An Incidence Geometry approach to Dictionary Learning*

Meera Sitharam† Mohamad Tarifi‡ Menghan Wang§

Abstract

We study the Dictionary Learning (aka Sparse Coding) problem of obtaining a sparse representation of data points, by learning *dictionary vectors* upon which the data points can be written as sparse linear combinations. We view this problem from a geometry perspective as the spanning set of a subspace arrangement, and focus on understanding the case when the underlying hypergraph of the subspace arrangement is specified. For this Fitted Dictionary Learning problem, we completely characterize the combinatorics of the associated subspace arrangements (i.e. their underlying hypergraphs), using incidence geometry. Specifically, a combinatorial rigidity-type theorem is proven that characterizes the hypergraphs of subspace arrangements that generically yield (a) at least one dictionary (b) a locally unique dictionary (i.e. at most a finite number of isolated dictionaries) of the specified size. We are unaware of prior application of combinatorial rigidity techniques in the setting of Dictionary Learning, or even in machine learning. We also provide a systematic classification of problems related to Dictionary Learning together with various approaches, assumptions required and performance.

1 Introduction

Dictionary Learning (aka Sparse Coding) is the problem of obtaining a sparse representation of data points, by learning *dictionary vectors* upon which the data points can be written as sparse linear combinations.

Problem 1 (Dictionary Learning) *A point set* $X = [x_1 \dots x_m]$ *in* $\mathbb{R}^d$ *is said to be* s-represented *by a* dictionary $D = [v_1 \dots v_n]$ *for a given* sparsity $s < d$, *if there exists* $\Theta = [\theta_1 \dots \theta_m]$ *such that* $x_i = D\theta_i$, *with* $\|\theta_i\|_0 \leq s$. *Given an X known to be s-represented by an unknown* dictionary D *of size* $|D| = n$, Dictionary Learning *is the problem of finding any dictionary* $\acute{D}$ *satisfying the properties of* D, *i.e.* $|\acute{D}| \leq n$, *and there exists* $\acute{\Theta}_i$ *such that* $x_i = \acute{D}\acute{\theta}_i$ *for all* $x_i \in X$.

*This research was supported in part by the research grant NSF CCF-1117695 and a research gift from SolidWorks.

†Department of Computer and Information Science and Engineering, University of Florida, `sitharam@cise.ufl.edu`

‡Google, `tarifi@google.com`

§Department of Computer and Information Science and Engineering, University of Florida, `menghan@cise.ufl.edu`

The Dictionary Learning problem arises in various contexts such as signal processing and machine learning. The dictionary under consideration is usually *overcomplete*, with $n > d$. However we are interested in asymptotic performance with respect to all four variables n, m, d, s. Typically, $m \gg n \gg d > s$. Both cases when s is large relative to d and when s is small relative to d are interesting.

1.1 Previous approaches and challenges

Several traditional Dictionary Learning algorithms work by *alternating minimization*, i.e. iterating the two steps of Vector Selection which finds a representation Θ of X in an estimated dictionary, and updating the Dictionary estimation by solving an optimization problem that is convex in D when Θ is known [25, 20, 19].

For an overcomplete dictionary, the general vector selection problem is ill defined and has been shown to be NP-hard [21]. One is then tempted to conclude that Dictionary Learning is also NP-hard. However, this cannot be directly deduced in general, since even though adding a witness D turns the problem into an NP-hard problem, it is possible that the Dictionary Learning solution produces a different dictionary $\acute{D}$. On the other hand, if D satisfies the condition of being a *frame*, i.e. for all θ such that $\|\theta\|_0 \leq s$, there exists a δ_s such that $(1 - \delta_s) \leq \frac{\|D\theta\|_2^2}{\|\theta\|_2^2} \leq (1 + \delta_s)$, it is guaranteed that the sparsest solution to the Vector Selection problem can be found via l_1 minimization [8, 7].

One popular alternating minimization method is the Method of Optimal Dictionary (MOD) [9], which uses a maximum likelihood formalism, and compute D via the pseudoinverse. Another method k-SVD [2] updates D by taking every atom in D and applying SVD to X and Θ restricted to only the columns that have contribution from that atom.

Though alternating minimization methods work well in practice, there is no theoretical guarantee that the their results will converge to a true dictionary. Several recent works give provable algorithms under stronger constraints on X and D. Spielman et. al [24] give an l_1 minimization based approach which is provable to find the exact dictionary D, but requires D to be a basis. Arora et. al [3] and Agarwal et. al [1] independently give provable non-iterative algorithms for learning approximation of overcomplete dictionaries. Both of their methods are based on an overlapping clustering approach to find data points sharing a dictionary vector, and then estimate the dictionary vectors from the clusters via SVD. However, their algorithms require the dictionaries to be *pairwise incoherent* which is much stronger than the frame property.

In this paper, we understand the Dictionary Learning problem from an intrinsically geometric point of view. Notice that each $x \in X$ lies in an s-dimensional subspace $supp_D(x)$, which is the span of s vectors $v \in D$ that form the *support* of x. The resulting *s-subspace arrangement* $S_{X,D} = \{(x, supp_D(x)) : x \in X\}$ has an underlying labeled (multi)hypergraph $H(S_{X,D}) = (\mathfrak{I}(D), \mathfrak{I}(S_{X,D}))$, where $\mathfrak{I}(D)$ denotes the index set of the dictionary D and $\mathfrak{I}(S_{X,D})$ is the set of (multi)hyperedges over the indices $\mathfrak{I}(D)$ corresponding to the labeled sets $(x, supp_D(x))$. The word "multi" appears because if $supp_D(x_1) = supp_D(x_2)$ for data points $x_1, x_2 \in X$ with $x_1 \neq x_2$, then that support set of dictionary vectors (resp. their indices) is multiply represented in $S_{X,D}$ (resp. $\mathfrak{I}(S_{X,D})$) as labeled sets $(x_1, supp_D(x_1))$ and $(x_2, supp_D(x_2))$. We denote the sizes of these multisets as $|S_{X,D}|$ (resp. $|\mathfrak{I}(S_{X,D})|$).

Note that there could be many dictionaries D for the same set X of data points and for each D, many possible subspace arrangements $S_{X,D}$ that are solutions to the Dictionary Learning problem.

2 Contributions

In this paper, we focus on the version of Dictionary Learning where the underlying hypergraph is specified.

Problem 2 (Fitted Dictionary Learning) *Let X be a given set of data points in $\mathbb{R}^d$. For an unknown dictionary $D = [v_1, \ldots, v_n]$ that s-represents X, we are given the hypergraph $H(S_{X,D})$ of the underlying subspace arrangement $S_{X,D}$. Find any dictionary $\acute{D}$ of size $|\acute{D}| \leq n$ consistent with the hypergraph $H(S_{X,D})$.*

Our contributions in this paper are as follows:

- As the *main result*, we use combinatorial rigidity techniques to obtain a complete characterization of the hypergraphs $H(S_{X,D})$ that generically yield (a) at least one solution dictionary D, and (b) a locally unique solution dictionary D (i.e. at most a finite number of isolated solution dictionaries) of the specified size (see Theorem 2). To the best of our knowledge, this paper pioneers the use of combinatorial rigidity for problems related to Dictionary Learning.

- We are interested in minimizing $|D|$ for general X. However, as a *corollary of the main result*, we obtain that if the data points in X are highly general, for example, picked uniformly at random from the sphere S^{d-1}, then when s is fixed, $|D| = \Omega(|X|)$ with probability 1 (see Corollary 4).

- As a *corollary to our main result*, we obtain an Dictionary Learning algorithm for sufficiently general data X, i.e. requiring sufficiently large dictionary size n (see Corollary 5).

- We provide a systematic classification of problems related to Dictionary Learning together with various approaches, conditions and performance (see Section 4).

Note that although our results are stated for uniform hypergraphs $H(S_{X,D})$ (i.e. each subspace in $S_{X,D}$ has the same dimension), they can be easily generalized to non-uniform underlying hypergraphs.
Remark on technical significance: in this paper, we follow [4] and [31] to give a complete combinatorial characterization for the Fitted Dictionary Learning problem, starting from the initial representation as a nonlinear algebraic system. For more details of technical challenges and significance, see Section 3.4.

3 Main Result: Combinatorial Rigidity Characterization for Dictionary Learning

In this section, we present the main result of the paper, i.e. a combinatorial characterization of the (multi)hypergraphs H such that the existence and local uniqueness of a dictionary D is guaranteed for generic X satisfying $H(S_{X,D}) = H$.

