
Improved analysis of randomized SVD for top-eigenvector approximation

Ruo-Chun Tzeng¹

Po-An Wang²

Florian Adriaens¹

Aristides Gionis¹

¹Division of Theoretical Computer Science

²Division of Decision and Control Systems
KTH Royal Institute of Technology, Sweden

Chi-Jen Lu³

³Institute of Information Science
Academia Sinica, Taiwan

Abstract

Computing the top eigenvectors of a matrix is a problem of fundamental interest to various fields. While the majority of the literature has focused on analyzing the reconstruction error of low-rank matrices associated with the retrieved eigenvectors, in many applications one is interested in finding one vector with high Rayleigh quotient. In this paper we study the problem of approximating the top-eigenvector. Given a symmetric matrix \mathbf{A} with largest eigenvalue λ_1 , our goal is to find a vector $\hat{\mathbf{u}}$ that approximates the leading eigenvector \mathbf{u}_1 with high accuracy, as measured by the ratio $R(\hat{\mathbf{u}}) = \lambda_1^{-1} \hat{\mathbf{u}}^T \mathbf{A} \hat{\mathbf{u}} / \hat{\mathbf{u}}^T \hat{\mathbf{u}}$. We present a novel analysis of the randomized SVD algorithm of Halko et al. (2011b) and derive tight bounds in many cases of interest. Notably, this is the first work that provides non-trivial bounds for approximating the ratio $R(\hat{\mathbf{u}})$ using randomized SVD with any number of iterations. Our theoretical analysis is complemented with a thorough experimental study that confirms the efficiency and accuracy of the method.

1 INTRODUCTION

Spectral methods, which typically rely on computing the leading eigenvectors of an appropriately-designed matrix, have been shown to provide high-quality solutions to a variety of problems in the fields of data analysis, optimization, clustering and learning (Kannan and Vempala, 2009). From a computational perspec-

tive, randomized approaches for spectral methods often give good estimates of leading eigenvectors and low-rank structures, opening up the possibility of dealing with truly large datasets (Halko et al., 2011a).

In this paper, we study the problem of approximating the leading eigenvector of a matrix while using a small amount of memory and making a limited number of passes over the input matrix. More concretely, given a symmetric matrix \mathbf{A} with largest eigenvalue λ_1 , our goal is to find a vector $\hat{\mathbf{u}}$ that maximizes the ratio

$$R(\hat{\mathbf{u}}) = \lambda_1^{-1} \frac{\hat{\mathbf{u}}^T \mathbf{A} \hat{\mathbf{u}}}{\hat{\mathbf{u}}^T \hat{\mathbf{u}}}. \quad (1)$$

Note that since λ_1 is fixed given \mathbf{A} , it can be omitted from the definition of R ; it is used only for convenience, to ensure that $R \leq 1$. Often, in different applications, in addition to having to select which matrix \mathbf{A} to use, it is also required that $\hat{\mathbf{u}} \in \mathcal{T} \subseteq \mathbb{R}^n$, where \mathcal{T} is typically a discrete subspace of \mathbb{R}^n . A common strategy in this case, is to first compute an approximation of the leading eigenvector in \mathbb{R}^n and then “round” the solution in \mathcal{T} . Below we outline some prominent examples of this scheme.

(1) The most direct example is PCA, where \mathbf{A} is the covariance matrix (Jolliffe, 1986); in this case $\mathcal{T} = \mathbb{R}^n$, and no rounding is required; (2) In the community-detection problem we can partition a network into two communities (and then recursively find more communities) by maximizing modularity (Newman, 2006), which can be mapped to our setting by taking \mathbf{A} to be the modularity matrix and $\mathcal{T} = \{\pm 1\}^n$; (3) The problem of finding k conflicting groups in signed networks can be formulated by taking \mathbf{A} to be the adjacency matrix of the signed network and $\mathcal{T} = \{0, -1, \ell\}^n$, for $\ell \in [k-1]$ (Bonchi et al., 2019; Tzeng et al., 2020); (4) For the fair densest subgraph, Anagnostopoulos et al. (2020) consider $\mathcal{T} = \{0, 1\}^n$ and obtain \mathbf{A} after projecting the adjacency matrix onto the subspace orthogonal

Proceedings of the 25th International Conference on Artificial Intelligence and Statistics (AISTATS) 2022, Valencia, Spain. PMLR: Volume 151. Copyright 2022 by the author(s).

to a given fairness labeling $z \in \{\pm 1\}^n$; (5) In several other applications, a solution to our problem is used as an intermediate step in the proposed method (Abdullah et al., 2014; Hopkins et al., 2016; Allen-Zhu and Li, 2016; Silva et al., 2018).

