x_3 - x_3 - x_2 \in \mathcal{I}$ and so $\text{in}_w(f_2 - f_1) = x_2 x_3 \in \text{in}_w(\mathcal{I})$. But clearly, $x_2 x_3 \notin (x_1^9)$. Hence $(\text{in}_w(f_1), \text{in}_w(f_2)) \subsetneq \text{in}_w(\mathcal{I})$.

The next fact, well known to experts in *Gröbner basis* theory and otherwise much less known, describes a situation where we can directly get the generators of $\text{in}_w(\mathcal{I})$ from the generators of \mathcal{I} .

Proposition E2: Let $w \in \mathbb{Z}_{>0}^n$ be a weight, and f_1, \dots, f_s be polynomials of $\mathbb{C}[x]$, such that $\text{in}_w(f_1), \dots, \text{in}_w(f_s)$ is a regular sequence. Then $\text{in}_w(\mathcal{I}) = (\text{in}_w(f_1), \dots, \text{in}_w(f_s))$.

REFERENCES

- [1] A. M. Legendre, *Nouvelles Méthodes Pour la détermination des Orbites des Comètes*. Paris, France: Firmin Didot, 1805.
- [2] C. F. Gauss, *Theoria Motus Corporum Coelestium in Sectionibus Conicis Solem Ambientum*. New York, NY, USA: Cambridge Univ. Press, 1809.
- [3] R. Maronna, "Principal components and orthogonal regression based on robust scales," *Technometrics*, vol. 47, no. 3, pp. 264–273, Aug. 2005.
- [4] Y. Wang, C. Dicle, M. Sznajder, and O. Camps, "Self scaled regularized robust regression," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 3261–3269.
- [5] G. Lerman and T. Maunu, "An overview of robust subspace recovery," *Proc. IEEE*, vol. 106, no. 8, pp. 1380–1410, Aug. 2018.
- [6] M. C. Tsakiris and R. Vidal, "Dual principal component pursuit," *J. Mach. Learn. Res.*, vol. 19, no. 18, pp. 1–49, 2018.
- [7] R. A. Maronna, R. D. Martin, and V. J. Yohai, *Robust Statistics: Theory and Methods* (Wiley Series in Probability and Statistics). Chichester, U.K.: Wiley, 2006.
- [8] D. J. Hsu, K. Shi, and X. Sun, "Linear regression without correspondence," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, 2017, pp. 1531–1540.
- [9] A. Pananjady, M. J. Wainwright, and T. A. Courtade, "Linear regression with shuffled data: Statistical and computational limits of permutation recovery," *IEEE Trans. Inf. Theory*, vol. 64, no. 5, pp. 3286–3300, May 2018.

