Electronic Colloquium on Computational Complexity, Revision 2 of Report No. 128 (2015)

Local Reconstruction of Low-Rank Matrices and Subspaces

Roee David* Elazar Goldenberg† Robert Krauthgamer‡

October 9, 2016

Abstract

We study the problem of *reconstructing a low-rank matrix*, where the input is an $n \times m$ matrix M over a field $\mathbb{F}$ and the goal is to reconstruct a (near-optimal) matrix M' that is low-rank and close to M under some distance function Δ. Furthermore, the reconstruction must be local, i.e., provides access to any desired entry of M' by reading only a few entries of the input M (ideally, independent of the matrix dimensions n and m). Our formulation of this problem is inspired by the local reconstruction framework of Saks and Seshadhri (SICOMP, 2010).

Our main result is a local reconstruction algorithm for the case where Δ is the normalized Hamming distance (between matrices). Given M that is ϵ-close to a matrix of rank $d < 1/\epsilon$ (together with d and ϵ), this algorithm computes with high probability a rank-d matrix M' that is $O(\sqrt{d\epsilon})$-close to M. This is a local algorithm that proceeds in two phases. The *preprocessing phase* reads only $\tilde{O}(\sqrt{d/\epsilon^3})$ random entries of M, and stores a small data structure. The *query phase* deterministically outputs a desired entry $M'_{i,j}$ by reading only the data structure and $2d$ additional entries of M.

We also consider local reconstruction in an easier setting, where the algorithm can read an entire matrix column in a single operation. When Δ is the normalized Hamming distance between vectors, we derive an algorithm that runs in polynomial time by applying our main result for matrix reconstruction. For comparison, when Δ is the truncated Euclidean distance and $\mathbb{F} = \mathbb{R}$, we analyze sampling algorithms by using statistical learning tools.

A preliminary version of this paper appears appears in ECCC, see: http://eccc.hpi-web.de/report/2015/128/

Key words: Sublinear-time algorithms, local reconstruction, low-rank matrix reconstruction, matrix rigidity, subspace approximation.

1 Introduction

Suppose our input is a large data matrix M guaranteed to be decomposable as the sum

$$M = \tilde{M} + S,$$

*Weizmann Institute of Science, Israel. Email: roee.david@weizmann.ac.il.

†Charles University in Prague, Czech Republic. The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013) / ERC Grant Agreement n. 616787. Email: elazargold@gmail.com.

‡Weizmann Institute of Science, Israel. Work supported in part by a US-Israel BSF grant #2010418 and an Israel Science Foundation grant #897/13. Email: robert.krauthgamer@weizmann.ac.il.

ISSN 1433-8092

where $\tilde{M}$ is a low-rank matrix and S is a sparse matrix. A common question arising in many application domains is whether $\tilde{M}$ can be recovered efficiently. This question actually has many variants — one has to specify the field used for operations (typically $\mathbb{R}$), the way S is generated to model noise (often in some random manner), and the output's quality measure (e.g., exact reconstruction of $\tilde{M}$). For example, major recent successes [CLMW11, CSPW11] in designing robust versions of Principal Component Analysis considered the above question for real matrices, assuming the singular vectors of $\tilde{M}$ are "well-spread", and the nonzeros of S are located at random entries (but can have arbitrary values), and their algorithms reconstruct $\tilde{M}$ exactly with high probability.

We study this problem when the matrices are over a field $\mathbb{F}$ and the noise S is adversarial. Under such weak conditions, the matrix $\tilde{M}$ need not be unique, and thus our goal is to reconstruct some low-rank matrix M' that is close in Hamming distance to the input M, called henceforth *Matrix Reconstruction*. Furthermore, our algorithm reconstructs M' in a *local* manner, i.e., every desired entry of M' can be computed fast, meaning with runtime that is independent of the matrix size, which in particular bounds the number of entries read from M (per entry of M'). See more in Section 1.1.

The above problem can be cast more generally as local reconstruction, a framework that was introduced by Saks and Seshadhri [SS10], as follows. Suppose the input is a large dataset W that ideally should satisfy some property $\mathcal{P}$, but due to corruptions, W is only guaranteed to be close to $\mathcal{P}$. The goal is to find W' that is close to W and does have property $\mathcal{P}$. Another instantiation of this framework is *Subspace Reconstruction*, where the input is a set of vectors $W \subset \mathbb{F}^n$ that are close (according to some distance) to some low-dimensional subspace, and the goal is to reconstruct such a subspace in a local manner.

As an application of our main result for matrix reconstruction, we obtain an algorithm for subspace reconstruction under Hamming distance between vectors (Section 1.2). The two problems are similar except that in subspace reconstruction, the basic object is a vector, which can be viewed as a whole column in a matrix. To further investigate the statistical aspects, we then study the same subspace reconstruction problem but under Euclidean distance between real vectors, i.e., $\mathbb{F} = \mathbb{R}$ (Section 1.4). This problem is known in the literature as *Subspace Approximation*, however all previous work studied it in the offline model, in contrast to our interest in local algorithms. We design local reconstruction algorithms by employing powerful statistical-learning machinery like Rademacher complexity. These tools cannot be applied to Matrix Reconstruction, our main problem, because they cannot handle the two-dimensional structure of matrices, in addition to being ill-suited for finite fields and Hamming distance.

1.1 Matrix Reconstruction

For two matrices A and B of the same dimensions, let $\Delta(A, B)$ denote the normalized Hamming distance between them. We say the matrices are ϵ-*close* if $\Delta(A, B) \leq \epsilon$, and otherwise we say they are ϵ-*far*. For what follows, we fix a field $\mathbb{F}$. The choice of Hamming distance between matrices is very natural when $\mathbb{F}$ is a finite field, but may also be relevant over the reals, e.g., to model faults or corruptions.

In the *Low-Rank Matrix Reconstruction* problem, the input is a matrix $M \in \mathbb{F}^{n \times m}$ together with $d \in \mathbb{N}$ and $\epsilon > 0$, with the guarantee that M is ϵ-close to some matrix $\tilde{M}$ of rank at most d. The goal is to find (reconstruct) a matrix M' of rank at most d that is ϵ'-close to M, for ϵ' as close as possible to ϵ. As the above setup does not determine $\tilde{M}$ uniquely (e.g., modifying a row of $\tilde{M}$

changes its distance by $1/n$), we define reconstruction as finding some matrix M' with guarantees similar to $\tilde{M}$.

Suppose we want to provide *direct access* to any desired entry of M', without reading all of M (analogously to local decoding, see e.g [GL89]). Specifically, how many queries to M are needed to reconstruct a low-rank M' only in a single location? Can it depend only on d and ϵ (but not on the matrix dimensions, n and m)? Formally, a reconstruction is called *local* if it consists of the following two phases. In the *preprocessing phase*, the algorithm accesses a few entries of M and creates a small data structure S. In the *query phase*, the algorithm is given $(i,j) \in [n] \times [m]$ and has to output $M'_{i,j}$ by using the data structure S and accessing a few more entries of M. We insist that the query phase is deterministic and does not modify S, hence invoking it multiple times would produce many entries of the same M'. In particular, the low-rank matrix M' is determined implicitly by the matrix M and the randomness of the preprocessing phase.[1] We assume that a single machine word can store either an index for a row/column or an element from the field $\mathbb{F}$. We suppress polylogarithmic factors by using the notation $\tilde{O}(f)$ as a shorthand for $f \cdot (\log f)^{O(1)}$.

Definition 1.1 (Local Reconstruction)**.** *An algorithm whose input is $M \in \mathbb{F}^{n \times m}$ together with $d \in \mathbb{N}$ and $\epsilon > 0$ is called an* ϵ'-reconstructor *(where ϵ' is a function of ϵ and d), if:*

1. *its output is a matrix $M' \in \mathbb{F}^{n \times m}$ of rank at most d; and*
2. *if M is ϵ-close to some $\tilde{M}$ of rank at most d, then $\Pr[\Delta(M, M') \leq \epsilon'] \geq 2/3$.*

An ϵ'-reconstructor *algorithm is called* local *with parameters $q_{\text{prep}}, t_{\text{prep}}, s, q_{\text{query}}, t_{\text{query}}$ (which are functions of ϵ and d) if it consists of two phases as above, where*

3. *the preprocessing phase runs in time t_{prep}, accesses at most q_{prep} entries of M, and creates a data structure S of size s words; and*
4. *the query phase runs in time t_{query} and accesses at most q_{query} entries of M.*

Theorem 1.2 (Main Theorem)**.** *For every field $\mathbb{F}$ there is a local $O(\sqrt{d\epsilon})$-reconstructor with parameters $q_{\text{prep}} = \tilde{O}\left(\sqrt{\frac{d}{\epsilon^3}}\right)$, $t_{\text{prep}} = \tilde{O}\left(\frac{1}{\epsilon}\left(\frac{(8e)^2 \ln^2 d}{d\epsilon}\right)^d\right)$, $s = O(d^2)$, $q_{\text{query}} = 2d$ and $t_{\text{query}} = O(d^2)$.*

This theorem, which we prove in Section 3, is largely incomparable to prior work because our setup is different. Our algorithm is *local* and thus its runtime for a single entry is independent of the matrix-size parameters n and m, whereas all previous algorithms we are aware of (mostly analysis of robust principle component, such as [CLMW11, CSPW11, XCM13]), require at least time nm to read the entire input matrix.

Let us make a few remarks about our algorithm's performance. First, its data structure S is very small. The main reason is that it implicitly employs a factorization of the output matrix M' (of rank at most d) as $M' = ABC$, where $A \in \mathbb{F}^{n \times d}$, $B \in \mathbb{F}^{d \times d}$, and $C \in \mathbb{F}^{d \times m}$. Second, the "closeness" between the output M' and the input M is $\epsilon' = O(\sqrt{d\epsilon})$. We suspect the factor d is necessary here, because our query phase uses just $O(d)$ queries, in which case the probability that at least one queried entry is "noisy" (i.e., where M and $\tilde{M}$ disagree) is about $d\epsilon$. Perhaps the square-root operator could be avoided, and we in fact obtain such an improved bound $\epsilon' = O(d\epsilon)$ when M is $\Omega(1)$-far from every rank $d-1$ matrix. This last condition holds with high probability

[1] This is exactly the viewpoint taken by [SS10], in which the preprocessing only determines "seed" coins, and the query phase does the rest of the work.

when $\tilde{M}$ is chosen uniformly at random from all matrices of rank at most d and ϵ is sufficiently small, see Section 3.2. Third, achieving optimal reconstruction, i.e., $\epsilon' = \epsilon$, is NP-hard even in the case $d = 2$. More precisely, computing for an input matrix M, a matrix M' of rank at most d that minimizes $\Delta(M', M)$, is NP-hard even when $d = 2$ and all computations are over $\mathbb{F}_2$. This was proved by Uri Feige (private communication), and we thank him for the permission to provide here a sketch of his proof. (The case $d = 1$ appears to be open, see [Des07, Section 4.2].) The proof is by reduction from the problem Hypercube 2-Segmentation (H2S), in which the input is vectors $v_1, \ldots, v_n \in \{0,1\}^d$ and the goal is to find two "medians" $c_1, c_2 \in \{0,1\}^d$ that minimize $\frac{1}{n}\sum_{i=1}^{n} \min\{\Delta(v_i, c_1), \Delta(v_i, c_2)\}$; this problem is known to be NP-hard, see [Fei14]. Given such an H2S instance, construct the matrix $M \in \{0,1\}^{n \times 7d}$ by letting each row i be the respective v_i followed by $6d$ ones. It can be verified that the optimal value of the H2S instance $\{v_i\}_i$ is equal to the optimal value of the matrix reconstruction instance M.[2]

Potential Applications. We provide two directions for potential applications of our main theorem, in addition to the application to Subspace Reconstruction (in Section 1.2), which we already mentioned.