Despite numerous pass-efficient algorithms for computing top eigenvectors proposed in the literature, prior analysis of $R(\hat{\mathbf{u}})$ have strong limitations when applied in practice. The main shortcoming is that most works provide *additive bounds* and require $\Omega(\ln n)$ passes to be meaningful (Simchowitz et al., 2018), whereas a smaller number of passes (constant or even a single pass) is critical in practical settings. It is unclear in the state-of-the-art whether $\Omega(\ln n)$ passes is necessary for previous methods, or whether such a bound is an artifact of the analysis.

In this paper we demonstrate that the requirement of $\Omega(\ln n)$ passes in the analysis of prior works is artificial. We show this by giving a *multiplicative bound* for $R(\hat{\mathbf{u}})$ achieved by the randomized SVD method (RSVD) of Halko et al. (2011b), which is one of the most prominent and widely-implemented pass-efficient algorithms (Pedregosa et al., 2011; Řehůřek and Sojka, 2010; Corporation, 2021; Erichson et al., 2019; Terray and Pinsard, 2021; Liutkus, 2021)

Our analysis shows that for any positive semidefinite matrix, RSVD returns with high probability a vector $\hat{\mathbf{u}}$ satisfying $R(\hat{\mathbf{u}}) = \Omega((d/n)^{1/(2q+1)})$ after $q \in \mathbb{N}$ iterations (Theorem 1), using $\mathcal{O}(dn)$ space for $d \in \mathbb{N}$, where typically $d \ll n$ (e.g., $d = \mathcal{O}(\ln n)$). Theorem 2 shows that our analysis is tight. Notably, our analysis subsumes the guarantee by prior works in the regime of $\Omega(\ln n)$ passes (Remark 1), and to the best of our knowledge, provides the first non-trivial guarantee of $R(\hat{\mathbf{u}})$ in the literature of pass-efficient algorithms for $o(\ln n)$ passes. Moreover, we show that under some natural conditions satisfied by real-world datasets, it is even possible to achieve $R(\hat{\mathbf{u}}) = \Omega(1)$ with a single pass (Remark 2).

Our core technical argument is a reduction from the optimization problem of maximizing R over a random subspace to the problem of estimating the projection length of a vector onto a random subspace. By using our technique, we derive the first non-trivial guarantee of $R(\hat{\mathbf{u}})$ for any number of passes for indefinite matrices (Theorem 4), under mild conditions (Assumption 1).

In addition, we propose an extension of the RSVD method, called **RandSum**, by using a random matrix sampled from Bernoulli(p) with mean $p \in (0, 1)$. While such a random matrix is rarely used in the literature of random projections, we show that there exist applications (Bonchi et al., 2019; Tzeng et al., 2020) espe-

cially suitable for this technique, and we show several properties of such a random matrix, which may be of independent interest.

2 RELATED WORK

For lack of space, we provide a brief overview of the existing literature, focusing on the most relevant works to our paper. For a general introduction on pass-efficient algorithms for matrix approximations, we refer the reader to Mahoney et al. (2011); Woodruff et al. (2014); Martinsson and Tropp (2020).

The study of $R(\hat{\mathbf{u}})$ for pass-efficient algorithms can be dated back to Kuczyński and Woźniakowski (1992) who analyzed two classical methods: the power method and the Lanczos method with random start. For any positive semidefinite matrix, they showed that the power method (respectively, Lanczos method) with random start, after $q \geq 2$ iterations returns an approximated top-eigenvector $\hat{\mathbf{u}}$ with $\mathbb{E}[R(\hat{\mathbf{u}})] \geq 1 - 0.871 \frac{\ln n}{q-1}$ (respectively, $\mathbb{E}[R(\hat{\mathbf{u}})] \geq 1 - 2.575 \left(\frac{\ln n}{q-1}\right)^2$).

The aforementioned methods are generalized to randomized SVD (Halko et al., 2011b) and block-Krylov methods (Musco and Musco, 2015), and a similar additive analysis of $R(\hat{\mathbf{u}})$ by Musco and Musco (2015) showed that for any positive semidefinite matrix, RSVD (respectively, randomized block-Krylov method) using $\mathcal{O}(nd)$ space (respectively, $\mathcal{O}(ndq)$ space) and after q iterations, returns an approximate top-eigenvector $\hat{\mathbf{u}}$ with $R(\hat{\mathbf{u}}) \geq 1 - \mathcal{O}(\frac{\ln n}{q})$ (respectively, $R(\hat{\mathbf{u}}) \geq 1 - \mathcal{O}((\frac{\ln n}{q})^2)$), with probability at least $1 - e^{-\Omega(d)}$.¹

The analysis of the previous works (Kuczyński and Woźniakowski, 1992; Musco and Musco, 2015) is tight, as shown by Simchowitz et al. (2018) for a class of methods (which include RSVD and block Krylov), which with high probability fail to find a vector $\hat{\mathbf{u}}$ with $R(\hat{\mathbf{u}}) \geq 23/24$ within $q = \mathcal{O}(\ln n)$ passes.