- [10] J. Unnikrishnan, S. Haghghatshoar, and M. Vetterli, "Unlabeled sensing with random linear measurements," *IEEE Trans. Inf. Theory*, vol. 64, no. 5, pp. 3237–3253, May 2018.
- [11] X. Song, H. Choi, and Y. Shi, "Permuted linear model for header-free communication via symmetric polynomials," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2018, pp. 661–665.
- [12] S. Thrun and J. J. Leonard, *Simultaneous Localization and Mapping*. Berlin, Germany: Springer, 2008, pp. 871–889.
- [13] J. Unnikrishnan and M. Vetterli, "Sampling and reconstruction of spatial fields using mobile sensors," *IEEE Trans. Signal Process.*, vol. 61, no. 9, pp. 2328–2340, May 2013.
- [14] J. Unnikrishnan and M. Vetterli, "Sampling high-dimensional bandlimited fields on low-dimensional manifolds," *IEEE Trans. Inf. Theory*, vol. 59, no. 4, pp. 2103–2127, Apr. 2013.
- [15] R. Ji, Y. Liang, L. Xu, and W. Zhang, "A concave optimization-based approach for joint multi-target track initialization," *IEEE Access*, vol. 7, pp. 108551–108560, 2019.
- [16] P. David, D. DeMenthon, R. Duraiswami, and H. Samet, "SoftPOSIT: Simultaneous pose and correspondence determination," *Int. J. Comput. Vis.*, vol. 59, no. 3, pp. 259–284, Sep. 2004.
- [17] M. Marques, M. Stosic, and J. Costeira, "Subspace matching: Unique solution to point matching with geometric constraints," in *Proc. IEEE 12th Int. Conf. Comput. Vis.*, Sep. 2009, pp. 1288–1294.
- [18] M. Slawski and E. Ben-David, "Linear regression with sparsely permuted data," *Electron. J. Statist.*, vol. 13, no. 1, pp. 1–36, 2019.
- [19] M. Slawski, E. Ben-David, and P. Li, "A two-stage approach to multivariate linear regression with sparsely mismatched data," 2019, *arXiv:1907.07148*. [Online]. Available: <http://arxiv.org/abs/1907.07148>
- [20] P. Lahiri and M. D. Larsen, "Regression analysis with linked data," *J. Amer. Stat. Assoc.*, vol. 100, no. 469, pp. 222–230, 2005.
- [21] C. Rose, I. Saira Mian, and R. Song, "Timing channels with multiple identical quanta," 2012, *arXiv:1208.1070*. [Online]. Available: <http://arxiv.org/abs/1208.1070>
- [22] X. Huang and A. Madan, "CAP3: A DNA sequence assembly program," *Genome Res.*, vol. 9, no. 9, pp. 868–877, Sep. 1999.
- [23] C. Rose and I. S. Mian, "Signaling with identical tokens: Upper bounds with energy constraints," in *Proc. IEEE Int. Symp. Inf. Theory*, Jun. 2014, pp. 1817–1821.
- [24] A. Narayanan and V. Shmatikov, "Robust de-anonymization of large sparse datasets," in *Proc. IEEE Symp. Secur. Privacy*, May 2008, pp. 111–125.
- [25] L. Keller, M. Jafari Siavoshani, C. Fragouli, K. Argyraki, and S. Diggavi, "Identity aware sensor networks," in *Proc. IEEE INFOCOM 28th Conf. Comput. Commun.*, Apr. 2009, pp. 2177–2185.
- [26] A. Balakrishnan, "On the problem of time jitter in sampling," *IRE Trans. Inf. Theory*, vol. 8, no. 3, pp. 226–236, Apr. 1962.
- [27] W. S. Robinson, "A method for chronologically ordering archaeological deposits," *Amer. Antiquity*, vol. 16, no. 4, pp. 293–301, 1951.
- [28] A. Abid, A. Poon, and J. Zou, "Linear regression with shuffled labels," 2017, *arXiv:1705.01342*. [Online]. Available: <http://arxiv.org/abs/1705.01342>
- [29] A. Abid and J. Zou, "Stochastic EM for shuffled linear regression," 2018, *arXiv:1804.00681*. [Online]. Available: <http://arxiv.org/abs/1804.00681>
- [30] X.-D. Luo and Z.-Q. Luo, "Extension of Hoffman's error bound to polynomial systems," *SIAM J. Optim.*, vol. 4, no. 2, pp. 383–392, 1994.
- [31] A. Pananjady, M. J. Wainwright, and T. A. Courtade, "Denosing linear models with permuted data," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2017, pp. 446–450.
- [32] S. Gleichman and Y. C. Eldar, "Blind compressed sensing," *IEEE Trans. Inf. Theory*, vol. 57, no. 10, pp. 6958–6975, Oct. 2011.
- [33] S. Haghghatshoar and G. Caire, "Signal recovery from unlabeled samples," *IEEE Trans. Signal Process.*, vol. 66, no. 5, pp. 1242–1257, Mar. 2018.
- [34] I. Dokmanić, "Permutations unlabeled beyond sampling unknown," *IEEE Signal Process. Lett.*, vol. 26, no. 6, pp. 823–827, Jun. 2019.
- [35] M. C. Tsakiris and L. Peng, "Homomorphic sensing," in *Proc. 36th Int. Conf. Mach. Learn.*, 2019, pp. 6335–6344.
- [36] M. C. Tsakiris, "Eigenspace conditions for homomorphic sensing," 2018, *arXiv:1812.07966*. [Online]. Available: <http://arxiv.org/abs/1812.07966>
- [37] L. Peng, X. Song, M. C. Tsakiris, H. Choi, L. Kneip, and Y. Shi, "Algebraically-initialized expectation maximization for header-free communication," in *Proc. ICASSP—IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2019, pp. 5182–5186.
- [38] G. Elhami, A. Scholefield, B. Bejar Haro, and M. Vetterli, "Unlabeled sensing: Reconstruction algorithm and theoretical guarantees," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2017, pp. 4566–4570.
- [39] R. Hartshorne, *Algebraic Geometry*. New York, NY, USA: Springer, 1977.
- [40] H. Matsumura, *Commutative Ring Theory* (Cambridge Studies in Advanced Mathematics). Cambridge, U.K.: Cambridge Univ. Press, 1989.
- [41] D. Eisenbud, *Commutative Algebra With a View Toward Algebraic Geometry*. New York, NY, USA: Springer, 1994.
- [42] M. C. Tsakiris and R. Vidal, "Filtrated algebraic subspace clustering," *SIAM J. Imag. Sci.*, vol. 10, no. 1, pp. 372–415, Jan. 2017.
- [43] M. C. Tsakiris and R. Vidal, "Algebraic clustering of affine subspaces," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 2, pp. 482–489, Feb. 2018.
- [44] D. Cox, J. Little, and D. O'Shea, *Ideals, Varieties, and Algorithms*. Cham, Switzerland: Springer, 2007.
- [45] W. Bruns and A. Conca, "Gröbner bases and determinantal ideals," in *Commutative Algebra, Singularities and Computer Algebra*. Dordrecht, The Netherlands: Springer, 2003, pp. 9–66.
- [46] J. Herzog and T. Hibi, *Monomial Ideals*. London, U.K.: Springer, 2011.
- [47] A. Conca, C. Krattenthaler, and J. Watanabe, "Regular sequences of symmetric polynomials," *Rend. Semin. Mat. Univ. Padova*, vol. 121, pp. 179–199, 2009.
- [48] D. Lazard, "Thirty years of polynomial system solving, and now?" *J. Symbolic Comput.*, vol. 44, no. 3, pp. 222–231, Mar. 2009.
- [49] J.-C. Faugere, M. Safey El Din, and T. Verron, "On the complexity of computing Gröbner bases for weighted homogeneous systems," *J. Symbolic Comput.*, vol. 76, pp. 107–141, Sep. 2016.
- [50] D. Bates, J. Hauenstein, S. Andrew, and C. Wampler. (2013). *Bertini: Software for Numerical Algebraic Geometry*. Accessed: 2019. [Online]. Available: <https://bertini.nd.edu/>
- [51] L. Kneip. (2015). *Polyjam*. Accessed: 2019. [Online]. Available: <https://github.com/laurentkneip/polyjam>
- [52] Z. Kukulova, M. Bujnak, and T. Pajdla, "Automatic generator of minimal problem solvers," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2008, pp. 302–315.