- Storage Reduction: Low-rank matrices clearly admit a succinct representation. We can partially extend this to matrices that are close to having a low rank, by applying our main theorem to compute a succinct approximation to the matrix in polynomial time (when d and $1/\epsilon$ are small enough).
- Sublinear-time graph algorithms: If the adjacency matrix of a graph G is close to having rank d over $\mathbb{F}_2$, then our main theorem can be used to quickly approximate the adjacency matrix by a matrix of rank d, which can actually be guaranteed to be symmetric. The latter matrix represents a vertex-weighted graph U with 2^d vertices, because such a matrix can have at most 2^d distinct rows and columns, and we can now apply on U algorithms that estimate various graph quantities, which in turn approximate these quantities in the graph G, much faster than applying it on G. This is a well-known approach that was used, for instance, to approximate the maximum cut in a dense graph [FK99], although their results apply to every graph (we assume the graph is close to low rank, which may lead to faster algorithms) and uses the cut-norm distance between matrices (instead of Hamming).

Related Work. The Matrix Reconstruction and Subspace Reconstruction problems follow the local reconstruction framework of Saks and Seshadhri [SS10], who studied monotonicity of functions $f : [n]^d \to \mathbb{R}$, and were inspired by (non-local) reconstruction of Ailon, Chazelle, Comandur and Liu [ACCL08]. A similar model of "repair" was proposed independently by Austin and Tao [AT10] in the context of hypergraphs.

[2] Given an optimal solution c_1, c_2 for the H2S instance, construct M' by taking M and replacing each v_i (in row i) with c_j that attains $\min\{\Delta(v_i, c_1), \Delta(v_i, c_2)\}$; then $\mathrm{rank}(M') \leq 2$. In the other direction, given optimal M', its rows must come from a set of the form $\{0, v, w, v+w\}$. This M' must contain a row i^* in which the "padding part" has at least $5d$ ones, as otherwise $\Delta(M, M') > (6d - 5d)n/(7dn) = 1/7$ and M' is suboptimal. It follows that in every row of M', the "padding part" has at least $5d$ ones, as otherwise this row can be replaced with row i^*. Consequently, the vector 0 does not appear in M', and moreover, at most two vectors from $\{v, w, v+w\}$ can appear in M', as otherwise some coordinate would show $1 + 1 = 1 \pmod 2$. Thus, M' has at most two distinct rows, and moreover "padding part" must be all ones, giving rise to two medians c_1, c_2.

The informal concept of approximating a matrix by a low-rank matrix is ubiquitous in the literature, and includes for example truncated Singular Value Decomposition (SVD), rigidity [Val77], cut-matrices [FK99], and ϵ-rank [Alo09], see also [ALSV13] and references therein. However, these often *do not* require the difference to be sparse.

1.2 Subspace Reconstruction under Hamming Distance

We now consider a different model, which actually demonstrates an application of Theorem 1.2. For two vectors $w, w' \in \mathbb{F}^n$, let $\Delta(w, w')$ be their normalized Hamming distance, and define distance from w to a set $A \subset \mathbb{F}^n$ as $\Delta(w, A) := \min_{a \in A} \Delta(w, a)$ (For simplicity, we assume the minimum is well-defined, as everything can be carried out also with an infimum).

In the *Low-Dimensional Subspace Reconstruction* problem, the input is a set of vectors $W \subset \mathbb{F}^n$ together with $d \in \mathbb{N}$ and $\epsilon > 0$, with the guarantee there is a d-dimensional subspace $\tilde{V}$ for which

$$\mathop{\mathbb{E}}_{w \in W}[\Delta(w, \tilde{V})] \leq \epsilon$$

where the expectation is over the uniform distribution of W (in other words, W is generated as an adversarial perturbation of some $\tilde{W} \subset \tilde{V}$). The goal is to find a subspace V' of dimension d such that $\mathbb{E}_{w \in W}[\Delta(w, V')] \leq \epsilon'$ for ϵ' that is close to ϵ.

This Subspace Reconstruction problem is equivalent to Matrix Reconstruction (the problem defined in Section 1.1), by simply writing the vectors of W as columns of a matrix M. However, from the perspective of *local* algorithms, it is a different problem because the basic object (to read, or to output) is now a whole vector rather than a scalar. Another, more subtle, difference is that now the algorithm inherently involves manipulations of n-dimensional vectors, and the runtime should preferably be polynomial in n. A more formal discussion follows.

As before, a *local* algorithm has two phases. In the *preprocessing phase*, it accesses a few vectors from the input W and creates a small data structure S that determines the output V' (represented by some basis of V'). In the *query phase*, the algorithm is given one vector $w \in W$ and has to output a corresponding vector $w' \in V'$ by using S (without further access to W). See Section 4 for a formal definition.

Applying Theorem 1.2 immediately yields a subspace reconstructor with $\epsilon' = O(\sqrt{d\epsilon})$ that is local, namely, its preprocessing phase queries $q_{\text{prep}} = \tilde{O}\left(\sqrt{d/\epsilon^3}\right)$ vectors from W, and the size of its data structure is $|S| = O(d^2)$ words. We point out that this reconstructor is explicit, meaning that the query phase's runtime is polynomial in n, which in general is a rather non-trivial problem (for example, when V' and w are given as input, it is NP-hard to find a vector of V' with minimum Hamming distance to w). See Section 4.1 for the precise statement and full details.

1.3 Technical Contribution

To illustrate our technical contribution, consider first a particularly easy setting of the Subspace Reconstruction problem, where $(1-\epsilon)$-fraction of the input $W \subset \mathbb{F}^n$ is *contained* in some d-dimensional subspace $\tilde{V}$. Recall that the goal is to find a d'-dimensional subspace V' containing at least $(1-\epsilon')$-fraction of W, while reading only a few vectors of W. Standard arguments using statistical learning (namely, the VC-dimension of all d-dimensional subspaces in $\mathbb{F}^n$ is d, see Section 4.4) imply that $O(d/\epsilon^2)$ random samples from W suffice (with high probability) to compute a subspace V' of dimension $d' \leq d$ achieving $\epsilon' = O(\epsilon)$. But a straightforward analysis shows that even $O(d/\epsilon)$ queries suffice to achieve $d' \leq d$ and $\epsilon' = O(\epsilon)$.

In the general setting, $W \subset \mathbb{F}^n$ is only *close* to a d-dimensional subspace $\tilde{V}$, i.e., $\mathbb{E}_{w \in W} \Delta(w, \tilde{V}) \leq \epsilon$. Notice that when Δ is the Hamming distance, this problem is no longer trivial (in particular, standard statistical learning tools seem inadequate), and here our study branches into two separate directions. One studies (Subspace Reconstruction) under non-Hamming distances, where statistical learning tools such as VC-dimension and Rademacher complexity are effective, see Section 1.4 for more details. A second direction develops a specialized machinery to solve Matrix Reconstruction under Hamming distance, as we explain next.

Our main result (for Matrix Reconstruction) relies on characterizing matrices of rank d in a *robust* way, i.e., it extends to matrices that are ϵ-*close* to having rank d, yet it is based on small random submatrices and is thus local. Specifically, we show that random sets of rows and columns, $R \subset [n], C \subset [m]$ of size $\text{poly}(d/\epsilon)$, typically contain subsets $\tilde{R} \subset R$, $\tilde{C} \subset C$, of size $|\tilde{R}| = |\tilde{C}| \leq d$, for which the matrix

$$M' \stackrel{\text{def}}{=} M_{[n],\tilde{C}} (M_{\tilde{R},\tilde{C}})^{-1} M_{\tilde{R},[m]} \tag{1}$$

satisfies $\Delta(M, M') \leq O(\sqrt{d\epsilon})$. Observe that $\tilde{R}, \tilde{C}$ implicitly determine a low-rank matrix M' that approximates M.

The implementation of the local reconstruction algorithm now follows easily. The preprocessing phase chooses random R and C, enumerates over all possible $\tilde{R} \subset R, \tilde{C} \subset C$, to find a choice that satisfies (1) (by testing it in a few random entries), and then store in the data structure S these $\tilde{R}, \tilde{C}$ and the corresponding submatrix $(M_{\tilde{R},\tilde{C}})^{-1}$. The query phase, when asked for entry (i, j), returns $M_{i,\tilde{C}} (M_{\tilde{R},\tilde{C}})^{-1} M_{\tilde{R},j}$, while reading from M only $|\tilde{R}| + |\tilde{C}|$ entries.

1.4 Subspace Reconstruction for Real Spaces

Subspace Reconstruction is appealing also under distance functions Δ other than Hamming distance. (It is easy to verify that that the problem is well-defined under any metric space $(\mathbb{F}^n, \Delta)$, and even the square of a metric.) Our main motivation to study other distances is to compare their performance, which can potentially lead to new techniques or directions. In addition, Euclidean and squared-Euclidean distances may be related to methods used in practice, such as Singular Value Decomposition.

Euclidean-like distances. We first restrict attention to distances on $\mathbb{R}^n$ that are invariant under unitary transformations of $\mathbb{R}^n$, and where all distances are bounded by 1, referring to these as *invariant* and *bounded* metrics, respectively. Primary examples are the truncated Euclidean distance

$$\min\{\|w - w'\|_2, 1\},$$

the Gaussian kernel $1 - e^{-\|w-w'\|^2}$, and the Laplacian kernel $1 - e^{-\|w-w'\|}$.

Our results in this context analyze sampling, which is a key aspect of the preprocessing phase, and show how effective is it to solve the problem on a (small) subset S chosen at random from the large input W. These results are information-theoretic; they analyze every possible algorithm that may be applied on S, without proposing any specific one. In statistical learning, such results are called generalization bounds, and indeed our proofs boil down to standard machinery like analyzing Rademacher complexity, see Section 4 for details.

Theorem 1.3. *Let Δ^{bi} be a metric on $\mathbb{R}^n$ that is bounded and invariant. Suppose $W \subset \mathbb{R}^n$ is finite and admits a d-dimensional subspace $\tilde{V} \subset \mathbb{R}^n$ such that $\mathbb{E}_{w \in W} \Delta^{bi}(w, \tilde{V}) = \epsilon$. Let S be a sample*

set of $k \geq d^2$ vectors, each drawn independently and uniformly from W. Then for every $\delta > 0$, with probability at least $1 - \delta$ (over the sample S), for every subspace $V' \subset \mathbb{R}^n$, $\dim(V') \leq k$,

$$\mathop{\mathbb{E}}_{w \in W}[\Delta^{bi}(w, V')] \leq 2\epsilon + \mathop{\mathbb{E}}_{w \in S}[\Delta^{bi}(w, V')] + 2\sqrt{\frac{2d^2 \log(ek/d^2)}{k}} + \sqrt{\frac{\log(2/\delta)}{2k}}.$$

One immediate consequence of Theorem 1.3 is a very simple bicriteria-approximation algorithm, as follows. The preprocessing phase just picks a sample set S by choosing independently at random $k = O((\frac{d}{\epsilon})^2 \log \frac{1}{\epsilon})$ vectors from W, and outputs $V' = \mathrm{span}(S)$. Its dimension is clearly at most k (which exceeds d), and by the theorem, with probability at least (say) $3/4$, we have $\epsilon' = \mathbb{E}_{w \in W}[\Delta^{bi}(w, V')] \leq 3\epsilon$. The query phase reports the orthogonal projection of any w onto this V', yielding a reconstructor that is clearly local and explicit.

Another natural choice for V' is to be a d-dimensional space that minimizes $\mathbb{E}_{w \in S}[\Delta^{bi}(w, V')]$. In this case, $\mathbb{E}_{w \in S}[\Delta^{bi}(w, V')] \leq \mathbb{E}_{w \in S}[\Delta^{bi}(w, \tilde{V})]$ and we can further use Hoeffding's inequality to bound the righthand-side, because its expectation (over S) is exactly $\mathbb{E}_S[\mathbb{E}_{w \in S}[\Delta^{bi}(w, \tilde{V})]] = \mathbb{E}_{w \in W}[\Delta^{bi}(w, \tilde{V})] = \epsilon$, and it is the average of iid bounded terms (because Δ^{bi} is bounded). Combining this with Theorem 1.3 and setting k as above would yield a true (not bicriteria) approximation, however, finding a minimizer V' might be computationally nontrivial, which motivates us to study squared Euclidean distances, as follows.