Since the magnitudes of the vectors in X or D are uninteresting, we treat the data and dictionary points as living in the projective $(d-1)$ space and use the same notation to refer to both original d-dimensional and projective $d-1$ dimensional versions when the meaning is clear from the context. We rephrase the Fitted Dictionary Learning problem as the following Pinned Subspace-Incidence problem for the convenience of applying machinery from incidence geometry.

Problem 3 (Pinned Subspace-Incidence Problem) *Let X be a given set of m points (pins) in $\mathbb{P}^{d-1}(\mathbb{R})$. For every pin $x \in X$, we are also given the hyperedge $supp_D(x)$, i.e, an index subset of an unknown set of points $D = \{v_1, \ldots, v_n\}$, such that x_i lies on the subspace spanned by $supp_D(x)$. Find any such set D that satisfies the given subspace incidences.*

3.1 Algebraic Representation

We represent the Pinned Subspace-Incidence problem in the tradition of geometric constraint solving [6, 23], and view the problem as finding the common solutions of a system of polynomial equations (finding a real algebraic variety).

Consider a pin x_k on the subspace spanned by points $v_1^k, v_2^k, \ldots, v_s^k$. Using homogeneous coordinates, we can write this incidence constraint by letting all the $s \times s$ minors of the $(d-1) \times s$ matrix

$$E^k = \left[\begin{array}{cccc} v_1^k - x_k & v_2^k - x_k & \ldots & v_s^k - x_k \end{array}\right]$$

be zero, where $v_i^k = (v_{i,1}^k, v_{i,2}^k, ..., v_{i,d-1}^k)$ and $x_k = (x_{k,1}, x_{k,2}, ..., x_{k,d-1})$. So each incidence can be written as $\binom{d-1}{s}$ equations, where any $d-s$ of them are independent.

As the hypergraph $H = H(S_{X,D})$ of the underlying subspace arrangement has m (multi)hyperedges, the pinned subspace-incidence problem now reduces to solving a system of $m\binom{d-1}{s}$ equations denoted as

$$(H, X)(D) = 0 \qquad (1)$$

where $(H, X)(D)$ is a vector valued function from $\mathbb{R}^{n(d-1)}$ to $\mathbb{R}^{m\binom{d-1}{s}}$ parameterized by X.

Without any pins, the points in D have in total $n(d-1)$ degrees of freedom, and every pin (hyperedge) potentially removes $(d-s)$ degrees of freedom.

We use the underlying (multi)hypergraph $H(S_{X,D}) = (\mathfrak{I}(D), \mathfrak{I}(S_{X,D}))$ to define a *pinned subspace-incidence framework* (H, X, D), where $X : \{x_1, \ldots, x_m\} \subseteq \mathbb{R}^{d-1} \to \mathfrak{I}(S_{X,D})$ is an assignment of a given set of pins x_k to edges $X(x_k) = supp_D(x_k) \in \mathfrak{I}(S_{X,D})$, and $D : \mathfrak{I}(D) \to \mathbb{R}^{d-1}$ is an embedding of each vertex j into a point $v_j \in \mathbb{R}^{d-1}$, such that each pin x_k lies on the subspace spanned by $\{v_1^k, v_2^k, \ldots, v_s^k\}$. Two frameworks (H_1, X_1, D_1) and (H_2, X_2, D_2) are *equivalent* if $H_1 = H_2$ and $X_1 = X_2$, i.e. they satisfy the same algebraic equations for the same labeled hypergraph and ordered set of pins. They are *congruent* if they are equivalent and $D_1 = D_2$.

The pinned subspace-incidence system $(H, X)(D)$ is *independent* if none of the algebraic constraints is in the ideal generated by the others. Generally, independence implies the existence of a solution D to the system $(H, X)(D)$, where X is fixed. The system is *rigid* if there exist at most finitely many (real or complex) solutions. The system is *minimally rigid* if it is both rigid and independent. Rigidity is often defined (slightly differently) for individual frameworks. A framework (H, X, D) is *rigid* (i.e. *locally unique*) if there is a neighborhood $N(D)$, such that any framework (H, X, D') equivalent to (H, X, D) with $D' \in N(D)$ is also congruent to (H, X, D). A rigid framework (H, X, D) is *minimally rigid* if it becomes flexible after removing any pin.

We are interested in characterizing minimal rigidity of the pinned subspace-incidence system and framework. However, checking independence relative to the ideal generated by the variety is computationally hard and best known algorithms, such as Grobner basis, are exponential in time and space [17]. However, the algebraic system $(H, X)(D)$ can be linearized at *generic* or *regular* (non-singular) points whereby independence and rigidity reduces to linear independence and maximal rank at *generic* frameworks.

In algebraic geometry, a property being generic intuitively means that the property holds on the open dense complement of an (real) algebraic variety. Formally,

Definition 1 *A framework (H, X, D) is generic w.r.t. a property Q if and only if (H, X, D) avoids an algebraic variety V_Q specific to Q. In other words, there exists a neighborhood $N(D)$ such that for all frameworks (H, X, D') with $D' \in N(D)$, (H, X, D') satisfies Q if and only if (H, X, D) satisfies Q.*

A property Q of frameworks is generic (i.e. becomes a property of the hypergraph alone) if for all graphs H, either all generic (w.r.t. Q) frameworks satisfies Q, or all generic (w.r.t. Q) frameworks do not satisfy Q.

Once an appropriate notion of genericity is defined, we can treat Q as a property of a hypergraph. The primary activity of the area of combinatorial rigidity is to give purely combinatorial characterizations of such generic properties Q.

3.2 Linearization as Rigidity Matrix and its Generic Combinatorics

Next we follow the approach taken by traditional combinatorial rigidity theory [4, 12] to show that rigidity and independence (based on nonlinear polynomials) of pinned subspace-incidence systems are generically properties of the underlying hypergraph $H(S_{X,D})$, and can furthermore be captured by linear conditions in an infinitesimal setting. Specifically, Lemma 1 shows that rigidity of a pinned subspace-incidence system is equivalent to the existence of a full rank *rigidity matrix*, obtained by taking the Jacobian of the algebraic system $(H, X)(D)$ at a regular point.

A *rigidity matrix* of a framework (H, X, D) is a matrix whose kernel is the infinitesimal motions (flexes) of (H, X, D). A framework is *infinitesimally independent* if the rows of the rigidity matrix are independent. A framework is *infinitesimally rigid* if the space of infinitesimal motion is trivial, i.e. the rigidity matrix has full rank. A framework is *infinitesimally minimally rigid* if it is both infinitesimally independent and rigid.

To define a rigidity matrix for a pinned subspace-incidence framework (H, X, D), we take the Jacobian $J_X(D)$ of the algebraic system $(H, X)(D)$ by taking partial derivatives w.r.t. the coordinates of v_i's. In the Jacobian, each vertex v_i has $d-1$ corresponding columns, and each pin / hyperedge has $\binom{d-1}{s}$ corresponding rows, of which any $d-s$ rows are independent and span the rest. This $m\binom{d-1}{s}$ by $n(d-1)$ matrix is called the *symmetric rigidity matrix* M of the framework. If we choose $d-s$ rows per hyperedge in M, the obtained matrix $\hat{M}$

is a rigidity matrix of size $m(d-s)$ by $n(d-1)$. The framework is infinitesimally rigid if and only if there is an $\hat{M}$ with full rank. Note that the rank of a generic matrix $\hat{M}$ is at least as large as the rank of any specific realization $\hat{M}(H, X, D)$.

Defining generic as non-singular, for a generic framework (H, X, D), infinitesimal rigidity is equivalent to generic rigidity.

Lemma 1 *If D and X are regular / non-singular with respect to the system $(H, X)(D)$, then generic infinitesimal rigidity of the framework (H, X, D) is equivalent to generic rigidity.*

Remark: Pinned subspace-incidence frameworks are generalizations of related types of frameworks, such as pin-collinear body-pin frameworks [14], direction networks [32], slider-pinning rigidity [27], the molecular conjecture in 2D [22], body-cad constraint system [13, 16], k-frames [31], and affine rigidity [11].

3.3 Statement of main results

We study the rigidity matrix to obtain the following combinatorial characterization of (a) sparsity / independence, i.e. existence of a dictionary, and (b) rigidity, i.e. the solution set being locally unique / finite, for a pinned subspace-incidence framework.

