

Matrix Decompositions and Sparse Graph Regularity*

Greg Bodwin and Santosh Vempala

Georgia Tech

Abstract

We introduce and study a matrix decomposition that is a common generalization of the singular value decomposition (SVD), cut decomposition, CUR decomposition, and others. For any given set of pairs $P \subseteq \mathbb{R}^m \times \mathbb{R}^n$ and matrix $A \in \mathbb{R}^{m \times n}$, we write A as a weighted sum of rank one matrices formed by some pairs in P . The resulting *projection value decomposition (PVD)* inherits several useful properties of the SVD; for example, the decomposition can be obtained by greedily peeling off rank one matrices, and the norm of the coefficients of the decomposition is the Frobenius norm of the matrix. Perhaps most interesting is that, in analogy with low-rank approximation from SVD, truncating the decomposition gives matrix approximations of small error.

When applied to the adjacency matrices of graphs, the PVD lets us derive the weak regularity lemma of Frieze and Kannan as well as Szemerédi’s strong regularity lemma. Whereas these regularity lemmas in their usual forms are nontrivial only for dense graphs on $\Omega(n^2)$ edges, our generalization implies extensions to a new class of sparse graphs which we call *cut pseudorandom*, which are roughly those with small leading coefficients in the appropriate PVD. It turns out that cut pseudorandomness unifies several important pseudorandomness concepts in prior work: we show that L_p upper regularity and a version of low threshold rank are both special cases, thus implying weak and strong regularity lemmas for these graph classes where only weak ones were previously known.

*Supported in part by NSF awards CCF-1717349, DMS-183932 and CCF-1909756.
 {gregory.bodwin,vempala}@cc.gatech.edu.

1 Introduction

Matrix decompositions are, broadly, methods to write a matrix as a linear combination of rank one matrices in ways that reveals something about its structure. The most famous example is the singular value decomposition (SVD), where for a rank r matrix $A \in \mathbb{R}^{m \times n}$ we write

$$A = \sigma_1 v_1 w_1^T + \cdots + \sigma_r v_r w_r^T, \quad (1)$$

greedily choosing singular values $\{\sigma_i\}$ and singular vector pairs $\{(v_i, w_i)\}$ to capture the most possible remaining mass of A . These choices then carry lots of well-known structural information about the matrix. One useful fact is that the first t terms in the sum give the best possible rank- t approximation to A in both the Frobenius norm and the spectral norm. This is fundamental in the area of *low-rank approximation* and has applications to principal component analysis (PCA), image processing, clustering, machine learning, etc. (see e.g., [6, 24]).

In some of these applications, what one really wants is an equation like (1) with restrictions on the values $\{v_i\}, \{w_i\}$ can take. This arises commonly in PCA, for example, where the matrix A holds data that might come in a restricted form (the *age* field must be a nonnegative integer, etc). It is often desirable to have the $\{v_i\}, \{w_i\}$ fit this form in order to be directly interpretable as data [30, 15]. The *CUR decomposition* [14, 25] addresses this in a specific way, essentially by requiring the $\{v_i\}$ s to be columns of A and the $\{w_i\}$ s to be rows, but it is also natural to consider a decomposition where $\{v_i\}, \{w_i\}$ can be any plausible columns, rows of data. Another place that restricted $\{v_i\}, \{w_i\}$ arise is in *graph regularity* and *graph limits* [28], where the $\{v_i\}, \{w_i\}$ must be binary to correspond to indicator vectors of cuts of a graph. A seminal work of Frieze and Kannan [18] introduced the corresponding *cut decomposition*. There are several other examples where particular $\{v_i\}, \{w_i\}$ restrictions are useful (we will mention some later).

This paper abstracts the problem of decomposing matrices into a linear combination of rank one matrices drawn from an arbitrarily restricted domain. Our main result is that a natural analog of low-rank approximation holds *in general* in any restricted domain, even if that domain is nonlinear or discrete. With appropriate choices of domain, this fact simultaneously generalizes low-rank approximation, cut approximation, and others, and it gives a new version of the CUR approximation (with a different type of error guarantee than usual). Much like the SVD is behind low-rank approximation, and cut decomposition is behind cut approximation, our generalized low-rank approximation is proved via a new *projection value decomposition (PVD)* which simultaneously generalizes these and seems of independent interest. The PVD has many properties analogous to the SVD, including a notion of *projection values* that inherit some important properties of singular values.

Our main application comes by applying the PVD to the adjacency matrices of graphs, which yields new sparse versions of the weak [18] and Szemerédi [34] regularity lemmas for a new class of input graphs that we call “cut pseudorandom.” Roughly, a graph is cut pseudorandom if the leading coefficients of the appropriate PVD are sufficiently small. Cut pseudorandomness helps unify and explain several important pseudorandomness concepts in prior work; in particular, we prove that L_p upper regularity [7, 8] which is fundamental in the theory of graph limits, and an unnormalized version of low threshold rank [5, 21, 22, 23, 20], a basic parameter in optimization, are both special cases. Thus we derive weak and strong (Szemerédi) regularity lemmas for these graph classes as corollaries; the weak ones were already known [7, 20], but the strong ones are new.

1.1 The Projection Value Decomposition

The *projection value decomposition (PVD)* of a matrix $A \in \mathbb{R}^{m \times n}$, with respect to a “pair set” $P \subseteq \mathbb{R}^m \times \mathbb{R}^n$, is obtained by the following greedy process (see Section 1.1 for more formality): in each step we choose the pair $(v, w) \in P$ maximizing the quantity

$$\frac{v^T A w}{\|v\|_2 \|w\|_2}$$

and then we remove the component of A with maximum Frobenius norm that can be expressed as a linear combination of the matrices $\{vw^T\}$, for all pairs (v, w) chosen so far. The *projection values* $\{\sigma_i\}$ are the Frobenius norms of the removed components. The name comes from the fact that we can

equivalently write the PVD by the equation

$$\Pi_P(A) = \sum_{i=1}^{mn} \Pi_{P^{(i)}}(A) - \Pi_{P^{(i-1)}}(A),$$

where $\Pi_P(A)$ is the projection of A onto the span of the rank 1 matrices made from pairs in P , and $P^{(i)}$ is the length i prefix of P with respect to the “greedy” ordering sketched above. The inner matrices in this sum all have rank 1, and the projection values are their Frobenius norms. We also note that it is natural to have $\Pi_P(A)$ on the left; P should be interpreted as a restriction of the domain, so matrices are only really defined up to their behavior on P .

In the special case $P = \mathbb{R}^m \times \mathbb{R}^n$ the PVD is exactly the SVD, and the projection values are the singular values, but in general it can be quite different when other pair sets are considered. Despite this, we show that several important properties of the SVD generalize to the PVD anyways: for example, the projection values are decreasing, and their L_2 norm is still exactly the Frobenius norm of the input matrix. In fact, it turns out that all the machinery used to get low-rank approximation via SVD generalizes to the PVD, allowing us to prove the following extension:

Theorem 1 (Main Result). *For a pair set $P \subseteq \mathbb{R}^m \times \mathbb{R}^n$, let*

$$\|A\|_P := \max_{(v,w) \in P | v,w \neq 0} \frac{v^T A w}{\|v\|_2 \|w\|_2}.$$

Then for any matrix $A \in \mathbb{R}^{m \times n}$, nonnegative integer r , and $k \geq 2$, letting ς_r be the first r projection values and $\hat{A} \in \mathbb{R}^{m \times n}$ the sum of the first r terms in the PVD, we have

$$\|A - \hat{A}\|_P \leq \frac{\|\varsigma_r\|_k}{(r+1)^{1/k}} \leq \frac{\|A\|_{F^{(k)}}}{(r+1)^{1/k}}$$

where $\|A\|_{F^{(k)}}$ is the so-called Schatten k -norm of A , which is the L_k norm of its singular values. (In particular, this means $\hat{A}$ is a linear combination of $\leq r$ rank 1 matrices of the form vw^T , $(v,w) \in P$.)

The main conceptual contribution of Theorem 1 is the use of restricted pair set P , which can be nonlinear or discrete (we emphasize that we are not assuming that it is a subspace of $\mathbb{R}^m \times \mathbb{R}^n$). Low-rank approximation by SVD corresponds exactly to the case $P = \mathbb{R}^m \times \mathbb{R}^n$. Low-rank approximation has the guarantee that $\hat{A}$ has rank r ; Theorem 1 uses the nonlinear generalization that $\hat{A}$ can be written as the sum of r outer products from the domain. Low-rank approximation also has the guarantee that $\|A - \hat{A}\|_2$ is small; Theorem 1 uses the nonlinear generalization that $A - \hat{A}$ essentially has small operator norm over the given domain. Thus, relative to low-rank approximation, Theorem 1 can be viewed as a tradeoff of stronger constraints on $\hat{A}$ for a weaker approximation guarantee. Various choices of P (plus some light algebra) give the following approximation theorems as special cases:

Corollary 2 (SVD). *For any $A \in \mathbb{R}^{m \times n}$, nonnegative integer r , and $k \geq 2$, there is a rank r matrix $\hat{A} \in \mathbb{R}^{m \times n}$ satisfying*

$$\|A - \hat{A}\|_2 \leq \frac{\|A\|_{F^{(k)}}}{(r+1)^{1/k}}.$$

Corollary 3 (CUR Decomposition). *For any $A \in \mathbb{R}^{m \times n}$, nonnegative integer r , and $k \geq 2$, there is $\hat{A} \in \mathbb{R}^{m \times n}$ that is a linear combination of r outer products of a column of A and a row of A , such that*

$$\left\| A (A - \hat{A})^T A \right\|_\infty \leq \|A\|_{F^{(k)}} \frac{c^* r^*}{(r+1)^{1/k}}$$

where c^*, r^* are the maximum lengths (L_2 norm) of a column, row in A and $\|\cdot\|_\infty$ denotes the maximum entry magnitude.