Squared Euclidean distance. In the non-metric but important case of squared Euclidean distance, we design a local reconstructor with $\epsilon' = O(\sqrt{\epsilon})$, under the assumption $\max_{w \in W} \|w\|_2 \leq 1$. This is a variant of Theorem 1.3, as stated below. For a vector $x \in \mathbb{R}^n$, a subspace $V \subset \mathbb{R}^n$, and $p > 0$, let us define $\ell_2^p(x, V) \stackrel{\text{def}}{=} \min_{v \in V} \|x - v\|_2^p$.

Theorem 1.4. *Suppose a finite $W \subset \{x \in \mathbb{R}^n : \|x\|_2 \leq 1\}$ admits a d-dimensional subspace $\tilde{V} \subset \mathbb{R}^n$ such that $\mathbb{E}_{w \in W}[\ell_2^2(w, \tilde{V})] = \epsilon$. Let S be a sample set of $k \geq d^2$ vectors, each drawn independently and uniformly from W. Then for every $\delta > 0$, with probability at least $1 - \delta$ (over the sample S), for every subspace $V' \subset \mathbb{R}^n$, $\dim(V') \leq k$,*

$$\mathop{\mathbb{E}}_{w \in W}[\ell_2^2(w, V')] \leq 8\sqrt{\epsilon} + \mathop{\mathbb{E}}_{w \in S}[\ell_2^2(w, V')] + 2\sqrt{\frac{2d^2 \log(ek/d^2)}{k}} + \sqrt{\frac{\log(2/\delta)}{2k}}.$$

One immediate application of the theorem is a straightforward sampling argument to speed up the computation of truncated SVD of W. Specifically, let the sample S have size $k = O(\frac{d^2}{\epsilon} \log \frac{1}{\epsilon})$, and compute truncated SVD for S (instead of for W). Then the output d-dimensional subspace V' minimizes $\epsilon'_S \stackrel{\text{def}}{=} \mathbb{E}_{w \in S}[\ell_2^2(w, V')]$, which we can bound using Hoeffding's inequality. Indeed, ϵ'_S is the average of iid terms, each bounded by $\ell_2^2(w, \tilde{V}) \leq \|w\|_2^2 \leq 1$, and has expectation $\mathbb{E}_S[\epsilon'_S] \leq \mathbb{E}_S[\mathbb{E}_{w \in S}[\ell_2^2(w, \tilde{V})]] = \epsilon$, thus, with probability at least (say) $1 - \delta$, we have $\epsilon'_S \leq \epsilon + \sqrt{\frac{2\ln(1/\delta)}{k}}$. Combined this with the theorem, with probability at least $1 - 2\delta$ we have $\mathbb{E}_{w \in W}[\ell_2^2(w, V')] \leq O(\sqrt{\epsilon}) + 2\sqrt{\frac{\log(2/\delta)}{k}}$.

Related Work. A related problem is Subspace Approximation, defined as follows for a parameter $p > 0$: Given a set $W \subset \mathbb{R}^n$ and some $d > 0$, find a d-dimensional subspace $V^* \subset \mathbb{R}^n$ that minimizes $\mathbb{E}_{w \in W}[\ell_2^p(V^*, w)]$. In the special case $p = 2$ one can just use truncated SVD, and the case of general p was studied e.g. in [Cla05, SV07, DV07, DTV11] (see also [HP06] for an even more L_p-fitting

problem). These results address the optimization problem of computing a subspace, while our result for Subspace Reconstruction under a bounded invariant metric focuses on sampling, completely avoiding the optimization problem. Another difference is that our results require the distances to be invariant (rather than restricted to Euclidean powers) and also bounded.

For squared-Euclidean distances, i.e., $p = 2$, the sampling approach was previously studied in [FKV04, Theorem 2] and [DKM06, Theorem 4]. Their analysis is optimized to sampling vectors from W non-uniformly (proportionally to their squared-length), and the result of applying it to uniform sampling is incomparable to ours.

1.5 Relation To Property Testing

A related problem, called low-rank testing, is to distinguish whether an input matrix M has rank at most d or is ϵ-far from all such matrices (here d and ϵ could be either fixed parameters or part of the input). This problem was introduced in [KS03], who showed a testing algorithm that reads a random submatrix of size $O(d/\epsilon) \times O(d/\epsilon)$, and the query complexity was improved to $O(d^2/\epsilon)$ in [LWW14] using an adaptive testing algorithm. The key lemma in [KS03] shows that if most submatrices of M of size $O(d/\epsilon) \times O(d/\epsilon)$ have rank at most d, then M itself is ϵ-close to a rank-d matrix. Our main result deals with a different regime, where M is *guaranteed* to be ϵ-close to a low-rank matrix, and the goal is to locally reconstruct such a low-rank matrix. To this end, we show that a typical random submatrix of size $O(d/\epsilon) \times O(d/\epsilon)$ suffices not just to learn about the distance to a low-rank matrix, but rather almost determines a low-rank approximation of M, as explained in Section 1.3.

One of the results in [DDG+15] analyzes the following algorithm for testing if a matrix has rank at most 1: pick a 2×2 submatrix of M and accept iff its rank is at most 1. They show that if M is ϵ-far from every rank-1 matrix then the test rejects with probability at least ϵ; this is an oblivious tester, i.e., the number of queries does not depend on ϵ. Using similar (but simpler) ideas to our local reconstruction, we can extend that oblivious tester to any rank d. Specifically, the tester picks a random $(d+1) \times (d+1)$ submatrix of M and accepts iff its rank is at most d. We can show that if M is ϵ-far from every matrix of rank at most d, the test rejects with probability at least ϵ^d. Details omitted.

2 Preliminaries

Let $M \in \mathbb{F}^{n \times m}$ be a matrix over a field $\mathbb{F}$. For $S \subset [n], T \subset [m]$, we denote by $M_{S,T}$ the submatrix of M confined to the rows S and columns T. For instance, we denote by $M_{S,[n]}$ ($M_{[m],T}$) the S-rows (T-columns) of M. For a row $i \in [n]$ (column $j \in [m]$) we denote by $M_{i,[m]}$ ($M_{[n],j}$) the i-th row (j-th column) of M.

Recall that $\mathrm{rank}(M)$ is the dimension of M's columns space. For $d, n, m \in \mathbb{N}$ we denote

$$\underset{n,m}{\textsc{rank}}(d) := \{M \in \mathbb{F}^{n \times m} \mid \mathrm{rank}(M) \leq d\},$$

and omit n, m whenever they are clear from the context.

Definition 2.1. *Let $\mathbb{F}$ be a field, and let $d, n, m \in \mathbb{N}$. We say that a matrix $M \in \mathbb{F}^{n \times m}$ is ϵ-close to $\textsc{rank}(d)$ if there exists a matrix $\tilde{M} \in \textsc{rank}_{n,m}(d)$, satisfying $\Delta(\tilde{M}, M) \leq \epsilon$. Otherwise, we say that M is ϵ-far from $\textsc{rank}(d)$.*

3 Local Reconstruction of Low-Rank Matrices

3.1 Reconstruction from Adversarial Noise

In this section we prove our main result, Theorem 1.2, restated below.

Theorem 1.2 (Main Theorem). *For every field $\mathbb{F}$ there is a local $O(\sqrt{d\epsilon})$-reconstructor with parameters $q_{\text{prep}} = \tilde{O}\left(\sqrt{\frac{d}{\epsilon^3}}\right)$, $t_{\text{prep}} = \tilde{O}\left(\frac{1}{\epsilon}\left(\frac{(8e)^2 \ln^2 d}{d\epsilon}\right)^d\right)$, $s = O(d^2)$, $q_{\text{query}} = 2d$ and $t_{\text{query}} = O(d^2)$.*

Perhaps surprisingly, the proof of this theorem evolves from the following easy fact about matrices of rank d (for completeness, we provide its proof in Appendix A). More precisely, the key technical step in our proof is Lemma 3.3, which can be viewed as a "robust" version of the following fact.

Fact 3.1. *Let $M \in \mathbb{F}^{n\times m}$, and let $R \subset [n], C \subset [m]$ be of size $|R| = |C| = \text{rank}(M)$. If $M_{R,C}$ is invertible,[3] then*

$$\forall (i,j) \in [n] \times [m], \qquad M_{i,j} = M_{i,C} \left(M_{R,C}\right)^{-1} M_{R,j}.$$

This fact motivates the following definition.

Definition 3.2. *Let $R \subset [n], C \subset [m]$ be of size $d \geq 1$. We say that the pair (R, C) is an ϵ-*core *of $M \in \mathbb{F}^{n\times m}$ if $M_{R,C}$ is invertible and*

$$\Pr_{(i,j)}[M_{i,j} = M_{i,C}(M_{R,C})^{-1}M_{R,j}] \geq 1 - \epsilon.$$

In the case $d = 0$, we say that $(R, C) = (\emptyset, \emptyset)$ is an ϵ-core of M if

$$\Pr_{(i,j)}[M_{i,j} = 0] \geq 1 - \epsilon.$$

For every M that is ϵ-close to $\text{RANK}(d)$, by Fact 3.1 there exists a pair (R, C) that is an ϵ-core of M. Our key step, stated in Lemma 3.3, finds an ϵ'-core efficiently, i.e., the number of queries depends only on d and ϵ.

Once we find (R, C) that constitutes an ϵ'-core of M, the preprocessing and the query phases can be implemented as follows. The preprocessing phase stores R, C and $(M_{R,C})^{-1}$ in the data structure S. In the query phase, when a query (i, j) is given, the algorithm returns $M_{i,C}(M_{R,C})^{-1}M_{R,j}$, which requires access to only $|R| + |C|$ entries of M. This algorithm clearly provides direct access to $M' \in \text{RANK}(d)$, and it remains to show that with high probability M' is ϵ'-close to M. A full description of the reconstructor algorithm and its analysis appears in Section 3.1.1. First we prove our main lemma.

Lemma 3.3 (Main Lemma). *Let $M \in \mathbb{F}^{n\times m}$ be ϵ-close to $\text{RANK}(d)$ with $\epsilon d < \frac{1}{324}$. Then for $\epsilon' = 18\sqrt{d\epsilon}$,*

$$\Pr_{\substack{R\subset[n],C\subset[m],\\ |R|=|C|=8\sqrt{d/\epsilon}\ln d}} \left[\exists \tilde{R} \subset R, \tilde{C} \subset C, \text{ s.t. } |\tilde{R}| = |\tilde{C}| \leq d \text{ and } (\tilde{R}, \tilde{C}) \text{ is } \epsilon'\text{-core of } M\right] \geq 11/12.$$

[3] Clearly, there always exist $R \subset [n], C \subset [m]$ of size $|R| = |C| = \text{rank}(M)$ such that $M_{R,C}$ is invertible.

Proof. Let $\tilde{M} \in \mathrm{rank}(d)$ be a matrix that is ϵ-close to M. A row r is called *noisy* if $\Delta(M_{r,[m]}, \tilde{M}_{r,[m]}) \geq 8\sqrt{\epsilon/d}$ and similarly for a column. We need the following claim, which we shall prove soon.