In the aforementioned works there are two limitations. First, the bounds of Kuczyński and Woźniakowski (1992) and Musco and Musco (2015) are additive, and unfortunately require $\Omega(\ln n)$ passes to be meaningful. In contrast, our analysis provides a multiplicative bound for $R(\hat{\mathbf{u}})$ and offers non-trivial guarantees for any number of passes. Second, the applicability of the methods of Kuczyński and Woźniakowski (1992) and Musco and Musco (2015) is limited to positive semidefinite matrices. Instead, we provide a sharp analysis of randomized SVD for positive semidefinite matrices and

¹Musco and Musco (2015) showed that the aforementioned results hold with constant probability, which could be improved to hold with probability $1 - e^{-\Omega(d)}$ by using stronger concentration results (Rudelson and Vershynin, 2010) in their proofs of Lemma 4 and Lemma 9.

show that our proof techniques generalize to indefinite matrices under mild conditions.

To complement our study, we briefly compare the measure $R(\hat{\mathbf{u}})$ with other classical metrics. Note here that, even though it is possible to convert an error guarantee for classical metrics (Xu et al., 2018; Drineas et al., 2018; Ghashami et al., 2016; Chen et al., 2017; Musco and Woodruff, 2017; Huang, 2018) into a lower bound for $R(\hat{\mathbf{u}})$ by matrix perturbation theory (Stewart and Guang Sun, 1990; Yu et al., 2015), the resulting bound is additive and depends on the eigengap. We also note that classical metrics typically compare the approximation $\hat{\mathbf{u}}$ to the top-eigenvector \mathbf{u}_1 of \mathbf{A} , however, such a comparison is not meaningful in our setting as small distance between $\hat{\mathbf{u}}$ and \mathbf{u}_1 ² is a sufficient but not necessary condition for having large $R(\hat{\mathbf{u}})$.

3 PRELIMINARIES

Let \mathbb{N} be the set of natural numbers excluding 0. Let \mathbb{R} be the set of real numbers, $\mathbb{S}^{m-1} = \{\mathbf{x} \in \mathbb{R}^m : \mathbf{x}^T \mathbf{x} = 1\}$, and $[m] = \{1, \dots, m\}$. Let $\text{range}(\mathbf{M})$ denote the column space of matrix \mathbf{M} , and $\|\cdot\|_F$ and $\|\cdot\|_2$ denote the Frobenius norm and the spectral norm, respectively. For a square matrix \mathbf{M} , let $\lambda_i(\mathbf{M})$ be its i -th largest eigenvalue and $\mathbf{u}_i(\mathbf{M})$ the corresponding eigenvector, and let $\sigma_i(\mathbf{M})$ be the i -th largest singular value. In all subsequent sections, we use boldface \mathbf{A} to denote the input matrix, and abbreviate $\lambda_i = \lambda_i(\mathbf{A})$, $\mathbf{u}_i = \mathbf{u}_i(\mathbf{A})$, and $\sigma_i = \sigma_i(\mathbf{A})$. We use $\langle \cdot, \cdot \rangle$ to denote the vector inner product. Finally, we use $\mathbf{1}_n = [1, \dots, 1]^T$ to denote the n -dimensional vector of all 1's and $\mathbf{0}_n = [0, \dots, 0]^T$ to denote the n -dimensional vector of all 0's.

For simplicity, we assume that the input matrix \mathbf{A} is real-valued and symmetric, with $\lambda_1 > 0$.

Definition 1 (Vector projection onto subspace). *Let $\mathbf{v} \in \mathbb{R}^n$ be a nonzero vector and $\mathcal{X} \subseteq \mathbb{R}^n$ be a nonempty subspace. The projection length of \mathbf{v} onto \mathcal{X} is given by $\cos \theta(\mathbf{v}, \mathcal{X})$, where*

$$\theta(\mathbf{v}, \mathcal{X}) = \cos^{-1} \left(\max_{\mathbf{x} \in \mathcal{X}} \frac{\langle \mathbf{v}, \mathbf{x} \rangle}{\|\mathbf{v}\|_2 \|\mathbf{x}\|_2} \right)$$

is the projection angle. For a matrix \mathbf{X} , we use $\theta(\mathbf{v}, \mathbf{X})$ to denote the projection angle of \mathbf{v} onto the range of \mathbf{X} .

It is well-known that projecting any vector $\mathbf{v} \in \mathbb{R}^n$ onto the range(\mathbf{S}) of a random matrix $\mathbf{S} \sim \mathcal{N}(0, 1)^{n \times d}$ results in $\cos^2 \theta(\mathbf{v}, \mathbf{S}) \approx d/n$ with high probability.

Lemma 1. (Hardt and Price, 2014) *Let $\mathbf{v} \in \mathbb{R}^n$ be a nonzero vector and $\mathbf{S} \sim \mathcal{N}(0, 1)^{n \times d}$, where $n, d \in \mathbb{N}$*

²More precisely, the distance between $\hat{\mathbf{u}}$ and the eigenspace associated with the largest eigenvalue λ_1 of \mathbf{A} .