Theorem 2 (Main Theorem) *A pinned subspace-incidence framework is generically minimally rigid if and only if the underlying hypergraph $H(S_{X,D}) = (\mathfrak{I}(D), \mathfrak{I}(S_{X,D}))$ satisfies $(d-s)|\mathfrak{I}(S_{X,D})| = (d-1)|\mathfrak{I}(D)|$ (i.e. $(d-s)|X| = (d-1)|D|$), and $(d-s)|E'| \leq (d-1)|V'|$ for every vertex induced subgraph $H' = (V', E')$. The latter condition alone ensures the independence of the framework.*

This combinatorial condition is actually $(d-1, 0)$-*tightness* of the hypergraph, which is a special case of the (k, l)-*tightness* condition that was widely studied in the geometric constraint solving and combinatorial rigidity literature before it was given a name in [15]. A hypergraph $H = (V, E)$ is $(k, 0)$-tight if $|E| = k|V|$, and for any $V' \subset V$, the induced subgraph $H' = (V', E')$ satisfies $|E'| \leq k|V'|$.

The proof of Theorem 2 adopts an approach by [31], in proving rigidity of k-frames, with the following outline:

- We obtain an *expanded multihypergraph* of $H(S_{X,D})$ by replacing each hyperedge with $(d-s)$ copies, in order to apply the $(k, 0)$-tightness condition.
- We show that for a specific form of the rows of a matrix defined on a *map-graph* (a relevant concept from hypergraph matroids corresponding to cycles of graphs), the determinant is not identically zero.
- Using a lemma from [26], which generalizes Tutte-Nash Williams [28, 18] and states that a hypergraph H is composed of k edge-disjoint map-graphs if and only if H is $(k, 0)$-tight, we apply Laplace decomposition to the $(d-1, 0)$-tight expanded multihypergraph as a union of $d-1$ maps. We then substitute in the result for map-graph and show that the determinant of the rigidity matrix is not identically zero, as long as the framework avoids a certain polynomial.
- The resulting polynomial is called the *pure condition* which characterizes the badly behaved cases (i.e. the conditions of non-genericity that the framework has to avoid for the combinatorial characterization to hold).

Example 1 *Figure 1 shows a pinned subspace-incidence framework with $d = 4, s = 2$. The expanded multihypergraph (replacing each hyperedge with 2 copies) satisfies $(3, 0)$-tightness condition, and the framework is minimally rigid.*

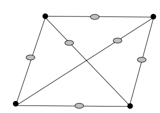


Figure 1: A minimally rigid pinned subspace-incidence framework of 6 pins and 4 vertices, with $d = 4, s = 2$.

One particular situation avoided by the pure condition is that there cannot be more than $s-1$ hyperedges containing the same set of vertices, namely, more than $s-1$ pins on the same subspace spanned by the dictionary vectors. Otherwise, s pins completely determine an s-subspace, whereby the vertices of the corresponding hyperedge have their degrees of freedom restricted and simple counterexamples to the characterization of the main theorem can be constructed.

Example 2 *Consider the framework in Figure 2 with $d = 3, s = 2$. There are $s = 2$ pins on each subspace. The expanded multihypergraph of the framework is $(2, 0)$-tight. However, the framework is obviously not minimally rigid.*

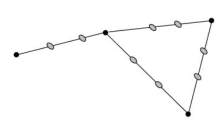


Figure 2: A pinned subspace-incidence framework of 8 pins and 4 vertices, with $d = 3, s = 2$, that violates the pure condition.

We relate the Fitted Dictionary Learning problem to the general Dictionary Learning problem, and give the lower bound of dictionary size for generic data points as a corollary to the main theorem.

Corollary 3 (Lower bound for generic data) *Given a set of m points $X = \{x_1, .., x_m\}$ in $\mathbb{R}^d$, generically there is a dictionary D of size n that s-represents X only if $(d - s)m \leq (d - 1)n$. Conversely, if $(d-s)m = (d-1)n$ and the supports of x_i (the nonzero entries of the θ_i's) are known to form a $(d-1, 0)$-tight hypergraph H, then generically, there is at least one and at most finitely many such dictionaries.*

Quantifying the term "generically" in Corollary 3 yields Corollaries 4 and 5 below, because the pure-conditions fail (i.e. the framework becomes non-generic) only on a measure-zero subset of the space of frameworks, and the number of possible underlying multi-hypergraphs is finite for a given set of pins.

Corollary 4 (Lower bound for highly general data) *Given a set of m points $X = \{x_1, .., x_m\}$ picked uniformly at random from the sphere S^{d-1}, a dictionary D that s-represents X has size at least $(\frac{d-s}{d-1})m$ with probability 1. In other words, $|D| = \Omega(X)$ if s and d are constants.*

Corollary 5 (Straightforward Learning Algorithm) *Given a set of m points $X = [x_1 \dots x_m]$ picked uniformly at random from the sphere S^{d-1}, we have a straightforward algorithm to construct a dictionary $D = [v_1 \dots v_n]$ that s-represents X, where $n = (\frac{d-s}{d-1})m$.*

The algorithm has two major parts: (1) constructing the underlying hypergraph $H(S_{X,D})$, and (2) constructing the s-subspace arrangement $S_{X,D}$ and the dictionary D. Part (1) starts from a minimal minimally rigid hypergraph H_0, and each following step appends a base structure B of $d - s$ vertices and $d - 1$ edges to the same set of base vertices in H_0. Both H_0 and B can be constructed using a modified version of the pebble game algorithm from [26]. Part (2) follows the graph construction of Part (1) by solving the size $O(d)$ algebraic system corresponding to B at each step, which takes $O(1)$ time when d is constant. Although there is more than one choice of solution for each step, since every graph construction step is based on the same set of base vertices, generically any choice will result in a successful solution for the entire sequence of steps.

3.4 Technical significance

In this paper, (1) we formulate the Pinned Subspace-Incidence problem as a nonlinear algebraic system $(H, X)(D)$. (2) We apply Asimow and Roth [4] to generically linearize $(H, X)(D)$, and (3) we apply White and Whiteley [31] to combinatorially characterize the rigidity of the underlying hypergraph $H(S_{X,D})$ and give the pure conditions. (4) Finally, in order to generalize the proof of (2) to hypergraphs, we use the map-graph characterization of Streinu and Theran [26], and adapt $H(S_{X,D})$ as expanded multihypergraphs. To our best knowledge, the only known results with a similar flavor are [13, 16] which characterize the rigidity of Body-and-cad frameworks. However, these results are dedicated to specific frameworks in 3D instead of arbitrary dimension subspace arrangements and hypergraphs, and their formulation process start directly with the linearized Jacobian.

4 Systematic classification of problems closely related to Dictionary Learning and previous approaches

By imposing a systematic series of increasingly stringent constraints on the input, we classify a whole set of independently interesting problems closely related to Dictionary Learning. A summarization of the input conditions and results of these different types of Dictionary Learning approaches can be found in Table 1 in Appendix A.

A natural restriction to the general Dictionary Learning is the following. We say that a set of data points X *lies on* a set S of s-dimensional subspaces if for all $x_i \in X$, there exists $S_i \in S$ such that $x_i \in S_i$.

Problem 4 (Subspace Arrangement Learning) *Let X be a given set of data points known to lie on a set S of s-dimensional subspaces of $\mathbb{R}^d$, with $|S| \leq k$.* Subspace arrangement learning *finds any subspace arrangement $\acute{S}$ of s-dimensional subspaces of $\mathbb{R}^d$ satisfying these conditions, i.e. $|\acute{S}| \leq k$, X lies on $\acute{S}$.*

There are several known algorithms for learning subspace arrangements. The Random Sample Consensus (RANSAC) method [29] learns subspace arrangements by isolating, one subspace at a time, via random sampling. Another method called Generalized PCA (GPCA) [30] uses techniques from algebraic geometry for subspace clustering, by factoring a homogeneous polynomial of degree k that is fitted to the points $\{x_1 \dots x_m\}$.

The next problem is obtaining a minimally sized dictionary from a subspace arrangement.

Problem 5 (Smallest Spanning Set for Arrangement) *Let S be a given set of s-dimensional subspaces of $\mathbb{R}^d$. Assume their intersections are known to be s-represented by a set I of vectors with $|I|$ at most n. Find any set of vectors $\acute{I}$ that satisfies these conditions.*

The smallest spanning set is not necessarily unique in general. This problem is closely related to the intersection semilattice of subspace arrrangement [5, 10].

When X contains sufficiently dense data to solve Problem 4, Dictionary Learning reduces to Problem 5, i.e. we can find D using the following two steps: (1)

Learn an s-subspace arrangement S for X. (2) Recover D by finding the smallest Spanning Set of S.