Manolis C. Tsakiris received the Dip.Eng. degree from the National Technical University of Athens, the M.S. degree from Imperial College London, and the Ph.D. degree from Johns Hopkins University, all in electrical and electronic engineering and computer science. Since August 2017, he has been an Assistant Professor with the School of Information Science and Technology, ShanghaiTech University. He pursues research in the area of subspace learning with a particular focus on problems of algebraic geometric nature. He is also researching pure aspects of commutative algebra.

Liangzu Peng received the bachelor's degree in engineering from Zhejiang University in 2017. He is currently a Graduate Student with the School of Information Science and Technology, ShanghaiTech University, advised by M. C. Tsakiris. He is interested in theoretical and algorithmic questions that arise when learning algebraic, geometric or combinatorial models from data, such as in computer vision, machine learning, and signal processing.

Aldo Conca has been a Professor of algebra with the University of Genova since 2000, where he currently serves as the Director of the Mathematics Department. He has authored more than 70 research articles in commutative algebra and its interactions with algebraic geometry and combinatorics, published in journals, such as *Advances in Mathematics*, *Mathematische Zeitschrift*, *Mathematische Annalen*, *Compositio Mathematica*, *Crelle's Journal*, *Journal of Algebra*, *TAMS*, *PAMS*, *American Journal of Mathematics*, *Journal of Algebra & Number Theory*, and *Journal of Pure and Applied Algebra*.

Laurent Kneip received the Dipl.Ing. degree from Friedrich-Alexander Universität Erlangen Nürnberg and the Ph.D. degree from ETH Zürich. He worked at the Autonomous Systems Lab, ETH Zürich. He currently holds a position as a tenure-track Assistant Professor with the School of Information Science and Technology, ShanghaiTech University, where he founded and directs the Mobile Perception Laboratory (MPL). He has published more than 70 articles mostly in top conferences and journals, such as ICCV, CVPR, ECCV, ICRA, TPAMI, and TRO. He is the main author of the open-source Project OpenGV. His long-term contributions in geometric computer vision have, for example, been awarded the Marr Prize (honorable mention) at ICCV in 2017. He was a former recipient of the Discovery Early Career Researcher Award (DECRA) from the Australian Research Council, hosted by the Australian National University, and a former Associate Investigator of the Australian Centre of Excellence for Robotic Vision.

Hayoung Choi received the B.S. and M.S. degrees from the Department of Mathematics, Kyungpook National University, Deagu, South Korea, in 2006 and 2012, respectively, and the Ph.D. degree from the Department of Mathematics, University of Wyoming, in 2015. He was previously a Post-Doctoral Researcher with the School of Information Science and Technology, ShanghaiTech University, in China. He is currently working as a Korea Research Fellow (KRF) with the Research Institute of Mathematics, Seoul National University, South Korea. His research concerns applied functional analysis, matrix analysis, large-scale data analysis, and applications of these topics to problems in machine learning, data mining, and communications.

Yuanming Shi (Member, IEEE) received the B.S. degree from Tsinghua University and the Ph.D. degree from The Hong Kong University of Science and Technology. He is currently a tenured Associate Professor with the School of Information Science and Technology, ShanghaiTech University. His research areas include optimization, statistics, machine learning, and signal processing and their applications to 6G, IoT, AI, and FinTech.