Algorithm 1: RSVD($\mathbf{A}, \mathcal{D}, q, d$)

```

 $\mathbf{Y} \leftarrow \mathbf{A}^q \mathbf{S}$  where  $\mathbf{S} \sim \mathcal{D}$ ;
 $\mathbf{Y} = \mathbf{Q} \mathbf{R}$ ;
 $\mathbf{B} \leftarrow \mathbf{Q}^T \mathbf{A} \mathbf{Q}$ ;
 $\hat{\mathbf{u}} = \mathbf{Q} \mathbf{u}_1(\mathbf{B})$ ;
return  $\hat{\mathbf{u}}$ ;

```

and $n \geq d$. Then,

$$\cos^2 \theta(\mathbf{v}, \mathbf{S}) = \Theta \left(\frac{d}{n} \right),$$

with probability at least $1 - e^{-\Omega(d)}$.

For completeness, we provide the proof of Lemma 1 in Appendix A.1. The proof idea is to observe that $\frac{\|\mathbf{s}^T \mathbf{v}\|_2}{\sigma_1(\mathbf{S})} \leq \cos \theta(\mathbf{v}, \mathbf{S}) \leq \frac{\|\mathbf{s}^T \mathbf{v}\|_2}{\sigma_d(\mathbf{S})}$ and use the concentration of the extreme singular values of a Gaussian random matrix.

More generally, Lemma 1 holds for any random matrix \mathbf{S} whose range is uniformly distributed with respect to the Haar measure on Grassmannian $\mathcal{G}_{n,d}$ of all the d -dimensional subspaces of \mathbb{R}^n , written as $\text{range}(\mathbf{S}) \sim \text{Uniform}(\mathcal{G}_{n,d})$. The reader may refer to Achlioptas (2001) and Halko et al. (2011b) for other choices of \mathbf{S} and Vershynin (2018, Section 5) for a general introduction to this phenomenon.

4 RANDOMIZED SVD

We briefly review the following variant of the randomized SVD (RSVD) algorithm, as proposed by Halko et al. (2011b), and shown in Algorithm 1. The algorithm returns an estimate $\hat{\mathbf{u}}$ of the leading eigenvector \mathbf{u}_1 of the input matrix \mathbf{A} . It uses $\mathcal{O}(dn)$ space and requires $q+1$ passes over the matrix \mathbf{A} , where $q \in \mathbb{N}$.³ The distribution \mathcal{D} is over $\mathbb{R}^{n \times d}$, and one particular instance of the algorithm sets $\mathcal{D} = \mathcal{N}(0, 1)^{n \times d}$. The algorithm begins with a random projection $\mathbf{Y} = \mathbf{A}^q \mathbf{S}$. The eigenvectors of \mathbf{A}^q are the same as \mathbf{A} , but the eigenvalues of \mathbf{A}^q have much stronger decay. Thus intuitively, by taking powers of the input matrix, the relative weight of the eigenvectors associated with the small eigenvalues is reduced, which is helpful in the basis identification for input matrices whose eigenvalues decay slowly. After projecting, the algorithm efficiently approximates the top-eigenvector of \mathbf{A} by

$$\hat{\mathbf{u}} \in \arg\max \{ \mathbf{v}^T \mathbf{A} \mathbf{v} : \mathbf{v} \in \text{range}(\mathbf{Y}) \cap \mathbb{S}^{n-1} \}. \quad (2)$$

³More precisely, RSVD requires q passes when $d = 1$ and $q+1$ passes when $d > 1$ as there is no need to compute $\mathbf{u}_1(\mathbf{B})$ when $d = 1$.

Indeed, any $\mathbf{v} \in \text{range}(\mathbf{Y})$ of unit length can be written as $\mathbf{v} = \mathbf{Q}\mathbf{a}$ for some $\mathbf{a} \in \mathbb{S}^{d-1}$, where \mathbf{Q} is an $n \times d$ orthonormal basis given by a QR decomposition of \mathbf{Y} . So it follows that

$$\max_{\mathbf{v} \in \text{range}(\mathbf{Y}) \cap \mathbb{S}^{n-1}} \mathbf{v}^T \mathbf{A} \mathbf{v} = \max_{\mathbf{a} \in \mathbb{S}^{d-1}} \mathbf{a}^T \mathbf{B} \mathbf{a} = \lambda_1(\mathbf{B}).$$

Thus, the vector $\hat{\mathbf{u}} = \mathbf{Q}\mathbf{u}_1(\mathbf{B})$ maximizes expression (2), and $\mathbf{u}_1(\mathbf{B})$ can be efficiently computed as the matrix \mathbf{B} is of dimension $d \times d$.