A natural restriction of this is when the data set X is given in *support-equivalence* classes. For a given subspace t in the subspace arrangement $S_{X,D}$ (respectively hyperedge h in the hypergraph's edge-set $\mathfrak{I}(S_{X,D})$), let $X_t = X_h \subseteq X$ be the equivalence class of data points x such that $span(supp_D(x)) = t$. We call the data points x in a same X_h as *support-equivalent*.

Problem 6 (Dictionary Learning for Partitioned Data) *Given data X partitioned into $X_i \subseteq X$,*
1. What is the minimum size of X and X_i's guaranteeing that there exists a locally unique dictionary D for a s-subspace arrangement $S_{X,D}$ satisfying $|D| \leq n$, and X_i represents the support-equivalence classes of X with respect to D?
2. How to find such a dictionary D?

With regard to the problem of minimizing $|D|$, very little is known for simple restrictions on X. For example the following question is open.

Question 1 *Given a general position assumption on X, are smaller dictionaries possible than indicated by Corollary 4? Conversely, what is the best lower bound on $|D|$ under such an assumption?*

The combinatorial characterization by Theorem 2 leads to the following question for general Dictionary Learning.

Question 2 *What is the minimum size of a data set X such that the Dictionary Learning Problem for X has a locally unique solution dictionary D of a given size? What are the geometric characteristics of such an X?*

A summarization of the input conditions and results of these different types of Dictionary Learning problems can be found in Table 1 in Appendix A.

5 Conclusion

In this paper, we approached Dictionary Learning from a geometric point of view.

We investigated Fitted Dictionary Learning theoretically using machinery from incidence geometry. Specifically, a combinatorial rigidity type theorem (our main result) is obtained which completely characterizes the subspace arrangements that are guaranteed to recover a finite number of dictionaries, using a purely combinatorial property on the supports. As corollaries of the main result, we gave lower bound for the size of dictionary when the data points are picked uniformly at random, and provided an algorithm for Dictionary Learning for such general data points. Additionally, we compare several closely related problems of independent interest, leading to different directions for future work.

References

[1] A. Agarwal, A. Anandkumar, and P. Netrapalli. Exact recovery of sparsely used overcomplete dictionaries. *arXiv preprint arXiv:1309.1952*, 2013.

[2] M. Aharon, M. Elad, and A. Bruckstein. k-svd: An algorithm for designing overcomplete dictionaries for sparse representation. *Signal Processing, IEEE Transactions on*, 54(11):4311–4322, 2006.

[3] S. Arora, R. Ge, and A. Moitra. New algorithms for learning incoherent and overcomplete dictionaries. *arXiv preprint arXiv:1308.6273*, 2013.

[4] L. Asimow and B. Roth. The rigidity of graphs. *Transactions of the American Mathematical Society*, 245:279–289, 1978.

[5] A. Björner. Subspace arrangements. In *First European Congress of Mathematics*, pages 321–370. Springer, 1994.

[6] B. Brüderlin and D. Roller. *Geometric constraint solving and applications*. Springer, 1998.

[7] E. J. Candès, J. Romberg, and T. Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *Information Theory, IEEE Transactions on*, 52(2):489–509, 2006.

[8] D. L. Donoho. Compressed sensing. *Information Theory, IEEE Transactions on*, 52(4):1289–1306, 2006.

[9] K. Engan, S. O. Aase, and J. Hakon Husoy. Method of optimal directions for frame design. In *Acoustics, Speech, and Signal Processing, 1999 IEEE International Conference on*, volume 5, pages 2443–2446. IEEE, 1999.

[10] M. Goresky and R. MacPherson. *Stratified morse theory*. Springer, 1988.

[11] S. J. Gortler, C. Gotsman, L. Liu, and D. P. Thurston. On affine rigidity. *Journal of Computational Geometry*, 4(1):160–181, 2013.

[12] J. E. Graver, B. Servatius, and H. Servatius. *Combinatorial rigidity*, volume 2. AMS Bookstore, 1993.

[13] K. Haller, A. Lee-St John, M. Sitharam, I. Streinu, and N. White. Body-and-cad geometric constraint systems. *Computational Geometry*, 45(8):385–405, 2012.

[14] B. Jackson and T. Jordán. Pin-collinear body-and-pin frameworks and the molecular conjecture. *Discrete & Computational Geometry*, 40(2):258–278, 2008.

[15] A. Lee, I. Streinu, and L. Theran. Graded sparse graphs and matroids. *J. UCS*, 13(11):1671–1679, 2007.

[16] A. Lee-St.John and J. Sidman. Combinatorics and the rigidity of {CAD} systems. *Computer-Aided Design*, 45(2):473 – 482, 2013. Solid and Physical Modeling 2012.

[17] J. Mittmann. Gröbner bases: Computational algebraic geometry and its complexity. 2007.

[18] C. S. J. Nash-Williams. Edge-disjoint spanning trees of finite graphs. *Journal of the London Mathematical Society*, 1(1):445–450, 1961.

[19] B. A. Olshausen and D. J. Field. Sparse coding with an overcomplete basis set: A strategy employed by V1? *Vision research*, 37(23):3311–3325, 1997.

[20] I. Ramirez, P. Sprechmann, and G. Sapiro. Classification and clustering via dictionary learning with structured incoherence and shared features. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 3501–3508. IEEE, 2010.

[21] G. rey Davis. *Adaptive nonlinear approximations*. PhD thesis, Citeseer, 1994.

[22] B. Servatius. Molecular conjecture in 2D, 2006. 16th Fall Workshop on Computational and Combinatorial Geometry.

[23] M. Sitharam. Combinatorial approaches to geometric constraint solving: Problems, progress, and directions. *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, 67:117, 2005.

[24] D. A. Spielman, H. Wang, and J. Wright. Exact recovery of sparsely-used dictionaries. In *Proceedings of the Twenty-Third international joint conference on Artificial Intelligence*, pages 3087–3090. AAAI Press, 2013.

[25] P. Sprechmann and G. Sapiro. Dictionary learning and sparse coding for unsupervised clustering. In *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*, pages 2042–2045. IEEE, 2010.

[26] I. Streinu and L. Theran. Sparse hypergraphs and pebble game algorithms. *European Journal of Combinatorics*, 30(8):1944–1964, 2009.

[27] I. Streinu and L. Theran. Slider-pinning rigidity: a maxwell–laman-type theorem. *Discrete & Computational Geometry*, 44(4):812–837, 2010.

[28] W. T. Tutte. On the problem of decomposing a graph into n connected factors. *Journal of the London Mathematical Society*, 1(1):221–230, 1961.

[29] R. Vidal. Subspace clustering. *Signal Processing Magazine, IEEE*, 28(2):52–68, 2011.

[30] R. Vidal, Y. Ma, and S. Sastry. Generalized principal component analysis (gpca). *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 27(12):1945–1959, 2005.

[31] N. White and W. Whiteley. The algebraic geometry of motions of bar-and-body frameworks. *SIAM Journal on Algebraic Discrete Methods*, 8(1):1–32, 1987.

[32] W. Whiteley. Some matroids from discrete applied geometry. *Contemporary Mathematics*, 197:171–312, 1996.

A Classification of Dictionary Learning problems

A summarization of the input conditions and results of different types of Dictionary Learning problems is given in Table 1.

B Proof of Results

In this section, we provide details and proof for the results in Section 3.

In the following, we denote a minor of a matrix A using the notation $A[R, C]$, where R and C are index sets of the rows and columns contained in the minor, respectively. In addition, $A[R, \cdot\,]$ represents the minor containing all columns and row set R, and $A[\,\cdot\,, C]$ represents the minor containing all rows and column set C.

B.1 Algebraic Representation

In this section, we provide the details in deriving the algebraic system of equations $(H, X)(D) = 0$ (1).

Consider a pin x_k on the subspace spanned by points $v_1^k, v_2^k, \ldots, v_s^k$. Using homogeneous coordinates, we can write this incidence constraint by letting all the $s \times s$ minors of the $(d-1) \times s$ matrix

$$E^k = \left[\begin{array}{cccc} v_1^k - x_k & v_2^k - x_k & \ldots & v_s^k - x_k \end{array}\right]$$

be zero, where $v_i^k = (v_{i,1}^k, v_{i,2}^k, \ldots, v_{i,d-1}^k)$ and $x_k = (x_{k,1}, x_{k,2}, \ldots, x_{k,d-1})$. So each incidence can be written as $\binom{d-1}{s}$ equations:

$$E^k[R(l), \cdot\,] = 0, \qquad 1 \le l \le \binom{d-1}{s} \tag{2}$$

where $R(l)$ enumerates all the s-subsets of rows of E^k. Note that only $d - s$ of these $\binom{d-1}{s}$ equations are independent, as the subspace spanned by $v_1^k, v_2^k, \ldots, v_s^k$ is a s-dimensional subspace in a d-dimensional space, which only has $s(d - s)$ degrees of freedom.