Corollary 4 (Cut Decomposition [18]). *For any $A \in \mathbb{R}^{m \times n}$, nonnegative integer r , and positive odd integer k , there is $\hat{A} \in \mathbb{R}^{m \times n}$, formed as a linear combination of r cut matrices (which are constant on an $S \times T$ block and zero elsewhere), such that*

$$\|A - \hat{A}\|_\square \leq \sqrt{mn} \frac{\|A\|_{F^{(k)}}}{(r+1)^{1/k}}$$

where $\|\cdot\|_{\square}$ is the cut norm, i.e.

$$\|A\|_{\square} := \max_{S \subseteq [m], T \subseteq [n]} \left| \sum_{i \in S, j \in T} A_{ij} \right|.$$

In Section 3.2 we further observe that the PVD, and in particular an analog of Theorem 1, extends to tensors. This recovers an additional tensor decomposition theorem in [13], which gives a polynomial time approximation scheme (PTAS) for certain maximum constraint satisfaction problems [13].

1.2 Applications to Graph Regularity

A major application of the cut decomposition is to *graph regularity*, where the high-level goal is to approximate an arbitrary graph by the union of a few (bipartite) random graphs. The area was pioneered by a classic lemma of Szemerédi [34], which guarantees the existence of a good approximation for any sufficiently dense input graph. This Szemerédi regularity lemma plays a fundamental role in modern combinatorics with many critical applications. It has also spawned a line of work on regularity lemmas themselves; notable examples include the weak regularity lemma of Frieze and Kannan [18] and the strong regularity lemma of Alon, Fischer, Krivelevich and Szegedy [2], among others. These lie at the center of a deep mathematical theory of graph limits [3, 28].

In their original forms these regularity lemmas are nontrivial only for *dense graphs*, meaning n -node graphs on $\Omega(n^2)$ edges. The subarea of sparse regularity tries to prove nontrivial extensions of these lemmas to *sparse graphs* on $o(n^2)$ edges. There are usually lower bounds stating that no extension is possible that holds for *every* sparse graph, so instead the goal is to define broad classes of sparse graphs and prove regularity lemmas for these classes specifically. Most often these classes are defined by a statistic of the graph required to be “close” to its expectation in a random graph, and so sometimes these graph classes are considered *pseudorandomness concepts*.

Let us next describe these ideas in more formality. Let $G = (V, E)$ be a graph and let $\Pi = \{V_1, \dots, V_q\}$ be a partition of its vertices into q parts. We define two measures of “irregularity,” which are essentially two different ways to quantify the amount the induced subgraphs between parts in Π look like random graphs. For vertex subsets $S, T \subseteq V$ let $e(S, T)$ count the number of edges with one endpoint in S and the other in T , and let their *density* be

$$d(S, T) := \frac{e(S, T)}{|S||T|}.$$

We then define:

$$\begin{aligned} \text{weakirr}(\Pi) &:= \max_{S, T \subseteq V} \left| e(S, T) - \sum_{V_i, V_j \in \Pi} |S \cap V_i| |T \cap V_j| d(V_i, V_j) \right|. \\ \text{szémirr}(\Pi) &:= \sum_{V_i, V_j \in \Pi} \max_{S \subseteq V_i, T \subseteq V_j} |e(S, T) - |S||T|d(V_i, V_j)|. \end{aligned}$$

For now it is not too important exactly how to interpret these definitions, just that they are both the natural irregularity measure for various applications. Corresponding to these are two (known) regularity lemmas:

Theorem 5 (Weak Regularity Lemma [18]). *For any $\varepsilon > 0$ and n -node graph $G = (V, E)$, there is a vertex partition Π into $|\Pi| = 2^{O(\varepsilon^{-2})}$ parts satisfying $\text{weakirr}(\Pi) \leq \varepsilon n^2$.*

Theorem 6 (Szemerédi Regularity Lemma [34, 29]). *For any $\varepsilon > 0$ and n -node graph $G = (V, E)$, there is a vertex partition Π into $|\Pi| = \text{tower}(O(\varepsilon^{-2}))$ parts¹ satisfying $\text{szémirr}(\Pi) \leq \varepsilon n^2$.*

For either type of irregularity, an easy counting argument shows that it is always $\leq |E|$ (for any vertex partition Π). Thus, the irregularity bounds of $\leq \varepsilon n^2$ promised by Theorems 5 and 6 are nontrivial only when $|E|$ is basically quadratic in n to begin with. A version that is nontrivial for sparser graphs would need an improved irregularity control of $\leq \varepsilon |E|$. To obtain this improvement, we propose the following pseudorandomness concept:

¹Recall that $\text{tower}(x) = 2^{2^{\dots}}$ with total height x .

Definition 1 (Cut Pseudorandomness). *For an n -node graph $G = (V, E)$ with adjacency matrix A , we say that G is (k, r, C) cut pseudorandom if, relative to P as all pairs of cut vectors in $\mathbb{R}^n$, the leading r projection values $\varsigma_r := (\sigma_1, \dots, \sigma_r)$ of A satisfy*

$$\|\varsigma_r\|_k \leq C \frac{|E|}{n}.$$

We will usually omit the parameter C to mean that G is cut pseudorandom for some absolute constant $C > 0$ (not depending on k, r, n , or $|E|$).

With this, we prove two new results, which constitute our main applications of Theorem 1. As we will see presently, the first theorem recovers weak regularity lemmas in the literature for important special cases, while the second provides strong (Szemerédi) regularity which is novel and of interest even in the special cases.

Theorem 7. *For all $\varepsilon > 0$ and $k \geq 2$, there is $r = O(\varepsilon^{-k})$ such that for any graph $G = (V, E)$ that is (k, r) cut pseudorandom, there is a vertex partition Π into $|\Pi| = 2^{O(\varepsilon^{-k})}$ parts satisfying*

$$\text{weakirr}(\Pi) \leq \varepsilon |E|.$$

Theorem 8. *For all $\varepsilon > 0$, there is $r = \text{tower}(O(\varepsilon^{-2}))$ such that for any graph $G = (V, E)$ that is $(2, r)$ cut pseudorandom, there is a vertex partition Π into $|\Pi| = \text{tower}(O(\varepsilon^{-2}))$ parts satisfying*

$$\text{szémirr}(\Pi) \leq \varepsilon |E|.$$

Note the technical difference that weak regularity holds for (k, r) cut pseudorandom graphs, while strong regularity requires $k = 2$. Some basic examples of cut pseudorandom graphs include:

- An Erdős-Rényi random graph on $\Omega(n^{1+1/k})$ edges, for a positive integer k , is (k, r) cut pseudorandom (for any r , with very high probability).
- All graphs on $\Omega(n^2)$ edges are $(2, r)$ cut pseudorandom.
- More generally, for any integer k , any graph on $\Omega(n^{1+1/k})$ edges with girth (shortest cycle length) $> 2k$ is (k, r) cut pseudorandom. Such graphs are only known to exist for $k \in \{1, 2, 3, 5\}$ [37, 36]; it is a popular but unresolved Erdős conjecture that they exist for other k [16]. We note that any graph with girth $> 2k$ can have only $O(n^{1+1/k})$ edges (folklore), so these graphs would exist only in a narrow density range.

In all the above examples, the cut pseudorandomness condition is satisfied even for the full vector of projection values; in fact, more strongly, the Schatten k -norm of A is bounded by $O(|E|/n)$. One way to see this is to observe that the sum of the $2k^{\text{th}}$ powers of all singular values of the adjacency matrix of a graph is equal to the number of closed $2k$ -walks in the graph, and the latter can be bounded for random graphs as $O(nd^k + d^{2k})$ where d is the average degree. In what follows we will see sparse regularity theorems for more general graph classes where we will need to use the fact that it suffices to bound the norm of the first r projection values only (rather than the full vector of projection values).

Let us give some more context on the intuitive meaning of cut pseudorandomness. For analogy, in low-rank approximation by the SVD, the quality of a rank- r approximation is controlled by the top singular value of $A - \hat{A}$, which is exactly the $(r+1)^{\text{th}}$ singular value σ_{r+1} of A . Instead of using σ_{r+1} directly, though, it is common (as in Corollary 2) to in turn control σ_{r+1} by observing that the singular values are decreasing, and therefore

$$\sigma_{r+1} \leq \frac{\|\langle \sigma_1, \dots, \sigma_{r+1} \rangle\|_k}{(r+1)^{1/k}}.$$

Similarly, for cut approximation by the PVD, the quality of a length r approximation is controlled by the $(r+1)^{\text{th}}$ projection value σ_{r+1} . These values are also decreasing, so again we can use

$$\sigma_{r+1} \leq \frac{\|\varsigma_{r+1}\|_k}{(r+1)^{1/k}}.$$

A (k, r) cut pseudorandom graphs is exactly one that admits this kind of L_k -certificate on the r^{th} projection value (to a certain quality). This obviously loses some generality, as there exist graphs where σ_{r+1} is suitably bounded but $\|\varsigma_{r+1}\|_k$ is not. But in a way, that's our point: we go on to show that many important pseudorandomness concepts in prior work are special cases of cut pseudorandomness, even with this loss of generality. Thus we are able to unify this prior work with a single pseudorandomness concept while simultaneously clarifying its limitations. We also push it a bit further: for the two pseudorandomness concepts we will discuss, sparse weak regularity lemmas were known, but Theorem 8 gives sparse Szemerédi regularity lemmas, which are new and of interest.