Claim 3.4. *Let $M \in \mathbb{F}^{n\times m}$ be ϵ-close to $\tilde{M}$ where $\mathrm{rank}(\tilde{M}) = d$ and let $\tilde{R} \subset R, \tilde{C} \subset C$ be each of size t. The pair $(\tilde{R}, \tilde{C})$ is called a super-core if*

- *$(\tilde{R}, \tilde{C})$ is $\sqrt{d\epsilon}$-core of $\tilde{M}$;*
- *$M_{\tilde{R},\tilde{C}} = \tilde{M}_{\tilde{R},\tilde{C}}$; and*
- *each $r \in \tilde{R}, c \in \tilde{C}$ is non-noisy.*

Then

$$\Pr_{R,C:\ |R|=|C|=8\sqrt{d/\epsilon}\ln d}\left[\ \exists \tilde{R} \subset R, \tilde{C} \subset C \text{ such that } (\tilde{R}, \tilde{C}) \text{ is a super-core}\right] \geq 11/12.$$

To use the claim, it suffices to show that every super-core $(\tilde{R}, \tilde{C})$ is a ϵ'-core for M. To see this, fix such $\tilde{R}, \tilde{C}$ and let M' defined by $M'_{i,j} = M_{i,\tilde{C}}(\tilde{M}_{\tilde{R},\tilde{C}})^{-1}M_{\tilde{R},j}$. Let us show that $\Delta(M, M') < \epsilon'$.

Indeed, since $\tilde{R}, \tilde{C}$ constitutes a $\sqrt{d\epsilon}$-core for $\tilde{M}$

$$\Pr_{(i,j)}[\tilde{M}_{i,j} \neq \tilde{M}_{i,\tilde{C}}(\tilde{M}_{\tilde{R},\tilde{C}})^{-1}\tilde{M}_{\tilde{R},j}] < \sqrt{d\epsilon}.$$

Next, since every $r \in \tilde{R}$ and $c \in \tilde{C}$ is non-noisy, by straightforward union bound,

$$\Pr_{(i,j)}\left[M_{i,\tilde{C}} \neq \tilde{M}_{i,\tilde{C}} \text{ and } M_{\tilde{R},j} \neq \tilde{M}_{\tilde{R},j}\right] < (|\tilde{R}| + |\tilde{C}|) \cdot 8\sqrt{\epsilon/d} \leq 16\sqrt{d\epsilon}.$$

When these two event do not occur $\tilde{M}_{i,j} = M'_{i,j}$, and we thus have $\Delta(M', \tilde{M}) < 17\sqrt{d\epsilon}$.

Recalling that $\Pr_{(i,j)}[M_{i,j} \neq \tilde{M}_{i,j}] = \Delta(M, \tilde{M}) < \epsilon$, we get by a union bound that $\Delta(M, M') < 18\sqrt{d\epsilon}$, which proves Lemma 3.3. □

Proof of Claim 3.4. Let $\tilde{M} \in \textsc{rank}(d)$ be a matrix that is ϵ-close to M. We shall consider the random choice of R, C as if it is done in d steps, adding at each step $8\ln d/\sqrt{\epsilon}$ rows and columns (respectively).

If $\tilde{M}$ is $\sqrt{d\epsilon}$-close to the 0-matrix, then the claim trivially holds with $\tilde{R} = \tilde{C} = \emptyset$. Otherwise, with probability at least $\sqrt{d\epsilon}$ over the choice of (i, j) it holds that $\mathrm{rank}(\tilde{M}_{i,j}) = 1$. Next observe that by averaging the fraction of noisy rows (and similarly columns) is bounded by $\sqrt{d\epsilon}/8$. Furthermore, the probability that $M_{i,j} \neq \tilde{M}_{i,j}$ is bounded by ϵ. Overall, using a union bound, all three events, namely $\mathrm{rank}(\tilde{M}_{i,j}) = 1$, both the i-th row and the j-th column are non-noisy, and $M_{i,j} = \tilde{M}_{i,j}$, happen with probability at least $\sqrt{d\epsilon} - 2\sqrt{d\epsilon}/8 - \epsilon \geq \sqrt{d\epsilon}/2$. Therefore, a pair (R_1, C_1) of cardinality $|R_1| = |C_1| = 8\ln(d)/\sqrt{d\epsilon}$ contains (i, j) such that $\mathrm{rank}(\tilde{M}_{i,j}) = 1$, both i and j are non-noisy and $M_{i,j} = \tilde{M}_{i,j}$ with probability at least $1 - (1 - \sqrt{d\epsilon}/2)^{8\ln d/\sqrt{d\epsilon}} \geq 1 - e^{-4\ln d} \geq 1 - \frac{e^{-3}}{d} \geq 1 - \frac{1}{6d}$.

Assuming that happens, choose such a pair (i, j) and let $\tilde{R} = \{i\}, \tilde{C} = \{j\}$. If $(\tilde{R}, \tilde{C})$ is now an $\sqrt{d\epsilon}$-core of $\tilde{M}$ then (R_1, C_1) is a super-core and the claim holds. Otherwise, with probability at least $\sqrt{d\epsilon}$ over $i \in [n] \setminus \tilde{R}, j \in [m] \setminus \tilde{C}$, it holds that $\mathrm{rank}(\tilde{M}_{\tilde{R}\cup\{i\},\tilde{C}\cup\{j\}}) = 2$. The probability that both the i-th row and the j-th column are non-noisy is at least $1 - \sqrt{d\epsilon}/4$. The probability that $M_{i,\tilde{C}} = \tilde{M}_{i,\tilde{C}}$, $M_{\tilde{R},j} = \tilde{M}_{\tilde{R},,j}$ and $M_{i,j} = \tilde{M}_{i,j}$ is at least $1 - 3\epsilon \geq 1 - \sqrt{d\epsilon}/4$. Therefore, for

a random set (R_2, C_2) of cardinality $|R_2| = |C_2| = 8 \ln d/\sqrt{d\epsilon}$, with probability at least $1 - \frac{1}{6d}$ the following holds: $(R_1 \cup R_2, C_1 \cup C_2)$ contains a pair (i, j) such that $\text{rank}(\tilde{M}_{\tilde{R}\cup\{i\},\tilde{C}\cup\{j\}}) = 2$ both i and j are non-noisy, and further $M_{i,\tilde{C}} = \tilde{M}_{i,\tilde{C}}$, $M_{\tilde{R},j} = \tilde{M}_{\tilde{R},,j}$ and $M_{i,j} = \tilde{M}_{i,j}$. Assume that this event happens, let (i, j) be such a pair, and set $\tilde{R} = \tilde{R} \cup \{i\}, \tilde{C} = \tilde{C} \cup \{j\}$.

If $(\tilde{R}, \tilde{C})$ is now a $\sqrt{d\epsilon}$-core of $\tilde{M}$ then $(\tilde{R}, \tilde{C})$ is a super-core and the claim holds. Otherwise, we proceed to pick (R_3, C_3) and so on. In each step $t < \text{rank}(\tilde{M})$, there are two possible cases. In the first case, $(R_1 \cup \ldots \cup R_{t-1}, C_1 \cup \ldots \cup C_{t-1})$ contains a super-core with probability $1 - \frac{1}{6d}$ (conditioned on previous steps). In the second case, with probability at least $\sqrt{d\epsilon}/4$ we can enlarge $\tilde{R}, \tilde{C}$ to size t such that $\text{rank}(\tilde{M}_{\tilde{R},\tilde{C}}) = t$ and in addition both the new row and column are non-noisy and $M_{\tilde{R},\tilde{C}} = \tilde{M}_{\tilde{R},\tilde{C}}$.

At step $t = \text{rank}(\tilde{M}) + 1$ (if reached), with probability at least $1 - \frac{1}{6d}$ (conditioned on success of previous steps) we have that $(R_1 \cup \ldots \cup R_t, C_1 \cup \ldots \cup C_t)$ contains $\tilde{R}, \tilde{C}$ with $\text{rank}(\tilde{M}_{\tilde{R},\tilde{C}}) = d$ and in addition both the new row and column are non-noisy and $M_{\tilde{R},\tilde{C}} = \tilde{M}_{\tilde{R},\tilde{C}}$. In this step we cannot find i, j such $\text{rank}(\tilde{M}_{\tilde{R}\cup\{i\},\tilde{C}\cup\{j\}}) = d + 1$. Nevertheless, by Fact 3.1, such a pair $(\tilde{R}, \tilde{C})$ constitutes a 0-core, and certainly $(\tilde{R}, \tilde{C})$ is a super-core. The success probability over all steps is at least $(1 - \frac{1}{6d})^d \geq \frac{5}{6}$, and the claim follows. □

In what follows we show that if M aside of being ϵ-close to $\text{RANK}(d)$ is also far from $\text{RANK}(d-1)$, then a typical choice of $\tilde{O}(\frac{d\delta}{\epsilon})$ rows and column would contain an $O(d\epsilon)$-core, compared to $O(\sqrt{d\epsilon})$ as in Claim 3.4. This in turn will be utilized for better reconstruction algorithm from random matrices, see details in Section 3.2.

Claim 3.5. *Let $M \in \mathbb{F}^{n\times m}$ be ϵ-close to $\tilde{M} \in \text{RANK}(d)$ and δ-far from $\text{RANK}(d-1)$ where $\delta > \max\{1/\sqrt{\epsilon}, 8d\epsilon\}$ and let $\tilde{R} \subset [n], \tilde{C} \subset [m]$ be each of size t. The pair $(\tilde{R}, \tilde{C})$ is called a super-core if*

- *$(\tilde{R}, \tilde{C})$ is ϵ/δ-core of $\tilde{M}$;*
- *$M_{\tilde{R},\tilde{C}} = \tilde{M}_{\tilde{R},\tilde{C}}$; and*
- *each $r \in \tilde{R}$ is non-noisy i.e., $\Delta(M_{r,[m]}, \tilde{M}_{r,[m]}) < 8\epsilon/\delta$ and similarly each $c \in \tilde{C}$.*

Then for random R, C of size $\frac{8d\delta \ln d}{\epsilon}$

$$\Pr_{R,C}[\ \exists \tilde{R} \subset R, \tilde{C} \subset C \text{ such that } (\tilde{R}, \tilde{C}) \text{ is a super-core}] \geq 11/12.$$

Proof Sketch of Claim 3.5. Let $\tilde{M} \in \text{rank}(d)$ be a matrix that is ϵ-close to M, since M is δ-far from $\text{RANK}(d-1)$, then by the triangle inequality $\tilde{M}$ is $\delta - \epsilon \geq \epsilon/2$-far from $\text{RANK}(d-1)$. As in the proof of Claim 3.4 we shall consider the random choice of R, C as if it is done in d steps, adding at each step $8 \log d\delta/\epsilon$ rows and columns (respectively).

The key idea is that on each step $t < d$, with probability at least $\delta/2$ we can enlarge $\tilde{R}, \tilde{C}$ such that $\text{rank}(\tilde{M}_{\tilde{R},\tilde{C}}) = t$ and in addition both the new row and column are non-noisy and i, j are such that $M_{i,\tilde{C}} = \tilde{M}_{i,\tilde{C}}$ and $M_{\tilde{R},j} = \tilde{M}_{\tilde{R},j}$ (provided that $(2t-1)\epsilon < \delta/4$). Therefore, by the same reasoning applied in the proof of Claim 3.4, we can conclude the proof, we omit the details. □

3.1.1 Proof of the main theorem

We can now prove Theorem 1.2. Let us first introduce the preprocessing and the query phases of reconstruction algorithm. As mentioned earlier, our data structure S will consist of subsets $\tilde{R} \subset [n], \tilde{C} \subset [m]$ and a matrix A.

Algorithm 1 Preprocessing Algorithm

Input: $M \in \mathbb{F}^{n\times m}$ (via direct access), $d \in \mathbb{N}$ and $\epsilon > 0$.
Output: subsets $\tilde{R} \subset [n], \tilde{C} \subset [m]$ of the same size $\ell \leq d$, and a matrix $A \in \mathbb{F}^{\ell\times\ell}$.