Given the hypergraph $H = H(S_{X,D})$ of the underlying subspace arrangement, the pinned subspace-incidence problem now reduces to solving a system of $m\binom{d-1}{s}$ equations (or, equivalently, $m(d - s)$ independent equations), each of the form (2). The system of equations sets a multivariate function $(H, X)(D)$ to 0:

$$(H, X)(D) = \begin{cases} \quad\ldots & \\ E^k[R(l), \cdot\,] & = \quad 0 \\ \quad\ldots & \end{cases} \tag{1}$$

When viewing X as a fixed parameter, $(H, X)(D)$ is a vector valued function from $\mathbb{R}^{n(d-1)}$ to $\mathbb{R}^{m\binom{d-1}{s}}$ parameterized by X.

Without any pins, the points in D have in total $n(d-1)$ degrees of freedom. In general, putting r pins on an s-dimensional subspace of d-dimensional space gives an $(s-r)$-dimensional subspace of a $(d-r)$-dimensional space, which has $(s-r)((d-r)-(s-r)) = (s-r)(d-s)$ degrees of freedom left. So every pin potentially removes $(d-s)$ degrees of freedom.

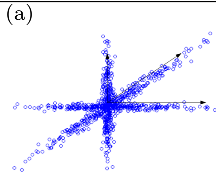
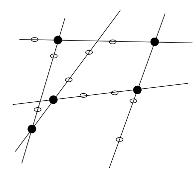
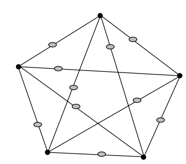
	Traditional Dictionary Learning			Dictionary Learning via subspace arrangement and spanning set	Dictionary Learning for Segmented Data	Fitted Dictionary Learning (this paper)
	Alternating Minimization Approaches	Spielman et. al [24]	Arora et. al [3], Agarwal et. al [1]			
Input and Conditions	D satisfies frame property	X generated from hidden dictionary D and certain distribution of Θ; D is a basis; $s \le d^{1/2}$	X generated from hidden dictionary D and certain distribution of Θ; D is pairwise incoherent; $s \le d^{1/2}$	X with promise that each subspace/dictionary support set is shared by sufficiently many of the data points in X	Partitioned / segmented Data X	Generic data points X (satisfying pure condition) with underlying hypergraph specified
Minimum m guaranteeing existence of a locally unique dictionary of a given size n	Question 2	$O(n \log n)$	$O(n^2 \log^2 n)$	Minimum number of points to guarantee a unique subspace arrangement that will give a spanning set of size n	Problem 6	$\frac{d-s}{d-1}n$ (Theorem 2); Unknown for general position data (Question 1)
Dictionary Learning algorithms	MOD, k-SVD, etc.	Algorithm from [24]	Algorithms from [3, 1]	Subspace Arrangement Learning Algorithms (Problem 4) and Spanning Set Finding (Problem 5)	Problem 6 and Spanning Set Finding (Problem 5)	Straightforward algorithm (Corollary 5)
Minimum m guaranteeing efficient dictionary learning	Unknown	$O(n^2 \log^2 n)$		Unknown	Unknown	Unknown
Illustrative example				(a) 	(b) 	(c) 

Table 1: Classification of Problems

B.2 Linearization as Rigidity Matrix and its Generic Combinatorics

Adapting [4], we now show that rigidity and independence (based on nonlinear polynomials) of pinned pubspace-incidence systems are generically properties of the underlying hypergraph $H(S_{X,D})$, and can furthermore be captured by linear conditions in an generic infinitesimal setting. Recall that a *rigidity matrix* of a framework (H, X, D) is a matrix whose kernel is the infinitesimal motions (flexes) of (H, X, D). A framework is *infinitesimally minimally rigid* if the rows of the rigidity matrix are independent, and the rigidity matrix has full rank.

To define a rigidity matrix for a pinned subspace-incidence framework (H, X, D), we take the Jacobian $J_X(D)$ of the algebraic system $(H, X)(D)$, by taking partial derivatives w.r.t. the coordinates of v_i's. In the Jacobian, each pin x_k has $\binom{d-1}{s}$ corresponding rows, and each vertex v_i has $d-1$ corresponding columns. Each equation $E^k[R(l), \cdot] = 0$ (2) gives the corresponding row in the Jacobian:

$$r^k(l) = [0, \dots, 0, 0, V^k_{1,1}(l), V^k_{1,2}(l), \dots, V^k_{1,d-1}(l), 0, 0, \\ \dots\dots \\ \dots, 0, 0, V^k_{s,1}(l), V^k_{s,2}(l), \dots, V^k_{s,d-1}(l), 0, 0, \dots, 0]$$

Let $D^k = \{v^k_i, 1 \le i \le s\}$ be the vertices of the hyperedge corresponding to x_k. Each vertex v^k_i has the entries $V^k_{i,1}(l), V^k_{i,2}(l), \dots, V^k_{i,d-1}(l)$ in its $d-1$ columns. For $j \in R(l)$, the entry $V^k_{i,j}(l)$ stands for the $(s-1)$-dimensional volume, of the $(s-1)$-simplex formed by the vertices $(D^k \setminus \{v^k_i\}) \cup \{x_k\}$, projected on the coordinates $R(l) \setminus \{j\}$, which is generically non-zero. All the other entries, including the terms $V^k_{i,j}(l)$ where $j \notin R(l)$, and the entries corresponding to vertices not on the hyperedge x_k, are zero. Notice that for every pair of vertices v^k_i and $v^k_{i'}$, the projected volumes on different coordinates all have the same ratio: $\frac{V^k_{i,j_2}(l)}{V^k_{i,j_1}(l)} = \frac{V^k_{i',j_2}(l)}{V^k_{i',j_1}(l)}$ for all $1 \le j_1, j_2 \le d-1, j_1 \in R(l), j_2 \in R(l)$. So we can divide each row $r^k(l)$ by $\sum_{i=1}^{s} V^k_{i,j^*}(l)$, where j^* is the smallest index in $R(l)$, and simplify $r^k(l)$ to

$$[0, \dots, 0, 0, b_1 a_1, b_2 a_1 \dots, b_{d-1} a_1 0, 0, \dots\dots, \\ \dots, 0, 0, b_1 a_{s-1}, b_2 a_{s-1}, \dots, b_{d-1} a_{s-1}, 0, 0, \\ \dots, 0, 0, b_1 \left(1 - \sum_{i=1}^{s-1} a_i\right), b_2 \left(1 - \sum_{i=1}^{s-1} a_i\right), \\ \dots, b_{d-1} \left(1 - \sum_{i=1}^{s-1} a_i\right), 0, 0, \dots, 0] \quad (3)$$

where the values of a_i and b_j are related to l and k, and $b_j = 0$ if $j \notin R(l)$.

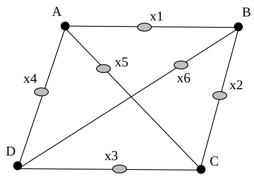


Figure 3: A pinned subspace-incidence framework of 6 pins and 4 vertices, with $d = 4, s = 2$.

Example 3 *Figure 3 shows a pinned subspace-incidence framework with $d = 4, s = 2$. If we denote $\alpha_{1,1} = A_1 - x_{1,1}, \alpha_{1,2} = A_2 - x_{1,2}, \beta_{1,2} = B_2 - x_{1,2}$, etc, the edge AB will have the following three rows in the Jacobian:*

$$\begin{bmatrix} \beta\alpha_{1,2} & -\beta\alpha_{1,1} & 0 & -\alpha_{1,2} & \alpha_{1,1} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \beta\alpha_{1,3} & 0 & -\beta\alpha_{1,1} & -\alpha_{1,3} & 0 & \alpha_{1,1} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \beta\alpha_{1,3} & -\beta\alpha_{1,2} & 0 & -\alpha_{1,3} & \alpha_{1,2} & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

and the corresponding rows in the simplified Jacobian has the following form

$$\begin{bmatrix} b_{1,1}a_1 & b_{1,2}a_1 & 0 & b_{1,1}(1-a_1) & b_{1,2}(1-a_1) & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ b_{2,1}a_2 & 0 & b_{2,2}a_2 & b_{2,1}(1-a_2) & 0 & b_{2,2}(1-a_2) & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & b_{3,1}a_3 & b_{3,2}a_3 & 0 & b_{3,1}(1-a_3) & b_{3,2}(1-a_3) & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

For a pinned subspace-incidence framework (H, X, D), we define the *symmetric rigidity matrix* M to be the simplified Jacobian matrix obtained above, of size $m\binom{d-1}{s}$ by $n(d-1)$, where each row has the form (3). If we choose $d-s$ rows per hyperedge in M, the obtained matrix $\hat{M}$ is a rigidity matrix of size $m(d-s)$ by $n(d-1)$. The framework is infinitesimally rigid if and only if there is an $\hat{M}$ with full rank. Note that the rank of a generic matrix $\hat{M}$ is at least as large as the rank of any specific realization $\hat{M}(H, X, D)$.