One important pseudorandomness concept we will cover is L_k upper regularity, developed in a pair of twin papers [7, 8] by Borgs, Chayes, Cohn, and Zhao:

Definition 2 (L_k Upper Regularity [7, 8]). *An n -node graph $G = (V, E)$ is (η, C) upper L_k regular if, for any vertex partition $\Pi = \{V_i\}_{i=1}^p$ into $p \leq \eta^{-1}$ parts,² we have*

$$\left(\sum_{V_i, V_j \in \Pi} \frac{|V_i||V_j|}{n^2} d(V_i, V_j)^k \right)^{1/k} \leq Cd(G)$$

where $d(G) = d(V, V)$ is the overall fraction of possible edges that are present in G . As before, we will omit the parameter C to mean that $C > 0$ is an absolute constant.

The case $k = \infty$ (which amends the above “ L_k norm” of densities to a max in the usual way) was studied first by Kohayakawa [26] and Rödl (unpublished), who obtained a sparse Szemerédi regularity lemma for these graphs:

Theorem 9 (Rephrased from [26]). *For all $\varepsilon > 0$, $C > 0$ there is $\eta > 0$ such that any (η, C) upper L_∞ regular graph G has a vertex partition Π into $|\Pi| = \text{tower}(\varepsilon^{-2})$ parts satisfying*

$$\text{szémirr}(\Pi) \leq \varepsilon|E|.$$

Many applications of the dense Szemerédi regularity lemma extend nicely to L_∞ upper regular graphs via this theorem (see surveys [19, 12]). In particular, the application of *graph limits* inspired the generalized definition of L_k upper regularity; Borgs et. al. [7, 8] layed out a comprehensive case that L_k upper regular graphs are perhaps the natural class in which to conduct limit theories. This included a *weak* regularity lemma for these graphs:

Theorem 10 (Rephrased from [7, 8]). *For all $\varepsilon > 0, k > 1, C > 0$, there is $\eta > 0$ such that any (η, C) upper L_k regular graph G has a vertex partition Π into $|\Pi| = 2^{\varepsilon^{-2}}$ parts satisfying*

$$\text{weakirr}(\Pi) \leq \varepsilon|E|.$$

We show:

Lemma 11. *For every positive integer r and $k > 1$ (or $k = \infty$), there is $\eta = 2^{O_k(-r)}$ such that any L_k upper η -regular graph G is $(2, r)$ cut pseudorandom.*

This Theorems 7 and 8 both apply to this graph class; the former was known in [7], while the latter was only known for $k = \infty$. A similar story holds for graphs of *low threshold rank* [5, 21, 22, 23, 20], which are roughly those whose “normalized” adjacency matrix only has a few large eigenvalues. We will work with the following closely related definition:

Definition 3 (Unnormalized Threshold Rank). *The unnormalized ε -threshold rank $t_\varepsilon(G)$ of an n -node graph $G = (V, E)$ with adjacency matrix eigenvalues $\{\lambda_1, \dots, \lambda_n\}$ is*

$$t_\varepsilon(G) := \sum_{\lambda_i > \varepsilon|E|/n} \left(\lambda_i \cdot \frac{n}{|E|} \right)^2.$$

²The inversion of η here is to maintain the parameter conventions of [7, 8]. Their definition requires every part in Π to have size $\geq \eta n$; this definition is equivalent up to constant factors in η , since we can merge smaller parts and control their contribution with a union bound.

This definition coincides with (normalized) threshold rank for degree-regular graphs, but differs a little for others (the name “threshold rank” makes much more sense in this normalized context, but we keep it for consistency). In [20], Oveis Gharan and Trevisan prove a weak regularity lemma for graphs of low (normalized) threshold rank, which can be easily adapted to the unnormalized setting:

Theorem 12 ([20]). *For all $\varepsilon > 0$, every graph $G = (V, E)$ of unnormalized ε -threshold rank $t_\varepsilon(G) = O(1)$ has a vertex partition into $|\Pi| = 2^{O(\varepsilon^{-2})}$ parts satisfying*

$$\text{weakirr}(\Pi) \leq \varepsilon|E|.$$

We show:

Lemma 13. *For every positive integer r and $k > 1$, there is $\varepsilon = O_k(r^{-1/2})$ such that any n -node graph $G = (V, E)$ with unnormalized threshold rank $t_\varepsilon(G) = O(1)$ is (k, r) cut pseudorandom.*

This recovers the above unnormalized weak regularity lemma of [20], and implies a new Szemerédi regularity lemma for these graphs as well. An application of the weak regularity lemma in [20] is a PTAS for MAX-CUT in input graphs of low threshold rank. For this reason (and intrinsic reasons), it is useful to have *algorithmic* versions of these regularity lemmas, and accordingly, algorithmic versions of the PVD. This is nontrivial: in fact, as described above, the greedy ordering of the set of cut vectors would require one to solve (normalized) max cut on arbitrary matrices, which is NP-hard. However, we show that it can be done by setting up an appropriate relaxation of the PVD, and then using the fact that one can approximate MAX-CUT on matrices to a constant factor in polynomial time [4].

Theorem 14. *The graph partitions in the above regularity lemmas (Theorems 7 and 8), as well as matrix decompositions satisfying Corollaries 2, 3, and 4 can all be computed in polynomial time.*

1.3 Other Related Work

Another important pseudorandomness concept is *quasirandom graphs* [11, 27], which are defined by any one of a large class of equivalent graph properties, all shared by Erdős-Rényi random graphs. In their original form these properties require the graph in question to be dense, but there is an interesting line of ongoing work developing extensions to the sparse setting (e.g. [9, 10]). In particular, in [1] the authors propose an extension of two properties to sparse graphs, show equivalence, and show that these properties imply a certain sparse regularity lemma. This regularity lemma uses a “normalized” measure of cut values, and thus it is not directly comparable to either of the weak or strong regularity lemmas mentioned earlier, but it is closer in spirit to the weak one. For more detail on the deep theory of quasirandom graphs and some other related pseudorandomness concepts, we refer to the survey [27].

Another sparse Szemerédi-type regularity lemma was proved by Scott [32], which remarkably holds for *all* sparse input graphs. To explain its relationship to our work and to the other lemmas mentioned here, we will need to back up a step in the exposition. The phrasing of the Szemerédi regularity lemma given in Theorem 6 is not the original one; rather, it is a recently-popular rephrasing due to [29] (see also [17]) that is essentially equivalent to the original for dense graphs. Szemerédi’s original phrasing is as follows. Let us say that a bipartite graph $B = (V_1 \cup V_2, E)$ is ε -regular if, for any $S \subseteq V_1, T \subseteq V_2$ of sizes $|S| \geq \varepsilon|V_1|, |T| \geq \varepsilon|V_2|$, we have

$$|d(S, T) - d(V_1, V_2)| < \varepsilon. \quad (2)$$

For $G = (V, E)$, a partition $\Pi = \{V_i\}_{i=1}^p$ of V is ε -regular if for all but εp^2 pairs of parts $V_i, V_j \in \Pi$, the bipartite subgraph $G[V_i, V_j]$ between these parts is ε -regular. We then have:

Theorem 15 (Szemerédi Regularity Lemma, Original Phrasing [34]). *For any $\varepsilon > 0$ and any $G = (V, E)$, there is an ε -regular partition Π of V into $|\Pi| = O_\varepsilon(1)$ parts such that the sizes of any two parts in Π differ by at most 1.*

The parameter ε plays three roles in this theorem: it controls the sizes of the subsets $|S|, |T|$ that are measured in a regular part, it controls the amount $d(S, T)$ can fluctuate from the average in a regular part, and it controls the number of parts allowed to be irregular. One could just as easily define $(\varepsilon_1, \varepsilon_2, \varepsilon_3)$ -regularity to disentangle these uses. This is in contrast to the phrasing in Theorem 6, in which ε only plays one role. Since the goal in sparse regularity is essentially to replace ε by $\varepsilon d(G)$, the

sparse extension of the phrasing in Theorem 6 is unambiguous, but there are a few reasonable sparse extensions of the phrasing in Theorem 15 depending on which of these uses of ε are or aren't changed (this difference in ambiguity is perhaps a conceptual advantage of the phrasing in Theorem 6). Scott proved, quite surprisingly, that the second use (only) can be replaced *for free*:

Theorem 16 ([32]). *The Szemerédi regularity lemma, as phrased in Theorem 15, continues to hold when (2) is replaced by*

$$|d(S, T) - d(V_1, V_2)| < \varepsilon d(G).$$

This gives a certain sparse regularity lemma, but the guarantee is weaker than the sparse version of Theorem 6. This is because one still allows εp^2 exceptional parts, and in principle one could hide most or all of the mass of the graph in these parts (unlike the dense setting, where the total hidden mass is easily controlled by a union bound). By contrast, Theorem 6 does not have exceptional parts in which to hide mass. Put another way, Theorem 16 cannot be naturally phrased using the irregularity measure `szémirr`; rather, it would use a different, more complex irregularity measure that involves selectively editing the graph first. In this way it may be more in line with the recent “approximate” regularity lemma of Moshkovitz and Shapira [31].