4.1 Analysis of RSVD

We now derive lower and upper bounds for $R(\hat{\mathbf{u}})$, where $\hat{\mathbf{u}}$ is the output of Algorithm 1, and $R(\mathbf{v}) = \lambda_1^{-1} \frac{\mathbf{v}^T \mathbf{A} \mathbf{v}}{\mathbf{v}^T \mathbf{v}}$ is defined for any nonzero vector $\mathbf{v} \in \mathbb{R}^n$. Note that due to expression (2), $\hat{\mathbf{u}}$ maximizes R over the column space $\text{range}(\mathbf{Y})$ of \mathbf{Y} . Since $\text{range}(\mathbf{Y}) = \{\mathbf{Y}\mathbf{a} : \mathbf{a} \in \mathbb{R}^d\}$, we can rewrite $R(\hat{\mathbf{u}})$ as

$$R(\hat{\mathbf{u}}) = \max_{\mathbf{v} \in \text{range}(\mathbf{Y}) \setminus \{\mathbf{0}_n\}} R(\mathbf{v}) = \max_{\mathbf{a} \in \mathbb{S}^{d-1}} R(\mathbf{Y}\mathbf{a}),$$

where the latter equality follows from the scale invariance of R . For notational convenience, we denote $R_{\mathbf{a}} = R(\mathbf{Y}\mathbf{a})$. After substituting $\mathbf{Y} = \mathbf{A}^q \mathbf{S}$ in the definition of R , we can evaluate $R_{\mathbf{a}}$ as

$$R_{\mathbf{a}} = \frac{1}{\lambda_1} \frac{(\mathbf{S}\mathbf{a})^T \mathbf{A}^{2q+1} (\mathbf{S}\mathbf{a})}{(\mathbf{S}\mathbf{a})^T \mathbf{A}^{2q} (\mathbf{S}\mathbf{a})}. \quad (3)$$

Since \mathbf{A} is real and symmetric, it has a real-valued eigen-decomposition $\mathbf{A} = \sum_{i=1}^n \lambda_i \mathbf{u}_i \mathbf{u}_i^T$, with $\{\mathbf{u}_i\}_{i=1}^n$ being orthonormal. Hence $\mathbf{A}^k = \sum_{i=1}^n \lambda_i^k \mathbf{u}_i \mathbf{u}_i^T$, for any $k \in \mathbb{N}$, and we further expand Equation (3) as

$$R_{\mathbf{a}} = \frac{1}{\lambda_1} \frac{\sum_i \lambda_i^{2q+1} \langle \mathbf{S}^T \mathbf{u}_i, \mathbf{a} \rangle^2}{\sum_i \lambda_i^{2q} \langle \mathbf{S}^T \mathbf{u}_i, \mathbf{a} \rangle^2} = \frac{\sum_i \alpha_i^{2q+1} \langle \mathbf{S}^T \mathbf{u}_i, \mathbf{a} \rangle^2}{\sum_i \alpha_i^{2q} \langle \mathbf{S}^T \mathbf{u}_i, \mathbf{a} \rangle^2}, \quad (4)$$

where $\alpha_i = \lambda_i / \lambda_1$, for all $i \in [n]$. This is well-defined since $\lambda_1 > 0$. For our analysis of $R(\hat{\mathbf{u}}) = \max_{\mathbf{a} \in \mathbb{S}^{d-1}} R_{\mathbf{a}}$, we first consider the case when \mathbf{A} is positive semidefinite (p.s.d.). The proof strategy and arguments serve as a building block for the indefinite case, discussed in Section 4.3.

4.2 Positive semidefinite matrices

Our first result, is a guarantee on the performance of RSVD, asserted by the following.

Theorem 1. *Let \mathbf{A} be a positive semidefinite matrix with $\lambda_1 > 0$ and $\hat{\mathbf{u}} = \text{RSVD}(\mathbf{A}, \mathcal{N}(0, 1)^{n \times d}, q, d)$ for any $q \in \mathbb{N}$. Then*

$$R(\hat{\mathbf{u}}) = \left(\Omega \left(\frac{d}{n} \right) \right)^{\frac{1}{2q+1}}$$

holds with probability at least $1 - e^{-\Omega(d)}$.