Defining generic as non-singular, we prove Lemma 1 showing that for a generic framework (H, X, D), infinitesimal rigidity is equivalent to generic rigidity.

Proof. [Proof Sketch of Lemma 1] First we show that if a framework is regular, infinitesimal rigidity implies rigidity. Consider the polynomial system $(H, X)(D)$ of equations. The Implicit Function Theorem states that there exists a function g, such that $D = g(X)$ on some open interval, if and only if the Jacobian $J_X(D)$ of $(H, X)(D)$ with respect to D has full rank. Therefore, if the framework is infinitesimally rigid, then the solutions to the algebraic system are isolated points (otherwise g could not be explicit). Since the algebraic system contains finitely many components, there are only finitely many such solution and each solution is a 0 dimensional point. This implies that the total number of solutions is finite, which is the definition of rigidity.

To show that generic rigidity implies generic infinitesimal rigidity, we take the contrapositive: if a generic framework is not infinitesimally rigid, we show that there is a finite flex. Let $\hat{M}$ be the $m(d-s)$ by $n(d-1)$ rigidity matrix obtained from the Jacobian $J_X(D)$ which has the maximum rank. If (H, X, D) is not infinitesimally rigid, then the rank r of $\hat{M}$ is less than $n(d-1)$. Let E^* be a set of edges in H such that $|E^*| = r$ and the corresponding rows in the Jacobian $J_X(D)$ are all independent. In $\hat{M}[E^*, \cdot]$, we can find r independent columns. Let D^* be the components of D corresponding to those r independent columns and $D^{*\perp}$ be the remaining components. The r-by-r submatrix $\hat{M}[E^*, D^*]$, made up of the corresponding independent rows and columns, is invertible. Then, by the Implicit Function Theorem, in a neighborhood of D there exists a continuous and differentiable function g such that $D^* = g(D^{*\perp})$. This identifies D', whose components are D^* and the level set of g corresponding to D^*, such that $(H, X)(D') = 0$. The level set defines the finite flexing of the framework. Therefore the system is not rigid. □

B.3 Required Hypergraph Properties

This section formally introduces the $(k, 0)$-sparsity condition that will be used for proving Theorem 2.

Definition 2 *A hypergraph $H = (V, E)$ is $(k, 0)$-sparse if for any $V' \subset V$, the induced subgraph $H' = (V', E')$ satisfies $|E'| \leq k|V'|$. A hypergraph H is $(k, 0)$-tight if H is $(k, 0)$-sparse and $|E| = k|V|$.*

This is a special case of the (k, l)-sparsity condition that was formally studied widely in the geometric constraint solving and combinatorial rigidity literature before it was given a name in [15]. A relevant concept from graph matroids is *map-graph*, defined as follows.

Definition 3 *An* orientation *of a hypergraph is given by identifying as the* tail *of each edge one of its endpoints. The* out-degree *of a vertex is the number of edges which identify it as the tail and connect v to $V - v$. A* map-graph *is a hypergraph that admits an orientation such that the out degree of every vertex is exactly one.*

The following lemma from [26] follows Tutte-Nash Williams [28, 18] to give a useful characterization of $(k, 0)$-tight graphs in terms of maps.

Lemma 6 *A hypergraph H is composed of k edge-disjoint map-graphs if and only if H is $(k, 0)$-tight.*

B.4 Proof of Main Theorem and Corollaries

In this section, we prove the main theorem, Theorem 2, a combinatorial characterization of the existence of finitely many solutions for a pinned subspace incidence framework. The proof adopts an approach by [31], in proving rigidity of k-frames.

First notice that the graph property from Theorem 2 is not directly a $(k, 0)$-tightness condition, so we modify the underlying hypergraph by duplicating each hyperedge into $(d-s)$ copies.

Definition 4 (Expanded multihypergraph) *Given the underlying hypergraph $H = (\Im(D), \Im(S_{X,D}))$ of a Pinned Subspace-Incidence problem, the* expanded multihypergraph *$\hat{H} = (V, \hat{E})$ of H is obtained by letting $V = \Im(D)$, and replacing each hyperedge in $\Im(S_{X,D})$ with $(d-s)$ copies in $\hat{E}$.*

The *rigidity matrix* $\hat{M}$ for a pinned subspace-incidence framework is a $(d-s)|\hat{E}|$ by $(d-1)|V|$ matrix according to the expanded multihypergraph $\hat{H}=(V,\hat{E})$, where each hyperedge $x_k \in \Im(S_{X,D})$ has $(d-s)$ rows, one for each copy. The $(d-s)$ rows are arbitrarily picked from the $\binom{d-1}{s}$ rows of x_k in the symmetric rigidity matrix M.

Theorem 2 can be restated on the expanded multihypergraph:

Theorem 7 *A pinned subspace-incidence framework is generically minimally rigid if and only if the underlying expanded multihypergraph is $(d-1,0)$-tight.*

Since Theorem 7 is equivalent to Theorem 2, we only need to prove Theorem 7 in the following.

We first consider the generic rank of particular matrices defined on a single map-graph.

Lemma 8 *A matrix N defined on a map-graph $H=(V,E)$, such that columns are indexed by the vertices and rows by the edges, where the row for hyperedge $x_k \in E$ has non-zero entries only at the s indices corresponding to $v_i^k \in x_k$, with the following pattern:*

$$[0,\dots,0,a_1^k,0,\dots,a_2^k,0,\dots\dots,0,a_{s-1}^k,0,\dots,1-\sum_{i=1}^{s-1}a_i^k,0,\dots,0] \tag{4}$$

is generically full rank.

Proof. According to the definition of a map-graph, the function $t: E \to V$ assigning a tail vertex to each hyperedge is a one-to-one correspondence. Without loss of generality, assume that for any x_k, the corresponding entry of $t(x_k)$ in N is $1-\sum_i a_i^k$ (notice that we can arbitrarily switch the variable names $a_1^k,\dots,a_{s-1}^k,1-\sum_i a_i^k$). The determinant of the map N is:

$$\det(N)=\pm\prod_k(1-\sum_i a_i^k)+\sum_\sigma \operatorname{sgn}(\sigma)\prod_{i=1}^n N[i,\sigma_i] \tag{5}$$

where σ enumerates all other permutations of $|N|$, excluding that of the first term $\pm\prod_k(1-\sum_i a_i^k)$.

Notice that each term $\prod_{i=1}^n N[i,\sigma_i]$ has at least one a_i^k as a factor. If we use the specialization with $a_i^k=0$ for all i and k, the summation over σ will all be zero, and $\det(N)$ will be $\pm\prod_k(1-\sum_i a_i^k)=\pm1$. So generically, N must have full rank. □

Now we are ready to prove the main theorem by decomposing the expanded multihypergraph as a union of $d-1$ maps, and applying Lemma 8.

Proof. [Proof of Main Theorem] First we show the only if direction. For a generically minimally rigid pinned subspace-incidence framework, the determinant of $\hat{M}$ is not identically zero. Since the number of columns is $n(d-1)$, it is trivial that $n(d-1)$ copied edges in $\hat{M}$, namely $n\frac{d-1}{d-s}$ pins, are necessary. It is also trivial to see that $(d-1,0)$-tightness is necessary, since any subgraph $H'=(V',E')$ of $\hat{H}$ with $|E'|>(d-1)|V'|$ is overdetermined and generically has no solution.

Next we show the if direction, that $n(d-1)$ copied edges arranged generically in a $(d-1,0)$-tight pattern imply infinitesimal rigidity.