2 Background

We start with some background terminology. For a matrix $A \in \mathbb{R}^{m \times n}$, the *Frobenius norm* is

$$\|A\|_F := \left(\sum_{i=1}^m \sum_{j=1}^n A_{ij}^2 \right)^{1/2}$$

or, equivalently, it is $\|v\|_2$ where $v = (\sigma_1, \dots, \sigma_m)$ is the vector of singular values of A . The associated *Frobenius inner product* is

$$\langle A, B \rangle_F := \sum_{i=1}^m \sum_{j=1}^n A_{ij} B_{ij} = \text{Tr}(AB^T).$$

The *Schatten k -norm* is

$$\|A\|_{F^{(k)}} := \|v\|_k$$

where again v is the singular values of A (so $\|A\|_{F^{(2)}} = \|A\|_F$). The *cut norm* of A is

$$\|A\|_{\square} := \max_{S \subseteq [m], T \subseteq [n]} \left| \sum_{s \in S, t \in T} A_{s,t} \right|;$$

as shorthand we will write this inner sum as

$$A(S, T) := \sum_{s \in S, t \in T} A_{s,t}.$$

A *cut vector* is a nonzero vector with 0/1 entries, and a *cut matrix* is a matrix that is constant on an $S \times T$ block and zero elsewhere, i.e. it is the outer product of two cut vectors times a scalar. The *spectral norm* of A is

$$\|A\|_2 := \max_{v \in \mathbb{R}^m, \|v\|_2=1} \|v^T A\|_2 = \max_{w \in \mathbb{R}^n, \|w\|_2=1} \|Aw\|_2,$$

or equivalently it is the largest singular value of A .

3 The Projection Value Decomposition

3.1 The Basic PVD

Relative to a *pair set* $P \subseteq \mathbb{R}^m \times \mathbb{R}^n$, we define the P -norm over matrices $A \in \mathbb{R}^{m \times n}$ as

$$\|A\|_P := \max_{(v,w) \in P \mid v,w \neq 0} \left| \frac{v^T A w}{\|v\|_2 \|w\|_2} \right|.$$

For some choices of P this is technically only a *seminorm* since we can possibly have $\|A\|_P = 0, A \neq 0$, but this is not really important. Homogeneity and the triangle inequality for this norm are easily verified (omitted). The *projection* of $A \in \mathbb{R}^{m \times n}$ onto $P \subseteq \mathbb{R}^m \times \mathbb{R}^n$, written $\Pi_P(A)$, is the projection of A (as a vector in $\mathbb{R}^{m \times n}$) into the subspace

$$\text{span} \{vw^T \mid (v, w) \in P\} \subseteq \mathbb{R}^{m \times n}.$$

For an ordered pair set $P \subseteq \mathbb{R}^m \times \mathbb{R}^n$, let $P^{(j)}$ denote the first j elements in the ordering. We define the *greedy ordering* of a pair set $P \subseteq \mathbb{R}^m \times \mathbb{R}^n$ inductively as follows: having determined $P^{(j)}$, the next element is

$$p_{j+1} := \arg \max_{p_{j+1} \in P} \left\| \Pi_{P^{(j)} \cup \{p_{j+1}\}}(A) - \Pi_{P^{(j)}}(A) \right\|_F.$$

To make this ordering uniquely defined, we should break ties in some canonical way, but for this paper it is not important how. In the following, unless explicitly stated otherwise we will assume that all pair sets are greedily ordered, and thus we reference subsets like $P^{(j)}$ directly. For $A \in \mathbb{R}^{m \times n}$, $P \subseteq \mathbb{R}^m \times \mathbb{R}^n$, the *projection values* of A with respect to P are

$$\sigma_j := \left\| \Pi_{P^{(j)}}(A) - \Pi_{P^{(j-1)}}(A) \right\|_F$$

for positive integers j . To phrase this as a matrix decomposition, we write

$$\Pi_P(A) = \sum_{i=1}^{mn} (\Pi_{P^{(i)}}(A) - \Pi_{P^{(i-1)}}(A)),$$

noting that the matrices in this sum are 1-dimensional, pairwise orthogonal, and the i^{th} one can be phrased as a linear combination of outer products from the first i pairs in P . We now prove some useful properties of the PVD.

Lemma 17. *For any $A \in \mathbb{R}^{m \times n}$, $P \subseteq \mathbb{R}^m \times \mathbb{R}^n$, the projection values of A are weakly decreasing.*

Proof. We compute:

$$\begin{aligned} \sigma_{j+1} &= \left\| \Pi_{P^{(j)} \cup \{p_{j+1}\}}(A) - \Pi_{P^{(j)}}(A) \right\|_F \\ &= \left\| \Pi_{P \setminus \{p_j\}} \left(\Pi_{P^{(j-1)} \cup \{p_{j+1}\}}(A) - \Pi_{P^{(j-1)}}(A) \right) \right\|_F \\ &\leq \left\| \Pi_{P^{(j-1)} \cup \{p_{j+1}\}}(A) - \Pi_{P^{(j-1)}}(A) \right\|_F \\ &\leq \left\| \Pi_{P^{(j-1)} \cup \{p_j\}}(A) - \Pi_{P^{(j-1)}}(A) \right\|_F \\ &= \sigma_j \end{aligned}$$

where the last inequality follows since p_j was chosen over p_{j+1} in the previous round. $\square$

With this, since $\mathbb{R}^{m \times n}$ is an mn -dimensional space, it follows that only the first mn projection values can be nonzero (and indeed, there are examples where all of the first mn projection values are nonzero, e.g. when A has all nonzero entries and P holds pairs of standard basis vectors in $\mathbb{R}^m, \mathbb{R}^n$). Thus in the sequel we consider the projection values to be only $\{\sigma_1, \dots, \sigma_{mn}\}$ and not the following zeroes. This also means that only the first mn pairs in the greedy ordering of P are relevant; the ordering does not actually need to be computed beyond this. One way we will use projection values is to control the corresponding P -norm:

Lemma 18. *For any $A \in \mathbb{R}^{m \times n}$, $P \subseteq \mathbb{R}^m \times \mathbb{R}^n$, and positive integer j , we have*

$$\sigma_j \geq \|A - \Pi_{P^{(j-1)}}(A)\|_P.$$

Proof. We compute:

$$\begin{aligned}
\|A - \Pi_{P^{(j-1)}}(A)\|_P &= \max_{(v,w) \in P | v,w \neq 0} \left| \frac{v^T (A - \Pi_{P^{(j-1)}}(A)) w}{\|v\|_2 \|w\|_2} \right| \\
&= \max_{(v,w) \in P | v,w \neq 0} \left| \frac{v^T \Pi_{P^{(j-1)} \cup \{(v,w)\}}(A - \Pi_{P^{(j-1)}}(A)) w}{\|v\|_2 \|w\|_2} \right| \\
&= \max_{(v,w) \in P | v,w \neq 0} \left| \frac{v^T (\Pi_{P^{(j-1)} \cup \{(v,w)\}}(A) - \Pi_{P^{(j-1)}}(A)) w}{\|v\|_2 \|w\|_2} \right| \\
&\leq \max_{(v,w) \in P | v,w \neq 0} \|\Pi_{P^{(j-1)} \cup \{(v,w)\}}(A) - \Pi_{P^{(j-1)}}(A)\|_2 \\
&= \max_{(v,w) \in P | v,w \neq 0} \|\Pi_{P^{(j-1)} \cup \{(v,w)\}}(A) - \Pi_{P^{(j-1)}}(A)\|_F \quad (\text{inner matrix has rank 1}) \\
&= \sigma_j. \quad \square
\end{aligned}$$

Lemma 19. For $A \in \mathbb{R}^{m \times n}$, $P \in \mathbb{R}^m \times \mathbb{R}^n$ with projection values $\varsigma = (\sigma_1, \dots, \sigma_{mn})$, for any $k \geq 2$ we have

$$\|\varsigma\|_k \leq \|\Pi_P(A)\|_{F^{(k)}}$$

with equality when $k = 2$.