Proof. If \mathbf{A} is p.s.d. we have $\alpha_i \geq 0$, and thus (assuming $q \in \mathbb{N}$) we can repeatedly apply the Cauchy-Schwarz inequality to Equation (4) and get

$$R_{\mathbf{a}} \geq \frac{\sum_i \alpha_i^{2q} \langle \mathbf{S}^T \mathbf{u}_i, \mathbf{a} \rangle^2}{\sum_i \alpha_i^{2q-1} \langle \mathbf{S}^T \mathbf{u}_i, \mathbf{a} \rangle^2} \geq \dots \geq \frac{\sum_i \alpha_i \langle \mathbf{S}^T \mathbf{u}_i, \mathbf{a} \rangle^2}{\sum_i \langle \mathbf{S}^T \mathbf{u}_i, \mathbf{a} \rangle^2}. \quad (5)$$

The key observation is that by repeatedly using Equation (5) results in

$$\begin{aligned} \sum_{i=1}^n \langle \mathbf{S}^T \mathbf{u}_i, \mathbf{a} \rangle^2 &\geq R_{\mathbf{a}}^{-1} \sum_{i=1}^n \alpha_i \langle \mathbf{S}^T \mathbf{u}_i, \mathbf{a} \rangle^2 \geq \dots \\ &\geq R_{\mathbf{a}}^{-(2q+1)} \sum_{i=1}^n \alpha_i^{2q+1} \langle \mathbf{S}^T \mathbf{u}_i, \mathbf{a} \rangle^2 \end{aligned}$$

which implies

$$R_{\mathbf{a}}^{2q+1} \geq \frac{\sum_{i=1}^n \alpha_i^{2q+1} \langle \mathbf{S}^T \mathbf{u}_i, \mathbf{a} \rangle^2}{\sum_{i=1}^n \langle \mathbf{S}^T \mathbf{u}_i, \mathbf{a} \rangle^2} \geq \frac{\langle \mathbf{S}^T \mathbf{u}_1, \mathbf{a} \rangle^2}{\sum_{i=1}^n \langle \mathbf{S}^T \mathbf{u}_i, \mathbf{a} \rangle^2}. \quad (6)$$

Finally, by $R(\hat{\mathbf{u}}) = \max_{\mathbf{a} \in \mathbb{S}^{d-1}} R_{\mathbf{a}}$ and Definition 1 we have

$$R(\hat{\mathbf{u}})^{2q+1} \geq \max_{\mathbf{a} \in \mathbb{S}^{d-1}} \frac{\langle \mathbf{S}^T \mathbf{u}_1, \mathbf{a} \rangle^2}{\sum_{i=1}^n \langle \mathbf{S}^T \mathbf{u}_i, \mathbf{a} \rangle^2} = \cos^2 \theta(\mathbf{u}_1, \mathbf{S}), \quad (7)$$

and invoking Lemma 1 proves the claim. \square

We offer a few remarks. First note that the fact that Equation (7) implies Theorem 1 can be proven by estimating $R_{\mathbf{a}}$ only on $\mathbf{a} = \frac{\mathbf{S}^T \mathbf{u}_1}{\|\mathbf{S}^T \mathbf{u}_1\|_2}$, since we essentially prove Lemma 1 on such a vector \mathbf{a} — see our discussion in Section 3 or Appendix A.1. Second, Equation (6) can also be shown by Hölder's inequality — see a simplified proof of Theorem 1 in Appendix A.2. Third, from Theorem 1, we see that increasing the number of passes q makes $R(\hat{\mathbf{u}})$ approaching to 1 exponentially fast, while increasing the dimension d leads to stronger concentration of $R(\hat{\mathbf{u}})$ around the slowly increased mean $\Omega((d/n)^{1/(2q+1)})$. Finally, we have:

Remark 1. *The guarantee by Theorem 1 can be written as $R(\hat{\mathbf{u}}) = e^{-\mathcal{O}(\ln n/(2q+1))} \geq 1 - \mathcal{O}(\ln n/q)$, and hence, subsumes the result of Musco and Musco (2015).*

One may wonder if our analysis is tight. The next theorem confirms the tightness of Theorem 1 up to a constant factor.

Theorem 2. *For any $q \in \mathbb{N}$, there exists a positive semidefinite matrix \mathbf{A} with $\lambda_1 > 0$, so that for $\hat{\mathbf{u}} = \text{RSVD}(\mathbf{A}, \mathcal{N}(0, 1)^{n \times d}, q, d)$, it holds*

$$R(\hat{\mathbf{u}}) = \mathcal{O} \left(\left(\frac{d}{n} \right)^{\frac{1}{2q+1}} \right),$$

with probability at least $1 - e^{-\Omega(d)}$.

We prove Theorem 2 in Appendix A.2 by considering the following eigenvalue distribution $\{\alpha_i\}$:

$$1 = \alpha_1 > \alpha_2 = \dots = \alpha_n = \left(\frac{d}{n}\right)^{\frac{1}{2q+1}}. \quad (8)$$

While our worst-case analysis is tight, Equation (8) rarely happens in practice. Instead, real-world matrices are often observed to have rapidly decaying singular values (Chakrabarti and Faloutsos, 2006; Eikmeier and Gleich, 2017). To take this consideration into account, we introduce the following definition to capture whether \mathbf{A} has at least power-law decay of its singular values $\{\sigma_i\}_{i \geq i_0}^n$.