We first group the columns according to the coordinates. In other words, we have $d-1$ groups C_j, where all columns for the first coordinate belong to C_1, all columns for the second coordinate belong to C_2, etc. This can be done by applying a Laplace expansion to rewrite the determinant of the rigidity matrix $\hat{M}$ as a sum of products of determinants (brackets) representing each of the coordinates taken separately:

$$\det(\hat{M})=\sum_\sigma\left(\pm\prod_j \det \hat{M}[R_j^\sigma,C_j]\right)$$

where the sum is taken over all partitions σ of the rows into $d-1$ subsets $R_1^\sigma,R_2^\sigma,\dots,R_j^\sigma,\dots,R_{d-1}^\sigma$, each of size $|V|$. Observe that for each $\hat{M}[R_j^\sigma,C_j]$,

$$\det(\hat{M}[R_j^\sigma,C_j])=(b_1^{\sigma j}\dots b_n^{\sigma j})\det(M'[R_j^\sigma,C_j])$$

for some coefficients $(b_1^{\sigma j}\dots b_n^{\sigma j})$, and each row of $\det(M'[R_j^\sigma,C_j])$ is either all zero, or of pattern (4). By Lemma 6, the expanded multihypergraph $\hat{H}$ can be decomposed into $(d-1)$ edge-disjoint maps. Each such decomposition has some corresponding row partitions σ, where each column group C_j corresponds to a map N_j, and R_j^σ contains rows corresponding to the edges in that map. Observe that $\hat{M}[R_j^\sigma,C_j]$ contains an all-zero row r, if and only if the row r has the jth coordinate entry being zero in $\hat{M}$. Recall for each hyperedge x_k, we are free to pick any $d-s$ rows to include in $\hat{M}$ from the $\binom{d-1}{s}$ rows in the symmetric rigidity matrix M. We claim that

Claim 1 *Given a map decomposition, we can always pick the rows of the rigidity matrix $\hat{M}$, such that there is a corresponding row partition σ^*, where none of the minors $\hat{M}[R_j^{\sigma^*},C_j]$ contains an all-zero row.*

Given a map decomposition, for any map N_j, there are $\binom{d-2}{s-1}$ among these $\binom{d-1}{s}$ rows with the jth coordinate being non-zero. Also, it is not hard to show that for all $2\le s\le d-1$, $\binom{d-2}{s-1}\ge d-s$. So for any N_j containing k_j copies of a particular hyperedge, since all the other maps can pick at most $(d-s)-k_j$ rows from its $\binom{d-2}{s-1}$ choices, it still has $\binom{d-2}{s-1}-((d-s)-k_j)\ge k_j$ choices. Therefore, given a map decomposition, we can always pick the rows in the rigidity matrix $\hat{M}$, such that there is a partition of each hyperedge's rows, where each map N_j get its required rows with non-zeros at coordinate j. This concludes the proof of the claim.

So by Lemma 8, the determinate of each such minor $\hat{M}[R_j^{\sigma^*},C_j]$ is generically non-zero. We conclude that

$$\det(\hat{M})=\sum_\sigma\left(\pm\prod_j\left((b_1^{\sigma j}\dots b_n^{\sigma j})\det\left(M'[R_j^\sigma,C_j]\right)\right)\right)$$

Observe that each term of the sum has a unique multi-linear coefficient $(b_1^{\sigma j}\dots b_n^{\sigma j})$ that generically do not cancel with any of the others since $\det(M'[R_j^\sigma,C_j])$ are independent of the b's. This implies that $\hat{M}$ is generically full rank, thus completes the proof. Moreover, substituting the values of $\det(M'[R_j^\sigma,C_j])$ from Lemma 8 gives the pure condition for genericity. □

Example 4 *Consider the pinned subspace-incidence framework in Example 3. The expanded multihypergraph satisfies* $(3,0)$*-tightness condition. The rigidity matrix* $\hat{M}$ *has the following form:*

$$\begin{bmatrix}
b_{1,1}a_1 & b_{1,2}a_1 & 0 & b_{1,1}\bar{a_1} & b_{1,2}\bar{a_1} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
b_{2,1}a_2 & 0 & b_{2,2}a_2 & b_{2,1}\bar{a_2} & 0 & b_{2,2}\bar{a_2} & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & b_{3,1}a_3 & b_{3,2}a_3 & 0 & b_{3,1}\bar{a_3} & b_{3,2}\bar{a_3} & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & b_{4,1}a_4 & 0 & b_{4,2}a_4 & b_{4,1}\bar{a_4} & 0 & b_{4,2}\bar{a_4} & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & b_{5,1}a_5 & b_{5,2}a_5 & 0 & b_{5,1}\bar{a_5} & b_{5,2}\bar{a_5} & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & b_{6,1}a_6 & 0 & b_{6,2}a_6 & b_{6,1}\bar{a_6} & 0 & b_{6,2}\bar{a_6} \\
b_{7,1}a_7 & b_{7,2}a_7 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & b_{7,1}\bar{a_7} & b_{7,2}\bar{a_7} & 0 \\
b_{8,1}a_8 & 0 & b_{8,2}a_8 & 0 & 0 & 0 & 0 & 0 & 0 & b_{8,1}\bar{a_8} & 0 & b_{8,2}\bar{a_8} \\
b_{9,1}a_9 & b_{9,2}a_9 & 0 & 0 & 0 & 0 & b_{9,1}\bar{a_9} & b_{9,2}\bar{a_9} & 0 & 0 & 0 & 0 \\
b_{10,1}a_{10} & 0 & b_{10,2}a_{10} & 0 & 0 & 0 & b_{10,1}\bar{a_{10}} & 0 & b_{10,2}\bar{a_{10}} & 0 & 0 & 0 \\
0 & 0 & 0 & b_{11,1}a_{11} & b_{11,2}a_{11} & 0 & 0 & 0 & 0 & b_{11,1}\bar{a_{11}} & b_{11,2}\bar{a_{11}} & 0 \\
0 & 0 & 0 & b_{12,1}a_{12} & 0 & b_{12,2}a_{12} & 0 & 0 & 0 & b_{12,1}\bar{a_{12}} & 0 & b_{12,2}\bar{a_{12}}
\end{bmatrix}$$

where $\bar{a_i}$ *stands for* $1 - a_i$.

After grouping the coordinates, it becomes

$$\left[\begin{array}{cccc|cccc|cccc}
b_{1,1}a_1 & b_{1,1}\bar{a_1} & 0 & 0 & b_{1,2}a_1 & b_{1,2}\bar{a_1} & 0 & 0 & 0 & 0 & 0 & 0 \\
b_{2,1}a_2 & b_{2,1}\bar{a_2} & 0 & 0 & 0 & 0 & 0 & 0 & b_{2,2}a_2 & b_{2,2}\bar{a_2} & 0 & 0 \\
0 & b_{3,1}a_3 & b_{3,1}\bar{a_3} & 0 & 0 & b_{3,2}a_3 & b_{3,2}\bar{a_3} & 0 & 0 & 0 & 0 & 0 \\
0 & b_{4,1}a_4 & b_{4,1}\bar{a_4} & 0 & 0 & 0 & 0 & 0 & 0 & b_{4,2}a_4 & b_{4,2}\bar{a_4} & 0 \\
0 & 0 & b_{5,1}a_5 & b_{5,1}\bar{a_5} & 0 & 0 & b_{5,2}a_5 & b_{5,2}\bar{a_5} & 0 & 0 & 0 & 0 \\
0 & 0 & b_{6,1}a_6 & b_{6,1}\bar{a_6} & 0 & 0 & 0 & 0 & 0 & 0 & b_{6,2}a_6 & b_{6,2}\bar{a_6} \\
b_{7,1}a_7 & 0 & 0 & b_{7,1}\bar{a_7} & b_{7,2}a_7 & 0 & 0 & b_{7,2}\bar{a_7} & 0 & 0 & 0 & 0 \\
b_{8,1}a_8 & 0 & 0 & b_{8,1}\bar{a_8} & 0 & 0 & 0 & 0 & b_{8,2}a_8 & 0 & 0 & b_{8,2}\bar{a_8} \\
b_{9,1}a_9 & 0 & b_{9,1}\bar{a_9} & 0 & b_{9,2}a_9 & 0 & b_{9,2}\bar{a_9} & 0 & 0 & 0 & 0 & 0 \\
b_{10,1}a_{10} & 0 & b_{10,1}\bar{a_{10}} & 0 & 0 & 0 & 0 & 0 & b_{10,2}a_{10} & 0 & b_{10,2}\bar{a_{10}} & 0 \\
0 & b_{11,1}a_{11} & 0 & b_{11,1}\bar{a_{11}} & 0 & b_{11,2}a_{11} & 0 & b_{11,2}\bar{a_{11}} & 0 & 0 & 0 & 0 \\
0 & b_{12,1}a_{12} & 0 & b_{12,1}\bar{a_{12}} & 0 & 0 & 0 & 0 & 0 & b_{12,2}a_{12} & 0 & b_{12,2}\bar{a_{12}}
\end{array}\right]$$

where the red rows inside each column group corresponding to a map decomposition of the expanded multihypergraph.