1: $\epsilon' = 36\sqrt{d\epsilon}$
2: $t = \frac{c}{\epsilon}\ln\left(\frac{\ln d}{d\epsilon}\right)$ for suitable constant $c > 0$ {thus $t = \tilde{O}(\frac{1}{\epsilon})$}
3: pick $T \subset [n] \times [m]$ by drawing t pairs independently uniformly at random
4: **if** $\Pr_{(i,j)\in T}[M_{i,j} = 0] \geq 1 - \epsilon'$ **then**
5: **return** $R = C = \emptyset$ and an empty matrix A
6: pick random $R \subset [n]$, $C \subset [m]$ of size $|R| = |C| = 8\sqrt{\frac{d}{\epsilon}}\ln d$
7: **for** $\ell = 1, \ldots, d$ **do**
8: **for all** $\tilde{R} \subset R$, $\tilde{C} \subset C$ of size ℓ **do**
9: **if** $\mathrm{rank}(M_{\tilde{R},\tilde{C}}) = \ell$ and $\Pr_{(i,j)\in T}[M_{i,j} = M_{i,\tilde{C}} \cdot (M_{\tilde{R},\tilde{C}})^{-1} \cdot M_{\tilde{R},j}] > 1 - \epsilon'$ **then**
10: **return** $(\tilde{R}, \tilde{C})$ and $M_{\tilde{R},\tilde{C}}$

Algorithm 2 Query Phase Algorithm

Input: location $(i, j) \in [n] \times [m]$, matrix $M \in \mathbb{F}^{n\times m}$ (via direct access), subsets $\tilde{R} \subset [n], \tilde{C} \subset [m]$ each of size $\ell \leq d$ and a matrix $A \in \mathbb{F}^{\ell\times\ell}$.
Output: $M'_{i,j}$.

1: **return** $M'_{i,j} = M_{i,\tilde{C}} A M_{\tilde{R},j}$

The next two lemmas, proved further below, analyze the effect of the random sample T in Algorithm 1.

Lemma 3.6 (Completeness)**.** *With probability at least* $5/6$ *over* R, C *and* T, *Algorithm 1 enumerates over some* $\tilde{R}$, $\tilde{C}$ *that satisfy the condition in Step 9.*

Lemma 3.7 (Soundness)**.** *With probability at least* $5/6$ *over* R, C *and* T, *every choice in Algorithm 1 of* $\tilde{R}$, *and* $\tilde{C}$, *that satisfies the condition in Step 9, satisfies also*

$$\Pr_{(i,j)\in[n]\times[m]}[M_{i,j} = M_{i,\tilde{C}} \cdot (M_{\tilde{R},\tilde{C}})^{-1} \cdot M_{\tilde{R},j}] \geq 1 - 2\epsilon'.$$

Let us use these two lemmas to complete the proof of Theorem 1.2. These lemmas clearly show that Algorithms 1 and 2 constitute a $2\epsilon'$-reconstructor. It is then immediate that the reconstructor is local, and its parameters are $(q_{\mathrm{prep}}, t_{\mathrm{prep}}, s, q_{\mathrm{query}}, t_{\mathrm{query}})$ as follows. Clearly, $t_{\mathrm{prep}} = O(d\binom{8\sqrt{d/\epsilon}\ln d}{d}^2 \cdot td^2) = \tilde{O}((\frac{8e\ln d}{\sqrt{d\epsilon}})^{2d}\epsilon^{-1})$, $s = d^2 + 2d, q_{\mathrm{query}} = 2d$ and $t_{\mathrm{query}} = O(d^2)$, and it thus remains to bound q_{prep}. Observe that for every $(i, j) \in T$ and every choice of $\tilde{R}, \tilde{C}$ we use the entries $M_{i,j}$, $M_{\tilde{R},j}$, $M_{i,\tilde{C}}$ and $M_{\tilde{R},\tilde{C}}$. We only need to query M in the entries $\{M_{i,j}\}_{(i,j)\in T}$, $\{M_{i,C}\}_{i|\exists j,(i,j)\in T}$, $\{M_R, j\}_{j|\exists i(i,j)\in T}$,

and $M_{R,C}$ and therefore $q_{\text{prep}} = t(|C|+|R|+1)+|R|\cdot|C| = O(t\cdot|R|) = O\left(\sqrt{\frac{d}{\epsilon}}\ln d\cdot\frac{1}{\epsilon}\ln(\frac{\ln d}{d\epsilon})\right) = \tilde{O}\left(\sqrt{\frac{d}{\epsilon^3}}\right)$. This proves Theorem 1.2, and it remains to prove the two lemmas.

Proof of Lemma 3.6. By Lemma 3.3, for a random choice of R, C with probability at least $11/12$, there exists $\tilde{R} \subset R, \tilde{C} \subset C$ such that $(\tilde{R}, \tilde{C})$ is $18\sqrt{d\epsilon}$-core of M. In such a case

$$\Pr_{(i,j)\in[n]\times[m]}[M_{i,j} = M_{i,\tilde{C}}\cdot(M_{\tilde{R},\tilde{C}})^{-1}\cdot M_{\tilde{R},j}] \geq 1 - 18\sqrt{d\epsilon} = 1-\epsilon'/2.$$

Fix this choice of $\tilde{R}$, and $\tilde{C}$. Since T is a random sample from $[n]\times[m]$,

$$\mathop{\mathbb{E}}_{T}\left[\Pr_{(i,j)\in T}[M_{i,j} = M_{i,\tilde{C}}\cdot(M_{\tilde{R},\tilde{C}})^{-1}\cdot M_{\tilde{R},j}]\right] = \Pr_{(i,j)\in[n]\times[m]}[M_{i,j} = M_{i,\tilde{C}}\cdot(M_{\tilde{R},\tilde{C}})^{-1}\cdot M_{\tilde{R},j}],$$

and now applying Höeffding bound, the event of additive deviation by $\epsilon'/2$, namely,

$$\Pr_{(i,j)\in T}[M_{i,j} = M_{i,\tilde{C}}\cdot(M_{\tilde{R},\tilde{C}})^{-1}\cdot M_{\tilde{R},j}] < 1-\epsilon',$$

occurs with probability (over the choice of T) at most $e^{-\Omega(t(\epsilon')^2)} < 1/12$, for a suitable $t = \Omega(1/\epsilon'^2)$. The claim follows. □

Proof of Lemma 3.7. Similarly to Lemma 3.6, for every $\tilde{R}$ and $\tilde{C}$, the probability (over T) that

$$\left|\Pr_{(i,j)\in T}[M_{i,j} = M_{i,\tilde{C}}\cdot(M_{\tilde{R},\tilde{C}})^{-1}\cdot M_{\tilde{R},j}] - \Pr_{(i,j)\in[n]\times[m]}[M_{i,j} = M_{i,\tilde{C}}\cdot(M_{\tilde{R},\tilde{C}})^{-1}\cdot M_{\tilde{R},j}]\right| > \epsilon' \quad (2)$$

is at most $e^{-\Omega(t(\epsilon')^2)} = e^{-\Omega(td\epsilon)}$.

Observe that Algorithm 1 enumerates over $\sum_{i=1}^{d}\binom{8\sqrt{d/\epsilon}\ln d}{i}^2 \leq d(\frac{8e\ln d}{\sqrt{d\epsilon}})^{2d} = e^{O(d\ln(\frac{\ln d}{d\epsilon}))}$ values for $\tilde{R}$ and $\tilde{C}$. The lemma follows by applying a union bound on all these values, using our choice of t. □

3.2 Reconstruction from Random Matrices

So far we have dealt with the reconstruction task under a worst case assumption, i.e., we only assumed that $\tilde{M}$ is of low rank. Can we do better when $\tilde{M}$ is a random low rank matrix? Turns out that in this case, with high probability we can do better.

Let $M = \tilde{M} + N$ be obtained by first picking a matrix $\tilde{M}$ (uniformly at random from $\textsc{rank}(d)$) and then adding some sparse matrix N i.e. $\Delta(N, \mathbf{0}) < \epsilon$, where N is adversarially chosen posterior of the choice of $\tilde{M}$. The crucial observation is that with high probability $\tilde{M}$ is δ-far from any matrix of $\text{rank}(d-1)$ (where $\delta > 0$ is some absolute constant), yielding that M is ϵ-close to $\textsc{rank}(d)$ but $(\delta-\epsilon)$-far from $\textsc{rank}(d-1)$.

We now show that under the stronger assumption that M is ϵ-close to $\textsc{rank}(d)$ but δ-far from $\textsc{rank}(d-1)$, and $\epsilon \ll \delta^2$, there exists a local reconstruction with much better parameters. In particular $\epsilon' = O(d\epsilon/\delta)$ compared to $O(d\sqrt{\epsilon})$ in Theorem 1.2.

Theorem 3.8. *Let $\mathbb{F}$ be a field. Let $M \in \mathbb{F}^{n\times m}$ be ϵ-close to* RANK(d) *and δ-far from* RANK$(d-1)$*, where $\delta > \max\{1/\sqrt{\epsilon}, 8d\epsilon\}$. Then there exists a local reconstructor for M with parameters $\epsilon' = O(d\epsilon/\delta)$, $q_{\text{prep}} = \tilde{O}\left(\frac{\delta d^2}{\epsilon^3}\right)$, $t_{\text{prep}} = \tilde{O}((\frac{8\delta \ln d}{\epsilon})^d)$, $s = O(d^2)$, $q_{\text{query}} = 2d$, and $t_{\text{query}} = O(d^2)$.*

The proof is similar to the proof of Theorem 1.2 in Section 3.1.1. However, it uses Claim 3.5 instead of Claim 3.4. In particular we have to modify Algorithm 1 line 2 so that $t = \frac{d}{\epsilon^2}\ln\left(\frac{8d\delta \ln d}{\epsilon}\right)$ and line 6 so that $|R| = |C| = \frac{8d\delta \ln d}{\epsilon}$ to get the parameters stated in Theorem 3.8.

4 Subspace Reconstruction

In this section, we study the Subspace Reconstruction problem, starting with a formal definition of the problem. We then show (in Section 4.1) that under Hamming distance, we can solve this problem as an immediate application of our Matrix Reconstruction result (Theorem 1.2).

We then study this problem for real spaces, where distances are Euclidean or similar. We start by presenting some tools from statistical learning, like Rademacher complexity, VC-dimension and generalization bounds, and then prove Theorems 1.3 and 1.4.

Definition 4.1. *An algorithm whose input is a list $w_1, \ldots, w_m \subset \mathbb{F}^n$ together with $d \in \mathbb{N}$ and $\epsilon > 0$ is called an* ϵ'-subspace reconstructor *(where ϵ' is a function of ϵ and d) if:*

1. *its output is a subspace V' of dimension at most d; and*
2. *if $\mathbb{E}_{i\in[m]}[\Delta(w_i, \tilde{V})] \le \epsilon$ for some subspace $\tilde{V}$ of dimension at most d, then with probability at least $2/3$ we have $\mathbb{E}_{i\in[m]}\Delta(w_i, V') \le \epsilon'$.*

An ϵ'-subspace reconstructor is called local with parameters $q_{\text{prep}}, t_{\text{prep}}, s, t_{\text{query}}$ *(which are functions of ϵ and d) if it consists of two phases as above, where:*

3. *the preprocessing phase runs in time t_{prep} and accesses at most q_{prep} input vectors, and creates a data structure S of size s words; and*
4. *the query phase runs in time t_{query}.*

Such a local reconstructor is called explicit *if t_{query} is polynomial in n.*

4.1 Subspace Reconstruction under Hamming Distance

Observe that given a local matrix reconstructor we can easily design a local subspace reconstructor, which has the same preprocessing phase (except that we now query whole vectors) and whose query algorithm executes, for each entry in the desired column, the matrix reconstructor's query algorithm. Thus, Theorem 1.2 immediately implies the following result for subspace reconstruction.

Theorem 4.2. *For every field $\mathbb{F}$ there is an $O(\sqrt{d\epsilon})$-subspace reconstructor (with respect to Hamming distance) with parameters $q_{\text{prep}} = \tilde{O}\left(\sqrt{\frac{d}{\epsilon^3}}\right)$, $t_{\text{prep}} = \tilde{O}\left(\frac{1}{\epsilon}(\frac{(8e)^2\ln^2 d}{d\epsilon})^d\right)$, $s = O(d^2)$, and $t_{\text{query}} = O(d^2 n)$.*

4.2 Generalization Bounds

For introduction to this subject we refer the reader to Chapter 3 in [MRT12].