Proof. Let us start with the case $k = 2$. We have:

$$\begin{aligned}
\|\varsigma\|_2^2 &= \sum_{i=1}^{mn} \sigma_i^2 \\
&= \sum_{i=1}^{mn} \|\Pi_{P^{(j)}}(A) - \Pi_{P^{(j-1)}}(A)\|_F^2.
\end{aligned}$$

The inner matrices in this summation are pairwise orthogonal (in Frobenius inner product). Thus we continue

$$\begin{aligned}
&= \left\| \sum_{i=1}^{mn} \Pi_{P^{(j)}}(A) - \Pi_{P^{(j-1)}}(A) \right\|_F^2 \\
&= \|\Pi_{P^{(mn)}}(A) - \Pi_{P^{(0)}}(A)\|_F^2 \\
&= \|\Pi_P(A)\|_F^2.
\end{aligned}$$

Now we turn to general $k \geq 2$. Let $v = (v_1, \dots, v_n)$ be the singular values of A (we assume w.l.o.g. that $n \leq m$), so the goal is to prove $\|v\|_k \geq \|\varsigma\|_k$. Since $\|v\|_2 = \|\varsigma\|_2$ from the above argument, the vector v (padded with zeroes) is a rotation of ς in $\mathbb{R}^{mn}$. Moreover, a standard fact is that (for any t) the leading t terms of the SVD give the best rank t approximation of A , which means

$$\|v_t\|_2 \geq \|\varsigma_t\|_2$$

for any index t , where v_t, ς_t denote the leading t singular, projection values. Since v, ς are sorted descendingly, it follows that there is an index $t \geq 1$ such that $v_t > \varsigma_t$ and $v_{t'} = \varsigma_{t'}$ for all $t' < t$. We argue that this property (together with the fact that v is a rotation of ς) implies that

$$\|v\|_k \geq \|\varsigma\|_k$$

as desired. To see this, repeat the following process until $\varsigma = v$: choose the last index z on which ς is nonzero, let t be the index as above, and rotate ς in the 2-dimensional subspace corresponding to indices t, z (thus preserving $\|\varsigma\|_2$) until either $\varsigma_t = v_t$ or $\varsigma_z = 0$. Then increment t or decrement z accordingly, and repeat. Each step shrinks ς_z while increasing ς_t , and it preserves $\|\varsigma\|_2$, which thus weakly increases $\|\varsigma\|_k$ (since $\varsigma_z \leq \varsigma_t$). Thus we transform ς into v while monotonically increasing $\|\varsigma\|_k$, and the lemma follows. $\square$

Let us generally write $\varsigma_r = \langle \sigma_1, \dots, \sigma_r \rangle$ for the vector of the leading r projection values. We then have the following truncation estimate for the projection value decomposition:

Theorem 20. Let $A \in \mathbb{R}^{m \times n}$, $P \subseteq \mathbb{R}^m \times \mathbb{R}^n$, r a nonnegative integer, $k \geq 2$, and let $\hat{A} = \Pi_{P(r)}(A) \in \mathbb{R}^{m \times n}$ (i.e. $\hat{A}$ is the sum of the leading r terms in the PVD). Then we have

$$\|A - \hat{A}\|_P \leq \frac{\|\zeta_{r+1}\|_k}{(r+1)^{1/k}} \leq \frac{\|A\|_{F^{(k)}}}{(r+1)^{1/k}}.$$

Proof. The latter inequality is immediate from Lemma 19, so let us prove the former. We compute:

$$\begin{aligned} \|A - \hat{A}\|_P^k &\leq \sigma_{r+1}^k && \text{Lemma 18} \\ &\leq \frac{\sum_{i=1}^{r+1} \sigma_i^k}{r+1} && \text{Lemma 17} \\ &= \frac{\|\zeta_{r+1}\|_k^k}{r+1} \end{aligned}$$

and the inequality now follows by taking the k^{th} root of both sides. $\square$

In the next section, we will discuss some other interesting ways to control $\|\zeta_{r+1}\|_k$ besides by $\|A\|_{F^{(k)}}$ as in Lemma 19. For now, we mention that one recovers the usual low rank approximation theorem via SVD by plugging in $k = 2$ and $P = \mathbb{R}^m \times \mathbb{R}^n$. One similarly recovers a version of CUR decomposition by plugging in $P = C \times R$ where C, R are the columns, rows of A respectively, or a version of cut decomposition by plugging in P as all pairs of cut vectors (this cut decomposition appears implicitly in the next section, but we will use Theorem 20 directly rather than writing down a corollary to make clear the role of the lemmas proved in this section).

3.2 Extension to Tensors

A tensor $T \in \mathbb{R}^{n_1 \times n_2 \times \dots \times n_s}$ is an s -dimensional array. It is often of interest to decompose a tensor into a small number of rank 1 tensors, i.e., outer products of vectors of the corresponding dimensionalities

$$T = \sum_i \alpha_i u_i^1 \otimes u_i^2 \otimes \dots \otimes u_i^s.$$

While there is no direct analog of SVD for tensors (e.g., one cannot generally have orthogonality of the vectors u_i^{ij}), we observe that the PVD itself extends naturally. For a tensor T as above, define its Frobenius norm as

$$\|T\|_F = \sqrt{\sum_{i_1, \dots, i_r} T_{i_1, \dots, i_s}^2}$$

and the s -form

$$T(x^1, \dots, x^s) = \sum_{i_1, i_2, \dots, i_s=1}^{n_1, \dots, n_s} T_{i_1, \dots, i_s} x_{i_1}^1 \dots x_{i_s}^s.$$

Then for a set of s -tuples of vectors $P \subseteq \mathbb{R}^{n_1} \times \dots \times \mathbb{R}^{n_s}$, we can define the P -norm of T as

$$\|T\|_P := \max_{(u^1, \dots, u^s) \in P \mid u^i \neq 0} \left| \frac{T(u^1, \dots, u^s)}{\|u^1\|_2 \dots \|u^s\|_2} \right|.$$

The span of a set of tuples P is the span of the vector space of the rank 1 tensors formed by outer products of each tuple in P , and the projection of T to P is the projection to its span, denoted as $\Pi_P(T)$. Then, analogous to matrices, for a given tensor T , we can define an ordering on the elements of P by greedily extracting rank 1 tensors with the highest P norm. Let the tuples in this ordering by p_j , the span of the first j be $P^{(j)}$ so that

$$p_{j+1} := \arg \max_{p_{j+1} \in P} \left\| \Pi_{P^{(j)} \cup \{p_{j+1}\}}(T) - \Pi_{P^{(j)}}(T) \right\|_F$$

and

$$\Pi_P(T) = \sum_{i=1}^{n_1 n_2 \dots n_s} \Pi_{P^{(i)}}(T) - \Pi_{P^{(i-1)}}(T).$$

Then Lemmas 17 and 18 hold for tensors with no change in their proofs, and Lemma 19 holds in the case $k = 2$ as well (the trouble with $k > 2$ is just that we are not aware of a clear definition of $\|\cdot\|_{F^{(k)}}$ for tensors, since the SVD is not clearly defined). We thus get the following partial extension of Theorem 20:

Theorem 21. *Let $T \in \mathbb{R}^{n_1 \times \dots \times n_s}$, $P \subseteq \mathbb{R}^{n_1} \times \dots \times \mathbb{R}^{n_s}$, r a nonnegative integer, $k \geq 2$, and let $\hat{T} \in \mathbb{R}^{n_1 \times \dots \times n_s} = \Pi_T(P^{(r)})$ (i.e. $\hat{T}$ is the sum of the leading r terms in the tensor PVD). Then we have*

$$\|T - \hat{T}\|_P \leq \frac{\|\varsigma_r\|_k}{(r+1)^{1/k}}.$$

This extends only one of the two inequalities stated in Theorem 20; for $k = 2$ we can further write

$$\frac{\|\varsigma_r\|_2}{(r+1)^{1/2}} \leq \frac{\|T\|_F}{(r+1)^{1/2}}.$$

This recovers the decomposition result of [13] for the case when $P = \mathbb{R}^{n_1} \times \dots \times \mathbb{R}^{n_s}$, and (since the PVD is efficiently computable in this case) it thus recovers the application to getting a PTAS for MAX-k-CSP for core-dense matrices and tensors (we refer to [13] for details). One can also get analogs of cut decomposition and CUR decompositions for tensors by plugging in appropriate choices of P .

4 Applications to Graph Regularity

We will now show how to recover and extend graph regularity lemmas using the PVD. Recall the following definition from the intro:

Definition 4 (Cut Pseudorandomness). *For an n -node graph $G = (V, E)$ with adjacency matrix A , we say that G is (k, r, C) cut pseudorandom if, relative to P as all pairs of cut vectors in $\mathbb{R}^n$, the leading r projection values $\varsigma_r := (\sigma_1, \dots, \sigma_r)$ of A satisfy*

$$\|\varsigma_r\|_k \leq C \frac{|E|}{n}.$$

We will usually omit the parameter C to mean that G is cut pseudorandom for some absolute constant $C > 0$ (not depending on k, r, n , or $|E|$).

4.1 Regularity Lemmas

We first give a *weak regularity lemma*, based on the following definition:

Definition 5 (Weak Irregularity [18]). *For $G = (V, E)$ and a partition $\Pi = \{V_i\}$ of V , we define*

$$\text{weakirr}(\Pi) := \max_{S, T \subseteq V} \left| e(S, T) - \sum_{V_i, V_j \in \Pi} |S \cap V_i| |T \cap V_j| d(V_i, V_j) \right|$$

(informally, this is the maximum amount a cut value changes in expectation if each bipartite subgraph $G[V_i, V_j]$ is replaced by a random bipartite graph of the same density).