The combinatorial characterization in Theorem 2 leads to the proof of Corollary 3, which gives a lower bound of dictionary size for generic data points in the general Dictionary Learning problem.

Proof. [Proof of Corollary 3 (Lower bound of dictionary size for generic data points)] We first prove one direction, that there is generically no dictionary of size $|D| = n$ if $(d-s)m > (d-1)n$. For any hypothetical s-subspace arrangement $S_{X,D}$, the expanded multihypergraph $\hat{H}(S_{X,D})$ - with the given bound for $|D|$ - cannot be $(d-1,0)$-sparse. Hence generically, under the pure conditions of Theorem 2, the rigidity matrix - of the s-subspace framework $H(S_{X,D})$ - with indeterminates representing the coordinate positions of the points in D - has dependent rows. In which case, the original algebraic system $(H,X)(D)$(whose Jacobian is the rigidity matrix) will not have a (complex or real) solution for D, with X plugged in.

The converse is implied from our theorem since we are guaranteed both generic independence (the existence of a solution) and generic rigidity (at most finitely many solutions). □

By characterizing the term "generically" in Corollary 3, we prove Corollary 4 which gives a lower bound of dictionary size for data points picked uniformly at random from the sphere S^{d-1}.

Proof. [Proof of Corollary 4 (Lower bound of dictionary size for highly general data points)] To quantify the term "generically" in Corollary 3, we note that the pure-conditions fail only on a measure-zero subset of the space of frameworks $S_{X,D}$. Since the number of possible multi-hypergraphs representing the s-subspace arrangements is finite for a given set of pins, it follows that except for a measure-zero subset of the space of pin-sets X, there is no (real or complex) solution to the algebraic system $(H,X)(D) = 0$ when $(d-s)m > (d-1)n$. Thus when X is picked uniformly at random from the sphere S^{d-1}, if $|D|$ is less than $\big((d-1)/(d-s)\big)|X|$, with probability 1, there is no solution. □

B.5 Straightforward Dictionary Learning Algorithm for highly general data points

In this section, we present the algorithm in Corollary 5, which constructs a dictionary of size $n = \left(\frac{d-s}{d-1}\right)m$, given m data points picked uniformly at random from the sphere S^{d-1}. The algorithm has two major parts: (1) constructing the underlying hypergraph $H(S_{X,D})$, and (2) constructing the s-subspace arrangement $S_{X,D}$ and the dictionary D.

(1) Algorithm for constructing the underlying hypergraph $H(S_{X,D})$ for a hypothetical s-subspace arrangement $S_{X,D}$:

1. We start by constructing a minimal minimally rigid hypergraph $H_0 = (V_0, E_0)$, using the *pebble game algorithm* introduced below. Here $|V_0| = k(d-s)$, $|E_0| = k(d-1)$, where k is the smallest positive integer such that $\binom{k(d-s)}{s}(s-1) \geq k(d-1)$, so it is possible to construct E_0 such that no more than $s-1$ hyperedges in E_0 containing the same set of vertices in V_0. The values $|V_0|$ and $|E_0|$ are constants when we think of d and s as constants.
2. We use the pebble game algorithm to append a set V_1 of $d-s$ vertices and a set E_1 of $d-1$ hyperedges to H_0, such that each hyperedge in E_1 contains at least one vertex from V_1, and the obtained graph H_1 is still minimally rigid. The subgraph B_1 induced by E_1 has vertex set $V_{B_1} = V_1 \bigcup V_B$, where $V_B \subset V_0$. We call the vertex set V_B the *base vertices* of the construction.
3. Each of the following construction step i appends a set V_i of $d-s$ vertices and a set E_i of $d-1$ hyperedges such that the subgraph B_i induced by E_i has vertex set $V_i \bigcup V_B$, and B_i is isomorphic to B_1. In other words, at each step, we directly append a basic structure the same as (V_1, E_1) to the base vertices V_B. It is not hard to verify that the obtained graph is still minimally rigid.

The *pebble game algorithm* by [26] works on a fixed finite set V of vertices and constructs a (k,l)-sparse hypergraph. Conversely, any (k,l)-sparse hypergraph on vertex set V can be constructed by this algorithm. This algorithm initializes by putting k pebbles on each vertex in V. There are two types of moves:

- *Add-edge*: adds an hyperedge e (e must contain at least $l+1$ pebbles), removes a pebble from a vertex v in e, and assign v as the tail of e;
- *Pebble-shift*: for a vertex $v_1 \in e$ which contains at least one pebble, let v_2 be the tail of e, move one pebble from v_1 to v_2, and change the tail of e to v_1.

At the end of the algorithm, if there are exactly l pebbles in the hypergraph, then the hypergraph is (k,l)-tight.

Our algorithm runs a slightly modified pebble game algorithm to find a $(d-1,0)$-tight expanded multihypergraph. We require that each add-edge move adding $(d-s)$ copies of a hyperedge e, so a total of $d-s$ pebbles are removed from vertices in e. Additionally, the multiplicity of a hyperedge (not counting the expanded copies) cannot exceed $(s-1)$. For constructing the basic structure of Stage 2, the algorithm initializes by putting $d-1$ pebbles on each vertex in V_1. In addition, an add-edge move can only add a hyperedge that contains at least one vertex in V_1, and a pebble-shift move can only shift a pebble inside V_1.

The pebble-game algorithm takes $O\left(s^2|V_0|\binom{|V_0|}{s}\right)$ time in Step 1 and $O\left(s^2\left(|V_0|+(d-s)\right)\binom{|V_0|+(d-s)}{s}\right)$ time in Step 2. Since the complete underlying hypergraph $H(S_{X,D})$ has $m=|X|$ edges, Step 3 will be iterated $O(m/(d-1))$ times, and each iteration takes constant time. Therefore the overall time complexity for constructing $H(S_{X,D})$ is $O\left(s^2\left(|V_0|+(d-s)\right)\binom{|V_0|+(d-s)}{s}+(m/(d-1))\right)$, which is $O(m)$ when d and s are regarded as constants.

(2) Algorithm for constructing the s-subspace arrangement $S_{X,D}$ and the dictionary D:

The construction of the s-subspace arrangement $S_{X,D}$ naturally follows from the construction of the underlying hypergraph. For the initial hypergraph H_0, we get a pinned subspace-incidence system $(H_0,X_0)(D_0)$ by arbitrarily choose $|X_0|=|E_0|$ pins from X. Similarly, for Step 2 and each iteration of Step 3, we form a pinned subspace-incidence system $(B_i,X_i)(D_i)$ by arbitrarily choosing $|X_i|=d-1$ pins from X.

Given X_0, we know that the rigidity matrix - of the s-subspace framework $H_0(S_{X_0,D_0})$ - with indeterminates representing the coordinate positions of the points in D_0 - generically has full rank (rows are maximally independent), under the pure conditions of Theorem 2; in which case, the original algebraic subsystem $(H_0,X_0)(D_0)$ (whose Jacobian is the rigidity matrix), with X_0 plugged in, is guaranteed to have a (possibly complex) solution and only finitely many solutions for D_0. Since the pure conditions fail only on a measure-zero subset of the space of pin-sets X_0, where each pin is in S^{d-1}, it follows that if the pins in X_0 are picked uniformly at random from S^{d-1} we know such a solution exists for D_0 (and S_{X_0,D_0}) and can be found by solving the algebraic system $H_0(S_{X_0,D_0})$.

Once we have solved $(H_0,X_0)(D_0)$, for each following construction step i, B_i is also rigid since coordintate positions of the vertices in V_B have been fixed (this actually follows from the generalization of our main result to non-uniform hypergraphs, by thinking of each vertex in V_B as being spanned by a single pin). So similarly, we know a solution exists for D_i (and S_{X_i,D_i}) and can be found by solving the algebraic system $B_i(S_{X_i,D_i})$, which is of constant size $O(d)$. Although there can be more than one choice of solution for each step, since every construction step is based on base vertices V_B, the solution of one step will not affect any other steps, so generically any choice will result in a successful solution for the entire construction sequence, and we obtain D by taking the union of all D_i's.

When we regard d and s as constants, the time complexity for Stage (2) is the constant time for solving the size $O(|V_0|)$ algebraic system $(H_0,X_0)(D_0)$, plus $O(m/(d-1))$ timed by the constant time for solving the size $O(d)$ system $(B_i,X_i)(D_i)$, that is $O(m)$. Therefore the overall time complexity of the dictionary learning algorithm is $O(m)$.