Let $\mathbf{D}$ be a distribution over a domain Z, usually the uniform distribution, and let $\mathbf{D}^k$ be the distribution of picking k elements from $\mathbf{D}$ independently. For $\mathbf{H}$ a set of subsets of Z, we call $\mathbf{H}$ an hypothesis set [4]. For $h \in \mathbf{H}$ let $g_h : z \in Z \to [0,1]$ be a loss function, we say $g = \{g_h | h \in \mathbf{H}\}$ is the class of loss functions associated with $\mathbf{H}$. Let $E[h,g] = \mathbb{E}_{z\sim\mathbf{D}}[g_h(z)]$ be the average loss of hypothesis h with respect to the loss function class g and let $\hat{E}_S[h,g] = \mathbb{E}_{s\in_R S}[g_h(s)]$ be similarly the empirical loss on the sample set $S \subset Z$. Define

$$\Phi(\mathbf{H}, g, Z, S) = \sup_{h\in\mathbf{H}} \left(E[h,g] - \hat{E}_S[h,g] \right).$$

The following theorem can be derived from the proof of Theorem 3.1 in [MRT12].

Theorem 4.3. *Let $\mathbf{H}$ be a hypothesis set and g be a class of loss functions. Then, for any $\delta > 0$, with probability $1-\delta$ over a sample set S of k elements from $\mathbf{D}$,*

$$\forall h \in \mathbf{H}, \quad E[h,g] \le \hat{E}_S[h,g] + \mathop{\mathbb{E}}_{S'\sim\mathbf{D}^k} \left[\Phi(\mathbf{H}, g, Z, S')\right] + \sqrt{\frac{\log\frac{2}{\delta}}{2k}}.$$

4.3 Rademacher Complexity

The Rademacher complexity captures the richness of a family of hypothesis functions by measuring the degree to which a hypothesis set can fit random noise. The following states the formal settings and definitions of the empirical Rademacher complexity.

Let $\mathbf{D}$ be a distribution over a domain Z. Let $\mathbf{H}$ be a family of hypothesis functions, and for each $h \in \mathbf{H}$ let $g_h : Z \to [0,1]$ denote the loss of hypothesis h on the element $z \in Z$. Let $g = \{g_h | h \in \mathbf{H}\}$ be a class of loss functions.

Let $S = (s_1, \ldots, s_k)$ be a fixed sample set of size k from $\mathbf{D}$.

Definition 4.4. The empirical Rademacher complexity[5] of $\mathbf{H}$ with respect to the sample set S is defined as

$$\hat{\mathcal{R}}_S(\mathbf{H}, g) = \mathop{\mathbb{E}}_{\sigma\in\{-1,1\}^k} \left[\sup_{h\in H} \frac{1}{k} \sum_{i=1}^{k} \sigma_i g_h(s_i) \right].$$

Definition 4.5. For any integer $k \ge 1$, the order k Rademacher complexity of $\mathbf{H}$ is the expectation of the empirical Rademacher complexity over all sample sets of size k drawn according to $\mathbf{D}$:

$$\mathcal{R}_k(\mathbf{H}, g) = \mathop{\mathbb{E}}_{S\sim\mathbf{D}^k} [\hat{\mathcal{R}}_S(\mathbf{H}, g)].$$

The following inequality is used to derive a generalization bound when the hypothesis set, (and the corresponding loss function) has a small Rademacher complexity (the proof of this inequality can be easily derived from the proof of Theorem 4.7 in [MRT12]).

[4] We deal with subsets, rather than a more general setting, to allow a more pleasant treatment of the VC-dimension definition, see Section 4.4.

[5] Here we use a definition with both loss function and hypothesis sets in order to make our arguments more clear.

Theorem 4.6.

$$\mathop{\mathbb{E}}_{S\sim\mathbf{D}^k}\left[\Phi\left(\mathbf{H},g,Z,S\right)\right]\leq 2\mathcal{R}_k(H,g).$$

The following generalization bound is an immediate corollary of Theorem 4.3 and Theorem 4.6.

Theorem 4.7. *Let* $\mathbf{H}$ *be a family of hypothesis sets and* g *be a class of loss function. Then, for any* $\delta>0$*, with probability* $1-\delta$ *over a sample set* S *of* k *elements from* $\mathbf{D}$*,*

$$\forall h\in H,\quad E\left[h,g\right]\leq\hat{E}_S\left[h,g\right]+2\mathcal{R}_S(\mathbf{H},g)+\sqrt{\frac{\log\frac{2}{\delta}}{2k}}.$$

4.4 VC-dimension

In this section we define the VC-dimension, which intuitively captures the richness of a hypothesis set. We note that the VC-dimension is affected only by the hypothesis set and not by any loss function class associated with it or by the distribution of sample points.

Definition 4.8. Given a set $S\subset Z$ we say $h\in\mathbf{H}$ *separates* $A\subseteq S$ if $h\cap S=A$. A hypothesis set *shatters* a set S if for every $A\subseteq S$ there exists $h\in\mathbf{H}$ such that h separates A.

Definition 4.9. The VC-dimension of a hypothesis set $\mathbf{H}$ is the size of the largest set that can be shattered by $\mathbf{H}$.

The following theorem can be found in [MRT12].

Theorem 4.10. *Let* $\mathbf{H}$ *be a hypothesis set with VC-dimension* d*. For all* $k\geq 1$*,*

$$\mathcal{R}_k\left(\mathbf{H},\{\mathbf{1}_h\}_{h\in\mathbf{H}}\right)\leq\sqrt{\frac{2d\log\frac{ek}{d}}{k}}.$$

Where $\mathbf{1}_h$ *is the zero one indicator function for the set* $h\subset Z$*.*

For the rest of this section assume $Z=\mathbb{F}^n$, where $\mathbb{F}$ is an arbitrary field. We also define the hypothesis set of d-dimensional subspaces.

Definition 4.11. $V_d=\{\mathbf{v}|\mathbf{v}\text{ is a }d\text{-dimensional subspace of }\mathbb{F}^n\}$.

Theorem 4.12. *The VC-dimension of* V_d *is* d*.*

Proof. To see that the VC-dimension of V_d is at least d let S be an independent set of d vectors in $\mathbb{F}^n$. For every $A\subseteq S$ we can *separate* A with $\text{span}\left(A\right)$, which is of dimension at most d. It follows that an independent set of d vectors in $\mathbb{F}^n$ is *shattered* by V_d.

Now we show that the VC-dimension of V_d is at most d. Assume by contradiction that there exists a set S of cardinality larger than d that is shattered by V_d. If $\dim\left(S\right)>d$ then no $\mathbf{v}\in V_d$ can separate all of S, i.e., $S\subseteq\mathbf{v}$ as otherwise $\dim\left(\mathbf{v}\right)\geq\dim\left(S\right)>d$. Thus assume $\dim\left(S\right)\leq d<|S|$. Hence there exists a set of points $A\subset S$ and a point $p'\in S\setminus A$ such that $p'\in\text{span}\left(A\right)$. But then for any $\mathbf{v}\in V_d$ if $A\subseteq\mathbf{v}$ then also $p'\in\mathbf{v}$, this means that A cannot be separated, a contradiction. □

4.5 Proof of Theorem 1.3

Let $W = \{w_1, w_2, ..., w_m\}, \Delta, \epsilon$ be as in Theorem 1.3. Let $\tilde{\mathbf{v}}$ be a subspace that attains

$$\epsilon = \min_{\mathbf{v}:\dim(\mathbf{v})=d} \mathop{\mathbb{E}}_{i\in[m]} [\Delta(w_i, \mathbf{v})] .$$

Define $\tilde{W} = \{\tilde{w}_1, \tilde{w}_2, ..., \tilde{w}_m\}$ to be such that $\tilde{w}_i \in \tilde{\mathbf{v}}$ is the closest vector to w_i, under distance Δ.

Theorem 4.13. *[Restatement of Theorem 1.3]. For any $\delta > 0$, with probability $1-\delta$ (for a random set S of k samples from $[m]$, where each sample set is drawn independently and uniformly), for every k-dimensional subspace $\boldsymbol{v}$ of $\mathbb{R}^n$,*

$$\mathop{\mathbb{E}}_{i\in[m]} [\Delta(w_i, \boldsymbol{v})] \leq \mathop{\mathbb{E}}_{i\in S} [\Delta(w_i, \boldsymbol{v})] + 2\sqrt{\frac{2d^2 \log \frac{ek}{d^2}}{k}} + 2\epsilon + \sqrt{\frac{\log \frac{2}{\delta}}{2k}} .$$

Proof of Theorem 4.13. Denote $W_S = \{w_i : i \in S\}$ and $\tilde{W}_S = \{\tilde{w}_i : i \in S\}$, and recall Definition 4.11. In order to use the generalization bounds presented above in Theorem 4.3, we bound $\mathbb{E}_S [\Phi(V_{d'}, \Delta, W, W_S)]$. In order to do so we first present a series of claims. The first claim shows, roughly speaking, that we can analyze as if the sampled points were from $\tilde{W}$ rather than from W.

Claim 4.14. $\mathbb{E}_S [\Phi(V_{d'}, \Delta, W, W_S)] \leq \mathbb{E}_S \left[\Phi\left(V_{d'}, \Delta, \tilde{W}, \tilde{W}_S\right)\right] + 2\epsilon.$

Proof. Fix S. Then by the triangle inequality,

$$\begin{aligned}
\Phi(V_{d'}, \Delta, W, W_S) &= \sup_{\mathbf{v}\in V_{d'}} \left(\mathbb{E}[\Delta(w_i, \mathbf{v})] - \hat{\mathbb{E}}_S[\Delta(w_i, \mathbf{v})]\right) \qquad (3)\\
&\leq \sup_{\mathbf{v}\in V_{d'}} \left(\mathop{\mathbb{E}}_{i\in[m]} \left[\Delta(w_i, \tilde{w}_i) + \Delta(\tilde{w}_i, \mathbf{v})\right] - \mathop{\mathbb{E}}_{i\in S} \left[\Delta(\tilde{w}_i, \mathbf{v}) - \Delta(\tilde{w}_i, w_i)\right]\right)\\
&= \sup_{\mathbf{v}\in V_{d'}} \left(\mathop{\mathbb{E}}_{i\in[m]} [\Delta(\tilde{w}_i, \mathbf{v})] - \mathop{\mathbb{E}}_{i\in S} [\Delta(\tilde{w}_i, \mathbf{v})]\right) + \mathop{\mathbb{E}}_{i\in[m]} [\Delta(w_i, \tilde{w}_i)] + \mathop{\mathbb{E}}_{i\in S} [\Delta(\tilde{w}_i, w_i)] .
\end{aligned}$$

Note that

$$E_S \left[\mathop{\mathbb{E}}_{i\in[m]} [\Delta(w_i, \tilde{w}_i)] + \mathop{\mathbb{E}}_{i\in S} [\Delta(\tilde{w}_i, w_i)]\right] = 2 \mathop{\mathbb{E}}_{i\in[m]} [\Delta(w_i, \tilde{w}_i)] \leq 2\epsilon.$$

Taking the expectation of (3) over S completes the proof. □

For $\mathbf{v}$ a d'-dimensional vector subspace of $\mathbb{R}^n$ and $t \in [0, 1]$, define

$$\mathbf{v}_t := \{x \in \mathbb{R}^n | \Delta(\mathbf{v}, x) > t\} .$$

Define the following hypothesis set

$$V'_{d',n} = \left\{\mathbf{v}_t | \ \mathbf{v} \text{ is a } d'\text{-dimensional vector subspace of } \mathbb{R}^n, \ t \in [0, 1]\right\} ,$$

when the second subscript n is clear from the context we omit it.