Definition 2. Let

$$i_0 = \begin{cases} \min_{j \in \mathcal{J}} j & \text{if } \mathcal{J} \neq \emptyset, \\ n & \text{otherwise,} \end{cases}$$

where $\mathcal{J} \subseteq [n]$ consists of all the integers $j \in [n]$ such that there exists $\gamma > 1/q$ and $C > 0$ satisfying $\sigma_i/\sigma_1 \leq C \cdot i^{-\gamma}$, for all $i \geq j$.

Theorem 3. Let \mathbf{A} be a positive semidefinite matrix, $\hat{\mathbf{u}} = \text{RSVD}(\mathbf{A}, \mathcal{N}(0, 1)^{n \times d}, q, d)$ for any $q \in \mathbb{N}$, and i_0 be defined as in Definition 2. Then

$$R(\hat{\mathbf{u}}) = \Omega\left(\left(\frac{d}{d + i_0}\right)^{\frac{1}{2q+1}}\right)$$

holds with probability at least $1 - e^{-\Omega(d)}$.

The proof of Theorem 3 can be found in Appendix A.2. The idea is to estimate $R_{\mathbf{a}}$ on $\mathbf{a} = \frac{\mathbf{S}^T \mathbf{u}_1}{\|\mathbf{S}^T \mathbf{u}_1\|_2}$ and check two possible cases. If i_0 is large, the analysis reduces to Theorem 1, while if i_0 is small, we invoke Bernstein-type inequalities and show that $R_{\mathbf{a}} = \Omega(1)$ with high probability. So, the overall guarantee of $R_{\mathbf{a}}$ is determined by the former case, and recalling $R(\hat{\mathbf{u}}) \geq \max_{\mathbf{a} \in \mathbb{S}^{d-1}} R_{\mathbf{a}}$ yields Theorem 3.

Remark 2. Theorem 3 subsumes Theorem 1 up to a constant factor as $d + i_0 = \mathcal{O}(n)$, and provides a much better guarantee if \mathbf{A} has singular values having at least power-law decay. In particular, if $i_0 = \mathcal{O}(d)$ then $R(\hat{\mathbf{u}}) = \Omega(1)$ with high probability, even with a single pass when $q = 1$ and $d = 1$.

4.3 Indefinite matrices

If \mathbf{A} has negative eigenvalues, the Inequality (5) in the proof of Theorem 1 is not valid anymore. Nevertheless, we expect to have a guarantee of $R(\hat{\mathbf{u}})$ similar to that of Theorem 1 if the negative eigenvalues are not too large. We introduce the following technical assumption.

Assumption 1. Assume there exists a constant $\kappa \in (0, 1]$ such that $\sum_{i=2}^n \lambda_i^{2q+1} \geq \kappa \sum_{i=2}^n |\lambda_i|^{2q+1}$.

An important observation is that Theorems 1 and 3 can be proved by estimating $R_{\mathbf{a}}$ only on one specific vector $\mathbf{a} = \frac{\mathbf{S}^T \mathbf{u}_1}{\|\mathbf{S}^T \mathbf{u}_1\|_2}$; see Section 4.2. Hence, it suffices to use the following lemma (proved in Appendix A.3) to generalize our results in Section 4.2 to indefinite matrices satisfying Assumption 1.

Lemma 2. Assume that matrix \mathbf{A} satisfies Assumption 1 and $\mathbf{S} \sim \mathcal{N}(0, 1)^{n \times d}$. There exists a constant $c_{\kappa} \in (0, 1]$ such that with probability at least $1 - e^{-\Omega(\sqrt{d}\kappa^2)}$, it holds

$$\sum_{i=1}^n \lambda_i^{2q+1} \langle \mathbf{S}^T \mathbf{u}_i, \mathbf{S}^T \mathbf{u}_1 \rangle^2 \geq c_{\kappa} \sum_{i=1}^n |\lambda_i|^{2q+1} \langle \mathbf{S}^T \mathbf{u}_i, \mathbf{S}^T \mathbf{u}_1 \rangle^2.$$

Lemma 2 essentially states that any indefinite matrix \mathbf{A} satisfying Assumption 1 has $R_{\mathbf{a}} = \Theta(\bar{R}_{\mathbf{a}})$ on such a vector $\mathbf{a} = \frac{\mathbf{S}^T \mathbf{u}_1}{\|\mathbf{S}^T \mathbf{u}_1\|_2}$, where

$$\bar{R}_{\mathbf{a}} = \frac{\sum_{i=1}^n |\lambda_i|^{2q+1} \langle \mathbf{S}^T \mathbf{u}_i, \mathbf{a} \rangle^2}{\sum_{i=1}^n \lambda_i^{2q} \langle \mathbf{S}^T \mathbf{u}_i, \mathbf{a} \rangle^2}. \quad (9)$$

The next theorem, proven in Appendix A.3, follows from Lemma 2 and the proof of Theorem 3.