We then prove:

Theorem 22 (Extended Weak Regularity Lemma [18]). *For any $\varepsilon > 0$, $k \geq 2$, there is $r = O(\varepsilon^{-k})$ such that for every (k, r) cut pseudorandom graph $G = (V, E)$, there is a partition $\Pi = \{V_i\}_{i=1}^p$ of V into $p = 2^{O(\varepsilon^{-k})}$ parts satisfying*

$$\text{weakirr}(\Pi) \leq \varepsilon |E|.$$

Proof. Applying Theorem 20 to the adjacency matrix A with P as all pairs of cut vectors and $r = \lceil \varepsilon^{-k} \rceil$, we have

$$\hat{A} = C_1 + \dots + C_r,$$

where each C_i is a cut matrix, so that

$$\|A - \hat{A}\|_P \leq \varepsilon \|C_r\|_k = O\left(\varepsilon \frac{|E|}{n}\right)$$

where the last equality follows by assumption that G is (k, r) cut pseudorandom. Note that

$$\|A - \hat{A}\|_P = \max_{v, w \in P} \left| \frac{v^T A w}{\|v\|_2 \|w\|_2} \right| \geq \frac{\max_{v, w \in P} |v^T A w|}{n} = \frac{\|A - \hat{A}\|_{\square}}{n},$$

so combined with the above, we can write

$$\|A - \hat{A}\|_{\square} = O(\varepsilon |E|).$$

Each C_i is the outer product of two cut vectors, each giving a bipartition of the dimensions $[n]$. Take Π to be the common refinement of all bipartitions used in any C_i , which thus has size

$$p \leq 2^{2r} \leq 2^{O(\varepsilon^{-k})}.$$

Now let $S, T \subseteq V$, and let $I_S, I_T \in \mathbb{R}^n$ be their respective binary indicator vectors. We compute:

$$\begin{aligned} \text{weakirr}(\Pi) &= \left| e(S, T) - \sum_{V_i, V_j \in \Pi} |S \cap V_i| |T \cap V_j| d(V_i, V_j) \right| \\ &= \left| I_S A I_T - I_S \hat{A} I_T \right| && I_S, I_T \text{ indicator vectors of } S, T \\ &= \left| I_S (A - \hat{A}) I_T \right| \\ &\leq \|A - \hat{A}\|_{\square} \\ &= O(\varepsilon |E|). \end{aligned}$$

The theorem now follows by a constant-factor reparametrization of ε to remove the constant factor hidden in the O . $\square$

We now give a new Szemerédi-type regularity lemma, which is more technically involved. Recall:

Definition 6 (Szemerédi Irregularity (implicit in [34], explicit in [29, 17])). *For $G = (V, E)$ and a partition $\Pi = \{V_i\}$ of V , we define*

$$\text{szémirr}(\Pi) := \sum_{V_i, V_j} \max_{S \subseteq V_i, T \subseteq V_j} |e(S, T) - |S||T|d(V_i, V_j)|.$$

Theorem 23. *For any $\varepsilon > 0$, there is $r = \text{tower}(O(\varepsilon^{-2}))$ such that for every $(2, r)$ cut pseudorandom graph $G = (V, E)$ with adjacency matrix A , there is a partition $\Pi = \{V_i\}_{i=1}^p$ of V into $p = \text{tower}(O(\varepsilon^{-2}))$ parts satisfying*

$$\text{szémirr}(\Pi) \leq \varepsilon |E|.$$

Proof. This builds on a generally analytic approach to the Szemerédi regularity lemma; see e.g. [33, 29]. Apply Theorem 20 to A with P as all pairs of cut vectors in $\mathbb{R}^m, \mathbb{R}^n$, giving projection values $\{\sigma_i\}$. Let Π be a vertex partition into p parts, let q be an integer, and let f an increasing function, all of which we will choose later in the proof (we will also choose r later). We then let

$$\hat{A} := \Pi_{P(q)}(A) \quad \text{and} \quad \hat{A}' := \Pi_{P(f(q))}(A),$$

so we can write

$$\begin{aligned}
\text{szémirr}(\Pi) &= \sum_{V_i, V_j \in \Pi} \max_{S \subseteq V_i, T \subseteq V_j} \left| \left(\widehat{A}(S, T) - |S||T|d(V_i, V_j) \right) + \left(\widehat{A}' - \widehat{A} \right)(S, T) + \left(A - \widehat{A}' \right)(S, T) \right| \\
&\leq \sum_{V_i, V_j} \max_{S, T} \left| \left(\widehat{A}(S, T) - |S||T|d(V_i, V_j) \right) \right| \\
&\quad + \sum_{V_i, V_j} \max_{S, T} \left| \left(\widehat{A}' - \widehat{A} \right)(S, T) \right| \\
&\quad + \sum_{V_i, V_j} \max_{S, T} \left| \left(A - \widehat{A}' \right)(S, T) \right|.
\end{aligned}$$

Our goal will be to control each of the three terms in this sum individually.

The First Term. To control the first term, we leverage our freedom to choose Π . Let Π be the common refinement of the bipartitions used in the pairs in $P^{(q)}$, which thus has $p \leq 4^q$ parts. Note that $\widehat{A}$ is then constant on the block corresponding to any pair of parts $V_i, V_j \in \Pi$. We thus have

$$\widehat{A}(S, T) = |S||T|d(V_i, V_j) \quad \text{for all } V_i, V_j \in \Pi, S \subseteq V_i, T \subseteq V_j$$

and so $\widehat{A}(S, T) - |S||T|d(V_i, V_j)$ is always zero, and so the first term in the sum is zero.

The Second Term. To control the second term $\left(\widehat{A} - \widehat{A}' \right)(S, T)$ in the sum, we will leverage our freedom to choose q and r . Let $f^{(i)}$ denote the function f iterated i times, and suppose we choose

$$r > f(\lceil \varepsilon^{-2} \rceil)(0).$$

Then by the pigeonhole principle there is $0 \leq i < \varepsilon^{-2}$ with

$$\sum_{j=f^{(i)}(0)+1}^{f^{(i+1)}(0)} \sigma_j^2 \leq \varepsilon^2 \sum_{j=1}^r \sigma_j^2 = O\left(\left(\varepsilon \cdot \frac{|E|}{n}\right)^2\right)$$

where this second expression comes from the assumption that G is (k, r) cut pseudorandom. We set $q := f^{(i)}(0)$ accordingly, and observe

$$\sum_{i=q+1}^{f(q)} \sigma_i^2 = \sum_{i,j=1}^n \left| \left(\widehat{A}' - \widehat{A} \right)_{ij} \right|^2$$

since by Lemma 19 these both measure $\left\| \widehat{A}' - \widehat{A} \right\|_F$. Then we compute:

$$\begin{aligned}
\sum_{i=q+1}^{f(q)} \sigma_i^2 &\geq \sum_{i,j=1}^n \left| \left(\widehat{A}' - \widehat{A} \right)_{ij} \right|^2 \\
&\geq \frac{1}{n^2} \left(\sum_{i,j=1}^n \left| \left(\widehat{A}' - \widehat{A} \right)_{ij} \right| \right)^2 && \text{Cauchy-Schwarz Inequality} \\
&\geq \frac{1}{n^2} \left(\sum_{V_i, V_j \in \Pi} \max_{S \subseteq V_i, T \subseteq V_j} \left| \left(\widehat{A}' - \widehat{A} \right)(S, T) \right| \right)^2
\end{aligned}$$

and so, rearranging, we have

$$\sum_{V_i, V_j \in \Pi} \max_{S \subseteq V_i, T \subseteq V_j} \left| \left(\widehat{A}' - \widehat{A} \right)(S, T) \right| = O(\varepsilon |E|).$$

Thus, up to constant-factor reparametrizations of ε , the second term only contributes a fraction of the allowed irregularity.

The Third Term. Finally, to control the third term $(A - \hat{A}') (S, T)$ in the sum, we leverage our freedom to choose f . We compute:

$$\begin{aligned}
\sum_{i,j=1}^p \max_{S \subseteq V_i, T \subseteq V_j} (A - \hat{A}') (S, T) &\leq p^2 \|A - \hat{A}'\|_{\square} \\
&\leq 16^q \|A - \hat{A}'\|_{\square} \\
&\leq 16^q n \|A - \hat{A}'\|_P \\
&\leq 16^q n \frac{\|\varsigma_r\|_2}{f(q)^{1/2}} && \text{Theorem 20} \\
&= \frac{16^q}{f(q)^{1/2}} O(|E|) && G \text{ is cut pseudorandom.}
\end{aligned}$$

This is a negligible fraction of the allowed irregularity, so long as we choose (say) $f(q) \geq 32^q$. Note that this requires us to choose f to be exponential, and the second part sets q by $O(\varepsilon^{-2})$ iterations of f , leading to the claimed tower-type size in $|\Pi|$ and r . $\square$

4.2 Relationship to Previous Pseudorandomness Concepts

We now show a relationship between cut pseudorandomness and other previous pseudorandomness concepts. We start with upper regularity.

Lemma 24. *For every positive integer r and $k > 1$, there is $\eta = 2^{O_k(-r)}$ such that any n -node graph $G = (V, E)$ that is L_k upper η -regular is $(2, r)$ cut pseudorandom.*

Proof. Let us start with the case $k = 2$. Let A be the adjacency matrix of G , and let

$$A = \sum_{i=1}^{mn} \Pi_{P(i)} (A) - \Pi_{P(i-1)} (A)$$

be the PVD of A with respect to P as all pairs of cut vectors in $\mathbb{R}^n$. We then have

$$\begin{aligned}
\|\varsigma_r\|_2^2 &= \sum_{i=1}^r \|\Pi_{P(i)} (A) - \Pi_{P(i-1)} (A)\|_F^2 \\
&= \left\| \sum_{i=1}^r \Pi_{P(i)} (A) - \Pi_{P(i-1)} (A) \right\|_F^2 && \text{inner matrices orthogonal} \\
&= \|\Pi_{P(r)} (A)\|_F^2.
\end{aligned}$$

Now, consider the vertex partition $\Pi = \{V_i\}$ that is the common refinement of all cut vectors used in $P^{(r)}$ (so $|\Pi| = 2^{O(r)}$), and let Q be the set of pairs of cut vectors corresponding to parts in this partition. Choose η such that $|Q| \leq \eta^{-1}$. We clearly have

$$\text{span} \{vw^T \mid (v, w) \in P^{(r)}\} \subseteq \text{span} \{vw^T \mid (v, w) \in Q\}$$

and thus we continue

$$\begin{aligned}
\|\varsigma_r\|_2^2 &\leq \|\Pi_Q (A)\|_F^2 \\
&= \sum_{V_i, V_j \in \Pi} \left(\frac{e(V_i, V_j)}{\sqrt{|V_i||V_j|}} \right)^2 \\
&= n^2 \cdot \sum_{V_i, V_j \in Q} \left(\frac{\sqrt{|V_i||V_j|}}{n} \cdot d(V_i, V_j) \right)^2 \\
&= O(n^2 d(G)^2) && G \text{ is } L_2 \text{ upper regular, } |Q| \leq \eta^{-1} \\
&= O(|E|^2/n^2)
\end{aligned}$$

and the lemma now follows by taking the square root of both sides.