The following claim shows that we can analyze a simpler loss function, the $0-1$ loss function rather than Δ, at the price of analyzing a more complex, VC-dimension wise, hypothesis set, $V'_{d'}$ rather than $V_{d'}$. The benefit of analyzing the $0-1$ loss function is that the VC-dimension generalization bound applies in the $0-1$ loss function case. We note that even in the case of $d'=2$ one can show that the VC-dimension of V'_2 is at least n as opposed to the VC-dimension of V_2, which by Theorem 4.12 is 2.

Claim 4.15. $\mathbb{E}_{S\subset[m]}\left[\Phi\left(V_{d'},\Delta,\tilde{W},\tilde{W}_S\right)\right]\leq\mathbb{E}_{S\subset[m]}\left[\Phi\left(V'_{d'},\{\mathbf{1}_{\boldsymbol{v}_t}\}_{\boldsymbol{v}_t\in V'_{d'}},\tilde{W},\tilde{W}_S\right)\right]$

Proof. A proof with similar arguments (albeit with a different statement) can be found in [MRT12, Chapter 10], and we provide the proof for completeness. Fix $S\in[m]$, then

$$\begin{aligned}
\Phi\left(V_{d'},\Delta,\tilde{W},S\right) &= \sup_{\mathbf{v}\in v_{d'}}\left(\mathop{\mathbb{E}}_{i\in[m]}\left[\Delta(\tilde{w}_i,\mathbf{v})\right]-\mathop{\mathbb{E}}_{i\in S}\left[\Delta(\tilde{w}_i,\mathbf{v})\right]\right)\\
&= \sup_{\mathbf{v}\in V_{d'}}\int_{t=0}^{1}\left(\Pr_{i\in[m]}\left[\Delta(\tilde{w}_i,\mathbf{v})>t\right]-\Pr_{i\in S}\left[\Delta(\tilde{w}_i,\mathbf{v})>t\right]\right)dt,\\
&\leq \sup_{\mathbf{v}\in V_{d'},t\in[0,1]}\left(\Pr_{i\in[m]}\left[\Delta(\tilde{w}_i,\mathbf{v})>t\right]-\Pr_{i\in S}\left[\Delta(\tilde{w}_i,\mathbf{v})>t\right]\right)\\
&= \sup_{\mathbf{v}\in V_{d'},t\in[0,1]}\left(\mathop{\mathbb{E}}_{i\in[m]}\left[\mathbf{1}_{\Delta(\tilde{w}_i,\mathbf{v})>t}\right]-\mathop{\mathbb{E}}_{i\in S}\left[\mathbf{1}_{\Delta(\tilde{w}_i,\mathbf{v})>t}\right]\right)\\
&= \Phi\left(V'_{d'},\{\mathbf{1}_{\mathbf{v}_t}\}_{\mathbf{v}_t\in V'_{d'}},\tilde{W},\tilde{W}_S\right).
\end{aligned}$$

where the second equality follows by the identity $\mathbb{E}[X]=\int_{t=0}^{\infty}\Pr[x>t]\,dt$ which holds for any positive random variable, the proof follows.

□

Define the hypothesis set of quadratic polynomials.

$$QP_n:=\left\{\{x\in\mathbb{R}^n\mid P(x)>0\}\;\mid\;P\text{ is an }n\text{-variate polynomial of degree }2\right\}.$$

Now we show an analysis of $V'_{d'}$. The analysis relies on the following geometric observation.

Claim 4.16. $V'_{d'}\subseteq QP_n$.

Proof. Let $\mathbf{v}_t\in V'_{d'}$. We show that $\mathbf{v}_t=\{x\in\mathbb{R}^n|\Delta(\mathbf{v},x)>t\}$ can be expressed as a quadratic polynomial strict inequality. Let $u'_1,...,u'_{d'}\in\mathbb{R}^n$ be an orthonormal basis for v and let $u_1,...,u_{n-d'}$ be an orthonormal basis for the orthogonal complement of $\mathbf{v}$,

$$v^{\perp}=\left\{x\in\mathbb{R}^n|\forall 0\leq i\leq d'\;\left\langle x,u'_i\right\rangle=0\right\}.$$

Define A_v to be the matrix with the i-th column equal u_i. It follows that $x\in\mathbf{v}_t$ if and only if

$$x^tA_v\left(A_v\right)^tx>t^2.$$

Note that the last inequality is a quadratic polynomial. □

The next definitions help to use the property that $\tilde{W}$ is contained in a d dimensional subspace of $\mathbb{R}^n$. Let $D = \text{span}(\tilde{W})$, as D is a subspace there exists $\psi : \mathbb{R}^n \to \mathbb{R}^d$, an isometry[6], (that is a linear map) with respect to D. Given any set or multi-set S of elements of D, not necessarily finite, we define $\psi(S)$ to be $\{\psi(s) \mid s \in S\}$. Let $\mathbf{H}$ be any hypothesis set on $\mathbb{R}^n$. Define a restriction operator as follows

$$Res_{\psi}(\mathbf{H}, D) =: \{\psi(h \cap D) \mid h \in \mathbf{H}\}.$$

Claim 4.17. *$Res_{\psi}(QP_n, D) = QP_d$.*

Proof. For simplicity assume D is such that $x \in D$ if and only if $(x)_i = 0$ for all $d+1 \leq i \leq n$. Use the following maping $\psi(y)_i = (y)_i$. Let

$$P = \sum_{1 \leq i \leq j \leq n} a_{i,j} x_i x_j ,$$

be a degree 2 polynomial such that $q = \{x \in \mathbb{R}^n | P(x) > 0\} \in QP_n$. It holds that

$$\psi(q) = \left\{ x \in \mathbb{R}^d \mid \sum_{1 \leq i \leq j \leq d} a_{i,j} x_i x_j > 0 \right\} \in QP_d .$$

Similar arguments proves for general D, details omitted. □

We note that it may be possible to present analysis directly on $V'_{d',n}$ but also note that $Res_{\psi}(V'_{d',n}, D)$ is a more complicated hypothesis set than $V'_{d',d}$. For example in the case where $n = 3$ and $d = 2$ it is easy to see that there are ellipsoids in $Res_{\psi}(V'_d, D)$, see Figure 1.

We will need the following fact, which can be found e.g. in [MRT12].

Lemma 4.18. *The VC-dimension of QP_d is at most d^2.*

We can now use the above to bound $\mathbb{E}_{S \subset [m]} [\Phi(V_k, \Delta, W, S)]$.

$$\begin{aligned}
\mathop{\mathbb{E}}_{S \subset [m]} [\Phi(V_{d'}, \Delta, W, W_S)] &\leq \mathop{\mathbb{E}}_{S \subset [m]} \left[\Phi\left(V_{d'}, \Delta, \tilde{W}, \tilde{W}_S\right) \right] + 2\epsilon && (4)\\
&\leq \mathop{\mathbb{E}}_{S \subset [m]} \left[\Phi\left(V'_{d'}, \{\mathbf{1}_{\mathbf{v}_t}\}_{\mathbf{v}_t \in V'_{d'}}, \tilde{W}, \tilde{W}_S\right) \right] + 2\epsilon && (5)\\
&\leq \mathop{\mathbb{E}}_{S \subset [m]} \left[\Phi\left(QP_n, \{\mathbf{1}_P\}_{P \in QP_n}, \tilde{W}, \tilde{W}_S\right) \right] + 2\epsilon && (6)\\
&= \mathop{\mathbb{E}}_{S \subset [m]} \left[\Phi\left(Res_{\psi}(QP_n, D), \{\mathbf{1}_P\}_{P \in Res_{\psi}(QP_n, D)}, \psi(\tilde{W}), \psi(\tilde{W}_S)\right) \right] + 2\epsilon && (7)\\
&= \mathop{\mathbb{E}}_{S \subset [m]} \left[\Phi\left(QP_d, \{\mathbf{1}_P\}_{P \in QP_d}, \psi(\tilde{W}), \psi(\tilde{W}_S)\right) \right] + 2\epsilon && (8)\\
&\leq 2\mathcal{R}_k\left(QP_d, \{\mathbf{1}_P\}_{P \in QP_d}\right) + 2\epsilon && (9)\\
&\leq 2\sqrt{\frac{2d^2 \log \frac{ek}{d^2}}{k}} + 2\epsilon. && (10)
\end{aligned}$$

[6] $\forall w_1, w_2 \in \mathbb{R}^n \quad \|\psi(w_1) - \psi(w_2)\|_2 = \|w_1 - w_2\|_2$

Here, Inequality 4 follows by Claim 4.14, Inequality 5 follows by Claim 4.15, Inequality 6 follows by Claim 4.16, Inequality 8 follows by Claim 4.17, Inequality 9 follows by Theorem 4.6 and Inequality 10 follows by Theorem 4.10 and Lemma 4.18.

Finally, we apply Theorem 4.3 and conclude that for any $\delta > 0$, with probability $1 - \delta$ over a sample set of k elements from $\mathbf{v}$,

$$\mathop{\mathbb{E}}_{i\in[m]}[\Delta(w_i, \mathbf{v})] \leq \mathop{\mathbb{E}}_{i\in S} \Delta(w_i, \mathbf{v}) + \sqrt{\frac{8d^2 \log \frac{ek}{d^2}}{k}} + 2\epsilon + \sqrt{\frac{\log \frac{2}{\delta}}{2k}}.$$

Note that we applied Theorem 4.3 for the hypothesis set V_k with respect to the loss function Δ. Other hypothesis sets were introduced only to bound Φ. □

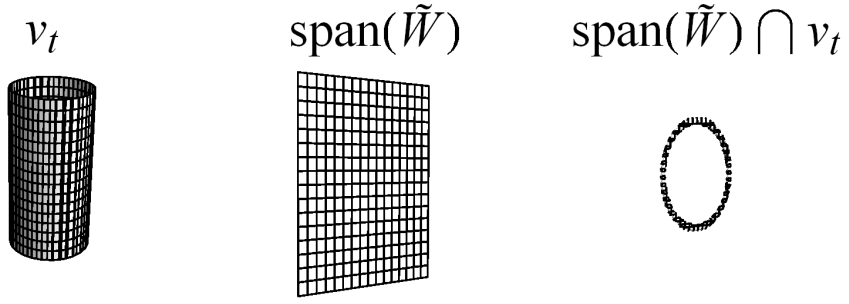

Figure 1: Ellipsoids in $Res_\psi(V'_d, D)$.

4.6 Proof of Theorem 1.4

Let $W = \{w_1, w_2, ..., w_m\}$ and ϵ be as in Theorem 1.4, and let $\mathbf{v}$ be a subspace that attains

$$\epsilon = \min_{\mathbf{v}:\dim(\mathbf{v})=d} \mathop{\mathbb{E}}_{i\in[m]}[\ell_2^2(w_i, \mathbf{v})].$$

Define $\tilde{W} = \{\tilde{w}_1, \tilde{w}_2, ..., \tilde{w}_m\}$ where each $\tilde{w}_i$ is the vector in $\mathbf{v}$ that has minimum ℓ_2^2 distance to w_i.

Theorem 4.19. *[Restatement of Theorem 1.4]. Let S be a set of k random samples, each drawn independently and uniformly from $[m]$. For every $\delta > 0$, with probability $1 - \delta$, for every k-dimensional subspace $\boldsymbol{v} \subset \mathbb{R}^n$,*

$$\mathop{\mathbb{E}}_{i\in[m]}[\ell_2^2(w_i, \boldsymbol{v})] \leq 8\sqrt{\epsilon} + \mathop{\mathbb{E}}_{i\in S}[\ell_2^2(w_i, \boldsymbol{v})] + 2\sqrt{\frac{2d^2 \log \frac{ek}{d^2}}{k}} + \sqrt{\frac{\log \frac{2}{\delta}}{2k}}.$$

The proof is similar to that of Theorem 1.3, except that we need the following variant of Claim 4.14 (one can check that the other claims apply also to ℓ_2^2). Denote $W_S = \{w_i : i \in S\}$ and $\tilde{W}_S = \{\tilde{w}_i : i \in S\}$, and recall Definition 4.11.

Claim 4.20. $\mathbb{E}_S\left[\Phi\left(V_{d'}, \ell_2^2, W, W_S\right)\right] \leq \mathbb{E}_S\left[\Phi\left(V_{d'}, \ell_2^2, \tilde{W}, \tilde{W}_S\right)\right] + 8\sqrt{\epsilon}$.

Proof. For every $w_i \in W$ and every subspace $\mathbf{v}$,

$$\ell_2^2(w_i, \mathbf{v}) = \min_{v \in \mathbf{v}} \|w_i - v\|_2^2.$$

The minimizer $v \in \mathbf{v}$ is an orthogonal projection of w_i, hence $\|v\|_2 \leq \|w_i\|_2$, and similarly $\|\tilde{w}_i\|_2 \leq \|w_i\|_2$. Writing

$$\begin{aligned}\|w_i - v\|_2^2 &= \|w_i - \tilde{w}_i + \tilde{w}_i - v\|_2^2 \\ &= \|w_i - \tilde{w}_i\|_2^2 + \|\tilde{w}_i - v\|_2^2 + 2\langle w_i - \tilde{w}_i, \tilde{w}_i - v\rangle,\end{aligned}$$

we now bound the inner-product using Cauchy-Schwarz, the triangle inequality, and our bounds from above, to get

$$\begin{aligned}|\langle w_i - \tilde{w}_i, \tilde{w}_i - v\rangle| &\leq \|w_i - \tilde{w}_i\|_2 \cdot \|\tilde{w}_i - v\|_2 \\ &\leq \|w_i - \tilde{w}_i\|_2 \cdot (\|\tilde{w}_i\|_2 + \|v\|_2) \\ &\leq \|w_i - \tilde{w}_i\|_2 \cdot 2\|w_i\|_2.\end{aligned}$$

Altogether, we obtain

$$\begin{aligned}\ell_2^2(w_i, \mathbf{v}) &\leq \|w_i - \tilde{w}_i\|_2^2 + \min_{v \in \mathbf{v}} \|\tilde{w}_i - v\|_2^2 + 4\|w_i - \tilde{w}_i\|_2 \cdot \|w_i\|_2, \\ \ell_2^2(w_i, \mathbf{v}) &\geq \|w_i - \tilde{w}_i\|_2^2 + \min_{v \in \mathbf{v}} \|\tilde{w}_i - v\|_2^2 - 4\|w_i - \tilde{w}_i\|_2 \cdot \|w_i\|_2.\end{aligned}$$

Using the last two inequalities, we can bound Φ for a fixed S by

$$\begin{aligned}\Phi\left(V_{d'}, \ell_2^2, W, W_S\right) &= \sup_{v \in V_{d'}} \left(E\left[\ell_2^2\left(w_i, v\right)\right] - \hat{E}_S\left[\ell_2^2\left(w_i, v\right)\right]\right) \\ &\leq \mathop{\mathbb{E}}_{i \in [m]} \|w_i - \tilde{w}_i\|_2^2 - \mathop{\mathbb{E}}_{i \in S} \|w_i - \tilde{w}_i\|_2^2 \\ &+ \sup_{v \in V_{d'}} \left(\mathop{\mathbb{E}}_{i \in [m]}\left[\min_{v \in \mathbf{v}} \|\tilde{w}_i - v\|_2^2\right] - \mathop{\mathbb{E}}_{i \in S}\left[\min_{v \in \mathbf{v}} \|\tilde{w}_i - v\|_2^2\right]\right) \\ &+ 4 \mathop{\mathbb{E}}_{i \in [m]}\left[\|w_i - \tilde{w}_i\|_2 \cdot \|w_i\|_2\right] + 4 \mathop{\mathbb{E}}_{i \in S}\left[\|w_i - \tilde{w}_i\|_2 \cdot \|w_i\|_2\right].\end{aligned}$$

Now take the expectation over S; then every expectation over $i \in S$, immediately becomes expectation over $i \in [m]$ (because $\mathbb{E}_S \mathbb{E}_{i \in S} f_i = \mathbb{E}_{i \in [m]} f_i$). In addition, using Cauchy-Schwartz,

$$\mathop{\mathbb{E}}_{i \in [m]}\left[\|w_i - \tilde{w}_i\|_2 \cdot \|w_i\|_2\right] \leq \left(\mathop{\mathbb{E}}_{i \in [m]} \|w_i - \tilde{w}_i\|_2^2 \cdot \mathop{\mathbb{E}}_{i \in [m]} \|w_i\|_2^2\right)^{1/2} \leq \sqrt{\epsilon}.$$

Altogether, $\mathbb{E}_S\left[\Phi(V_{d'}, \ell_2^2, W, W_S)\right] \leq \mathbb{E}_S\left[\Phi(V_{d'}, \ell_2^2, \tilde{W}, \tilde{W}_S)\right] + 8\sqrt{\epsilon}$. □

A Proof of Fact 3.1

For convenience we recall Fact 3.1.

Fact 3.1. *Let $M \in \mathbb{F}^{n\times m}$, and let $R \subset [n], C \subset [m]$ be of size $|R| = |C| = \mathrm{rank}(M)$. If $M_{R,C}$ is invertible,*[7] *then*

$$\forall (i,j) \in [n] \times [m], \qquad M_{i,j} = M_{i,C} \left(M_{R,C}\right)^{-1} M_{R,j}.$$

Proof. Since every matrix of rank d has a sub-matrix $A \in \mathbb{F}^{d\times d}$ of (full) rank d, the fact follows immediately from the following assertion: Let $A \in \mathbb{F}^{d\times d}$ be a matrix of rank d, let $x, y \in \mathbb{F}^d$ be a row vectors, let $z \in \mathbb{F}$, and suppose $A' = \begin{bmatrix} A & y \\ x^T & z \end{bmatrix}$ has rank d; then $z = x^T A^{-1} y$.

To prove this assertion, suppose towards contradiction that $z' \stackrel{\text{def}}{=} x^T A^{-1} y \neq z$. Then

$$\begin{aligned} \mathrm{rank}(A') &= \mathrm{rank}\left(\begin{bmatrix} A & y \\ x^T & z \end{bmatrix}\right) = \mathrm{rank}\left(\begin{bmatrix} A & y & y \\ x^T & z & z' \end{bmatrix}\right) \\ &\geq \mathrm{rank}\left(\begin{bmatrix} A & \vec{0} \\ x^T & z - z' \end{bmatrix}\right) = \mathrm{rank}\left(\begin{bmatrix} A & \vec{0} \\ \vec{0}^T & 1 \end{bmatrix}\right) = d + 1, \end{aligned}$$

a contradiction.

□

References

[ACCL08] N. Ailon, B. Chazelle, S. Comandur, and D. Liu. Property-preserving data reconstruction. *Algorithmica*, 51(2):160–182, April 2008. doi:10.1007/s00453-007-9075-9.

[Alo09] N. Alon. Perturbed identity matrices have high rank: Proof and applications. *Comb. Probab. Comput.*, 18(1-2):3–15, March 2009. doi:10.1017/S0963548307008917.

[ALSV13] N. Alon, T. Lee, A. Shraibman, and S. Vempala. The approximate rank of a matrix and its algorithmic applications: Approximate rank. In *Proceedings of the Forty-fifth Annual ACM Symposium on Theory of Computing*, STOC '13, pages 675–684. ACM, 2013. doi:10.1145/2488608.2488694.

[AT10] T. Austin and T. Tao. Testability and repair of hereditary hypergraph properties. *Random Struct. Algorithms*, 36(4):373–463, July 2010. doi:10.1002/rsa.v36:4.

[Cla05] K. L. Clarkson. Subgradient and sampling algorithms for l_1 regression. In *Proceedings of the sixteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 257–266. SIAM, 2005.

[CLMW11] E. J. Candès, X. Li, Y. Ma, and J. Wright. Robust principal component analysis? *J. ACM*, 58(3):11:1–11:37, June 2011. doi:10.1145/1970392.1970395.

[CSPW11] V. Chandrasekaran, S. Sanghavi, P. A. Parrilo, and A. S. Willsky. Rank-sparsity incoherence for matrix decomposition. *SIAM Journal on Optimization*, 21(2):572–596, 2011. doi:10.1137/090761793.

[DDG+15] R. David, I. Dinur, E. Goldenberg, G. Kindler, and I. Shinkar. Direct sum testing. In *Proceedings of the 2015 Conference on Innovations in Theoretical Computer Science, ITCS*, pages 327–336, 2015. doi:10.1145/2688073.2688078.

[Des07] A. J. Deshpande. *Sampling-based algorithms for dimension reduction*. PhD thesis, Massachusetts Institute of Technology, 2007. Available from: http://hdl.handle.net/1721.1/38935.

[7] Clearly, there always exist $R \subset [n], C \subset [m]$ of size $|R| = |C| = \mathrm{rank}(M)$ such that $M_{R,C}$ is invertible.

[DKM06] P. Drineas, R. Kannan, and M. W. Mahoney. Fast Monte Carlo algorithms for matrices II: Computing a low-rank approximation to a matrix. *SIAM J. Comput.*, 36(1):158–183, July 2006. doi:10.1137/S0097539704442696.

[DTV11] A. Deshpande, M. Tulsiani, and N. K. Vishnoi. Algorithms and hardness for subspace approximation. In *Proceedings of the twenty-second annual ACM-SIAM symposium on Discrete Algorithms*, pages 482–496. SIAM, 2011. doi:10.1137/1.9781611973082.39.

[DV07] A. Deshpande and K. Varadarajan. Sampling-based dimension reduction for subspace approximation. In *Proceedings of the thirty-ninth annual ACM symposium on Theory of computing*, pages 641–650. ACM, 2007.

[Fei14] U. Feige. NP-hardness of hypercube 2-segmentation. *CoRR*, abs/1411.0821, 2014. arXiv:1411.0821.

[FK99] A. Frieze and R. Kannan. Quick approximation to matrices and applications. *Combinatorica*, 19(2):175–220, 1999. doi:10.1007/s004930050052.

[FKV04] A. Frieze, R. Kannan, and S. Vempala. Fast monte-carlo algorithms for finding low-rank approximations. *J. ACM*, 51(6):1025–1041, November 2004. doi:10.1145/1039488.1039494.

[GL89] O. Goldreich and L. A. Levin. A hard-core predicate for all one-way functions. In *Proceedings of the Twenty-first Annual ACM Symposium on Theory of Computing*, STOC '89, pages 25–32. ACM, 1989. doi:10.1145/73007.73010.

[HP06] S. Har-Peled. How to get close to the median shape. In *Proceedings of the twenty-second annual symposium on Computational geometry*, pages 402–410. ACM, 2006.

[KS03] R. Krauthgamer and O. Sasson. Property testing of data dimensionality. In *14th Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 18–27, January 2003.

[LWW14] Y. Li, Z. Wang, and D. P. Woodruff. Improved testing of low rank matrices. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '14, pages 691–700. ACM, 2014. doi:10.1145/2623330.2623736.

[MRT12] M. Mohri, A. Rostamizadeh, and A. Talwalkar. *Foundations Of Machine Learning*. MIT Press, 2012.

[SS10] M. Saks and C. Seshadhri. Local monotonicity reconstruction. *SIAM J. Comput.*, 39(7):2897–2926, May 2010. doi:10.1137/080728561.

[SV07] N. D. Shyamalkumar and K. Varadarajan. Efficient subspace approximation algorithms. In *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '07, pages 532–540. SIAM, 2007.

[Val77] L. G. Valiant. Graph-theoretic arguments in low-level complexity. In *Mathematical foundations of computer science*, Lecture Notes in Computer Sciece, pages 162–176. Springer, 1977.

[XCM13] H. Xu, C. Caramanis, and S. Mannor. Outlier-robust PCA: the high-dimensional case. *IEEE Transactions on Information Theory*, 59(1):546–572, 2013. doi:10.1109/TIT.2012.2212415.

ECCC ISSN 1433-8092

https://eccc.weizmann.ac.il