It is observed in [7] that $L_{k' > k}$ upper regularity is a strictly stronger condition than L_k upper regularity, so this proof implicitly covers all $k > 2$. For $1 < k < 2$, by Hölder's inequality a graph that is L_k upper η -regular is L_2 upper $\text{poly}_k(\eta)$ -regular, and the lemma follows. $\square$

Now we turn to graphs of low threshold rank.

Lemma 25. *For any positive integer r , there is $\varepsilon = O(r^{-1/2})$ such that any n -node graph $G = (V, E)$ with unnormalized threshold rank $t_\varepsilon(G) = O(1)$ is $(2, r)$ cut pseudorandom.*

Proof. Let $v_r = (\sigma_1, \dots, \sigma_r)$ be the leading r singular values. We have $\|\varsigma_r\|_2 \leq \|v_r\|_2$, so it suffices to bound $\|v_r\|_2$. By definition we have

$$\sum_{i \mid v_i > \varepsilon |E|/n} \left(v_i \cdot \frac{n}{|E|} \right)^2 = O(1)$$

and so

$$\sum_{i \mid v_i > \varepsilon |E|/n} v_i^2 = O\left(\frac{|E|^2}{n^2}\right).$$

By a union bound, we have

$$\sum_{i \mid v_i \leq \varepsilon |E|/n} v_i^2 = O\left(r \varepsilon^2 \frac{|E|^2}{n^2}\right).$$

Thus, as long as we choose $\varepsilon \leq r^{-1/2}$, we have

$$\|\varsigma_r\|_2 \leq \|v_r\|_2 = O\left(\frac{|E|}{n}\right). \quad \square$$

Combined with Theorems 22 and 23, these implies weak and strong regularity lemmas for these graph classes, recovering and extending the weak regularity lemmas of [7], [20] and the strong regularity lemma of [26].

5 A Functional PVD

Whether or not one can *compute* the PVD in polynomial time depends on the complexity of finding

$$\max_{(v,w) \in P} \frac{v^T A w}{\|v\|_2 \|w\|_2}$$

for an input matrix A , which must be repeated up to mn times (with respect to changing A) to compute the relevant prefix of the greedy ordering of P . The complexity of this problem depends on the structure of P , of course, but let us discuss the runtime for the three decompositions mentioned in this paper. When $P = \mathbb{R}^m \times \mathbb{R}^n$, corresponding to the SVD, this is easily computed as the top singular vector pair which can be found in matrix multiplication time. For the CUR decomposition, P has size $|P| = O(mn)$, so there is a clear polynomial time maximization algorithm that brute forces over P . The most interesting case is the cut decomposition, where this maximization problem is generally NP-hard. Since one cannot compute an *exact* PVD here, we will develop an *approximate* PVD that is more amenable to *approximating* the maximizing pair. Let

$$f : \mathbb{R}^m \rightarrow [0, 1], g : \mathbb{R}^n \rightarrow [0, 1],$$

and for $P \subseteq \mathbb{R}^m \times \mathbb{R}^n$ let

$$\|A\|_{P,f,g} := \max_{(v,w) \in P \mid v,w \neq 0} \left| f(v)g(w) \frac{v^T A w}{\|v\|_2 \|w\|_2} \right|.$$

The (f, g) -PVD of $A \in \mathbb{R}^{m \times n}$ with respect to P is then defined exactly as before, but with respect to the ordering of P where the $(j+1)^{\text{th}}$ element is

$$p_{j+1} := \arg \max_{(v,w) \in P} f(v)g(w) \left\| \Pi_{P^{(j)} \cup \{(v,w)\}}(A) - \Pi_{P^{(j)}}(A) \right\|_F.$$

With this, we have the following generalizations of the lemmas in Section 1.1. Let $\{\sigma_i\}$ be the projection values, and let (v_i, w_i) be the i^{th} projection pair in the above ordering of P .

Lemma 26. *The quantity $\sigma_j f(v_j)g(w_j)$ is non-increasing in j .*

Proof. We compute:

$$\begin{aligned}
\sigma_{j+1} f(v_{j+1})g(w_{j+1}) &= f(v_{j+1})g(w_{j+1}) \left\| \Pi_{P^{(j)} \cup \{p_{j+1}\}}(A) - \Pi_{P^{(j)}}(A) \right\|_F \\
&= f(v_{j+1})g(w_{j+1}) \left\| \Pi_{P \setminus \{p_j\}} \left(\Pi_{P^{(j-1)} \cup \{p_{j+1}\}}(A) - \Pi_{P^{(j-1)}}(A) \right) \right\|_F \\
&\leq f(v_{j+1})g(w_{j+1}) \left\| \Pi_{P^{(j-1)} \cup \{p_{j+1}\}}(A) - \Pi_{P^{(j-1)}}(A) \right\|_F \\
&\leq f(v_j)g(w_j) \left\| \Pi_{P^{(j-1)} \cup \{p_j\}}(A) - \Pi_{P^{(j-1)}}(A) \right\|_F \\
&= f(v_j)g(w_j)\sigma_j
\end{aligned}$$

where the last inequality follows since p_j was chosen over p_{j+1} in the previous round. $\square$

Lemma 27. *For any positive integer j , we have*

$$\sigma_j f(v_j)g(w_j) \geq \|A - \Pi_{P^{(j-1)}}(A)\|_{P,f,g}.$$

Proof. We compute:

$$\begin{aligned}
\sigma_j &= \|\Pi_{P^{(j)}}(A) - \Pi_{P^{(j-1)}}(A)\|_F \\
&= \|\Pi_{P^{(j)}}(A) - \Pi_{P^{(j-1)}}(A)\|_2 \\
&= \max_{v \in \mathbb{R}^m, w \in \mathbb{R}^n} \left| \frac{v^T (\Pi_{P^{(j)}}(A) - \Pi_{P^{(j-1)}}(A)) w}{\|v\|_2 \|w\|_2} \right| \\
&\geq \max_{(v,w) \in P | v,w \neq 0} \left| \frac{v^T (\Pi_{P^{(j)}}(A) - \Pi_{P^{(j-1)}}(A)) w}{\|v\|_2 \|w\|_2} \right| \\
&\geq (f(v_j)g(w_j))^{-1} \cdot \max_{(v,w) \in P | v,w \neq 0} \left| f(v)g(w) \frac{v^T (\Pi_{P^{(j)}}(A) - \Pi_{P^{(j-1)}}(A)) w}{\|v\|_2 \|w\|_2} \right| \quad \text{range of } f, g \text{ is } [0, 1] \\
&= (f(v_j)g(w_j))^{-1} \max_{(v,w) \in P | v,w \neq 0} \left| f(v)g(w) \frac{v^T (A - \Pi_{P^{(j-1)}}(A)) w}{\|v\|_2 \|w\|_2} \right| \\
&= (f(v_j)g(w_j))^{-1} \|A - \Pi_{P^{(j-1)}}(A)\|_{P,f,g}
\end{aligned}$$

and the lemma follows by rearranging. $\square$

Lemma 19 still holds without change, since its proof does not use anything about the ordering of P . Putting these together, we prove:

Theorem 28. *Let r be a nonnegative integer, $k \geq 2$, and $\hat{A} = \Pi_{P^{(r)}}(A)$. Then we have*

$$\|A - \hat{A}\|_{P,f,g} \leq \frac{\|\zeta_{r+1}\|_k}{(r+1)^{1/k}} \leq \frac{\|A\|_{F^{(k)}}}{(r+1)^{1/k}}.$$

Proof. Again the latter inequality is immediate from Lemma 19, so we focus on the former one. We have:

$$\begin{aligned}
\|A - \hat{A}\|_{P,f,g}^k &\leq (f(v_{r+1})g(w_{r+1})\sigma_{r+1})^k && \text{Lemma 27} \\
&\leq \frac{\sum_{i=1}^{r+1} (f(v_i)g(w_i)\sigma_i)^k}{r+1} && \text{Lemma 26} \\
&\leq \frac{\sum_{i=1}^{r+1} \sigma_i^k}{r+1} && \text{range of } f, g \text{ is } [0, 1]
\end{aligned}$$

and now the theorem follows by taking the k^{th} root of both sides. $\square$

Finally, let us say why this PVD formulation with f, g gives algorithmic versions of the results in this paper. Imagine for a second that we could compute efficiently the max cut of a matrix *exactly* (we can't, of course, it's NP-hard) – that is, given $A \in \mathbb{R}^{m \times n}$, we can find

$$\arg \max_{S \subseteq [m], T \subseteq [n]} |A(s, t)|.$$

Letting

$$f(v) := \frac{\|v\|_2}{\sqrt{m}}, g(w) := \frac{\|w\|_2}{\sqrt{n}},$$

and taking P to be all pairs of cut vectors in $\mathbb{R}^m, \mathbb{R}^n$, we have

$$\|A\|_{\square} = \frac{\|A\|_{P, f, g}}{\sqrt{mn}}$$

and the corresponding PVD would then be efficiently computable. In particular, this gives

$$\|A - \hat{A}\|_{\square} \leq \sqrt{mn} \cdot \frac{\|s_{r+1}\|_k}{(r+1)^{1/k}} \leq \sqrt{mn} \cdot \frac{\|A\|_{F^{(k)}}}{(r+1)^{1/k}}$$

which is more in line with the phrasing used previously in this paper. But that said, since we can't compute max cuts exactly, we instead use the following result of Alon and Naor:

Theorem 29 ([4]). *There is an algorithm that, on a matrix $A \in \mathbb{R}^{m \times n}$, runs in polynomial time and returns $S \subseteq [m], T \subseteq [n]$ satisfying*

$$|A(S, T)| > 0.56 \|A\|_{\square}.$$

There are now two valid ways to proceed. One reasonable approach is to define f, g as before, and then observe that Theorem 28 holds up to constant factors when the PVD ordering only chooses $p_{j+1} =: (v, w)$ maximizing

$$f(v)g(w) \|\Pi_{P^{(j)} \cup \{(v, w)\}}(A) - \Pi_{P^{(j)}}(A)\|_F$$

up to constant factors, as in Theorem 29. This does not require any major modifications to the above proofs. An alternate, more direct approach is to define the PVD ordering by repeatedly using Theorem 29 to find the next pair in P , and then define f, g as above, post-hoc, rescaled by a suitable constant on each chosen pair (and leaving the f, g values on all other pairs unchanged) to make the $(j+1)^{st}$ chosen pair the *true* maximizer of the quantity

$$f(v)g(w) \|\Pi_{P^{(j)} \cup \{(v, w)\}}(A) - \Pi_{P^{(j)}}(A)\|_F.$$

Then one can apply the “exact” results of this section as stated above. More generally, this approach works whenever one has an approximation algorithm for finding a norm-maximizing pair.

References

- [1] ALON, N., COJA-OGHLAN, A., HÀN, H., KANG, M., RÖDL, V., AND SCHACHT, M. Quasi-randomness and algorithmic regularity for graphs with general degree distributions. *SIAM Journal on Computing* 39, 6 (2010), 2336–2362.
- [2] ALON, N., FISCHER, E., KRIVELEVICH, M., AND SZEGEDY, M. Efficient testing of large graphs. *Combinatorica* 20 (05 2001).
- [3] ALON, N., FISCHER, E., NEWMAN, I., AND SHAPIRA, A. A combinatorial characterization of the testable graph properties: It's all about regularity. *SIAM J. Comput.* 39 (01 2009), 143–167.
- [4] ALON, N., AND NAOR, A. Approximating the cut-norm via grothendieck's inequality. In *Proceedings of the thirty-sixth annual ACM symposium on Theory of computing* (2004), ACM, pp. 72–80.
- [5] BARAK, B., RAGHAVENDRA, P., AND STEURER, D. Rounding semidefinite programming hierarchies via global correlation. In *2011 IEEE 52nd Annual Symposium on Foundations of Computer Science* (2011), IEEE, pp. 472–481.

- [6] BLUM, A., HOPCROFT, J., AND KANNAN, R. *Foundations of Data Science*. Cambridge University Press, 2020.
- [7] BORGS, C., CHAYES, J., COHN, H., AND ZHAO, Y. An l^p theory of sparse graph convergence i: Limits, sparse random graph models, and power law distributions. *Transactions of the American Mathematical Society* (01 2014).
- [8] BORGS, C., CHAYES, J., COHN, H., AND ZHAO, Y. An l^p theory of sparse graph convergence ii: Ld convergence, quotients, and right convergence. *The Annals of Probability* 46 (08 2014).
- [9] CHUNG, F., AND GRAHAM, R. Sparse quasi-random graphs. *Combinatorica* 22, 2 (2002), 217–244.
- [10] CHUNG, F., AND GRAHAM, R. Quasi-random graphs with given degree sequences. *Random Structures & Algorithms* 32, 1 (2008), 1–19.
- [11] CHUNG, F. R. K., GRAHAM, R. L., AND WILSON, R. M. Quasi-random graphs. *Combinatorica* 9, 4 (1989), 345–362.
- [12] CONLON, D., FOX, J., AND ZHAO, Y. Extremal results in sparse pseudorandom graphs. *Advances in Mathematics* 256 (04 2012).
- [13] DE LA VEGA, W. F., KARPINSKI, M., KANNAN, R., AND VEMPALA, S. Tensor decomposition and approximation schemes for constraint satisfaction problems. In *Proceedings of the thirty-seventh annual ACM symposium on Theory of computing* (2005), ACM, pp. 747–754.
- [14] DRINEAS, P., KANNAN, R., AND MAHONEY, M. W. Fast monte carlo algorithms for matrices III: computing a compressed approximate matrix decomposition. *SIAM J. Comput.* 36, 1 (2006), 184–206.
- [15] DRINEAS, P., AND MAHONEY, M. W. Randnla: randomized numerical linear algebra. *Commun. ACM* 59, 6 (2016), 80–90.
- [16] ERDŐS, P. Extremal problems in graph theory. In *Proceedings of the Symposium on Theory of Graphs and its Applications* (1963), p. 2936.
- [17] FOX, J., AND MIKLÓS LOVÁSZ, L. A tight lower bound for szemerédi’s regularity lemma. *Combinatorica* (03 2014).
- [18] FRIEZE, A., AND KANNAN, R. Quick approximation to matrices and applications. *Combinatorica* 19, 2 (1999), 175–220.
- [19] GERKE, S., AND STEGER, A. The sparse regularity lemma and its applications. *Surveys in combinatorics* 327 (2005), 227–258.
- [20] GHARAN, S. O., AND TREVISAN, L. A new regularity lemma and faster approximation algorithms for low threshold rank graphs. *Theory of Computing* 11 (2015), 241–256.
- [21] GURUSWAMI, V., AND SINOP, A. K. Lasserre hierarchy, higher eigenvalues, and approximation schemes for graph partitioning and quadratic integer programming with psd objectives. In *2011 IEEE 52nd Annual Symposium on Foundations of Computer Science* (2011), IEEE, pp. 482–491.
- [22] GURUSWAMI, V., AND SINOP, A. K. Faster sdp hierarchy solvers for local rounding algorithms. In *2012 IEEE 53rd Annual Symposium on Foundations of Computer Science* (2012), IEEE, pp. 197–206.
- [23] GURUSWAMI, V., AND SINOP, A. K. Approximating non-uniform sparsest cut via generalized spectra. In *Proceedings of the twenty-fourth annual ACM-SIAM symposium on Discrete algorithms* (2013), SIAM, pp. 295–305.
- [24] KANNAN, R., AND VEMPALA, S. S. Spectral algorithms. *Foundations and Trends in Theoretical Computer Science* 4, 3-4 (2009), 157–288.

- [25] KANNAN, R., AND VEMPALA, S. S. Randomized algorithms in numerical linear algebra. *Acta Numer.* 26 (2017), 95–135.
- [26] KOHAYAKAWA, Y. Szemerédi’s regularity lemma for sparse graphs. In *Foundations of computational mathematics*. Springer, 1997, pp. 216–230.
- [27] KRIVELEVICH, M., AND SUDAKOV, B. Pseudo-random graphs. In *More sets, graphs and numbers*. Springer, 2006, pp. 199–262.
- [28] LOVÁSZ, L. *Large Networks and Graph Limits*. American Mathematical Society Colloquium Publications, 01 2012.
- [29] LOVÁSZ, L., AND SZEGEDY, B. Szemerédi’s lemma for the analyst. *Geometric and Functional Analysis* 17 (04 2007), 252–270.
- [30] MAHONEY, M. W., AND DRINEAS, P. CUR matrix decompositions for improved data analysis. *Proc. Natl. Acad. Sci. U.S.A.* 106, 3 (2009), 697–702.
- [31] MOSHKOVITZ, G., AND SHAPIRA, A. A sparse regular approximation lemma. *Transactions of the American Mathematical Society* 371, 10 (2019), 6779–6814.
- [32] SCOTT, A. Szemerédi’s regularity lemma for matrices and sparse graphs. *Combinatorics, Probability and Computing* 20, 3 (2011), 455–466.
- [33] SZEGEDY, B. Limits of kernel operators and the spectral regularity lemma. *European Journal of Combinatorics* 32 (03 2010).
- [34] SZEMERÉDI, E. Regular partitions of graphs. Tech. rep., STANFORD UNIV CALIF DEPT OF COMPUTER SCIENCE, 1975.
- [35] TAO, T. The spectral proof of the Szemerédi regularity lemma. <https://terrytao.wordpress.com/2012/12/03/the-spectral-proof-of-the-szemerédi-regularity-lemma/>, 2012.
- [36] TITS, J. Sur la trialité et certains groupes qui s’ en déduisent. *Publications Mathématiques de l’IHÉS* 2 (1959), 13–60.
- [37] WENGER, R. Extremal graphs with no c_4 ’s, c_6 ’s, or c_{10} ’s. *Journal of Combinatorial Theory, Series B* 52 (05 1991), 113–116.