Theorem 4. Assume that matrix \mathbf{A} satisfies Assumption 1. Let $\hat{\mathbf{u}} = \text{RSVD}(\mathbf{A}, \mathcal{N}(0, 1)^{n \times d}, q, d)$ for any $q \in \mathbb{N}$. Then,

$$R(\hat{\mathbf{u}}) = \Omega\left(c_{\kappa} \left(\frac{d}{d + i_0}\right)^{\frac{1}{2q+1}}\right),$$

with probability at least $1 - e^{-\Omega(\sqrt{d}\kappa^2)}$.

Remark 3. As discussed in Section 3, all the theorems shown in this section, i.e., Theorems 1, 2, 3, and 4, can be easily extended to any random matrix \mathbf{S} satisfying $\mathbf{S} \sim \text{Uniform}(\mathcal{G}_{n,d})$.

5 EXTENSION: COMBINING WITH PROJECTION FROM BERNOULLI

In this section, we propose an extension of Randomized SVD, which we name **RandSum**, and show as Algorithm 2. In **RandSum**, half of the columns of \mathbf{S} are replaced with i.i.d. samples from a Bernoulli distribution with mean $p \in (0, 1)$.⁴ We can show that the guarantee achieved by the **RandSum** algorithm for $R(\hat{\mathbf{u}})$ is no worse than that by the RSVD algorithm, since

⁴ $\text{Bernoulli}(p)^{n \times d}$ does not belong to the class of distributions mentioned in Section 3 to which Lemma 1 applies.

Algorithm 2: RandSum (\mathbf{A} , q , d , p)

$$\mathbf{S}_1 \sim \mathcal{N}(0, 1)^{n \times \lceil \frac{d}{2} \rceil}, \mathbf{S}_2 \sim \text{Bernoulli}(p)^{n \times \lfloor \frac{d}{2} \rfloor};$$

$$\mathbf{S} \leftarrow [\mathbf{S}_1 \quad \mathbf{S}_2];$$

return RSVD(\mathbf{A} , \mathbf{S} , q , d);

half of the columns of \mathbf{S} come from a normal distribution. To study the additional benefits due to the submatrix drawn from the Bernoulli, we derive the following lemma as an analog of Lemma 1 for a Bernoulli random matrix. The proof is in Appendix B.1.

Lemma 3. Let $\mathbf{v} \in \mathbb{S}^{n-1}$, $d \leq n/3$, and $\mathbf{S} \sim \text{Bernoulli}(p)^{n \times d}$ for a constant $p \in (0, 1)$. Then,

$$\cos^2 \theta(\mathbf{v}, \mathbf{S}) = \Omega\left(\frac{\max\{1, \langle \mathbf{v}, \mathbf{1}_n \rangle^2\}}{n}\right)$$

holds with probability at least $1 - e^{-\Omega(d)}$.

The next theorem, which holds for any p.s.d. matrix \mathbf{A} , is a direct consequence of Lemmas 1 and 3 and applying the techniques introduced in Theorem 1. The proof is in Appendix B.2.

Theorem 5. Let \mathbf{A} be a positive semidefinite matrix with $\lambda_1 > 0$, and $\hat{\mathbf{u}} = \text{RandSum}(\mathbf{A}, q, d, p)$ for any constant $p \in (0, 1)$ and integer $d \geq 2$. Then,

$$R(\hat{\mathbf{u}}) = \left(\Omega\left(\frac{\max\{d, \langle \mathbf{u}_1, \mathbf{1}_n \rangle^2\}}{n}\right) \right)^{\frac{1}{2q+1}}$$

holds with probability at least $1 - e^{-\Omega(d)}$.

Theorem 5 shows that $R(\hat{\mathbf{u}}) = \Theta(1)$ with high probability when $\langle \mathbf{u}_1, \mathbf{1}_n \rangle^2 = \Theta(n)$, which is achievable as the maximum possible value of $\langle \mathbf{u}_1, \mathbf{1}_n \rangle^2$ is n .

Remark 4. For certain tasks such as conflicting-group detection (Bonchi et al., 2019; Tzeng et al., 2020), one could expect to have large $\langle \mathbf{u}_1, \mathbf{1}_n \rangle^2$, since $\langle \mathbf{u}_1, \mathbf{1}_n \rangle^2$ naturally corresponds to the size of the subgraph, which is located by \mathbf{u}_1 .⁵ However, for tasks such as community detection, $\langle \mathbf{u}_1, \mathbf{1}_n \rangle^2 \approx 0$ is often the case.

Finally, we consider the generalization of Theorem 5 to indefinite matrices. To derive Lemma 4, the analog of Lemma 2 for Bernoulli random matrices, we introduce Assumption 2, where (i) is merely for the ease of presentation and (ii) generalizes Assumption 1 as $\xi_i = 1$ for $\mathbf{S} \sim \mathcal{N}(0, 1)^{n \times d}$. The proof of Lemma 4 can be found in Appendix B.3.

⁵We say that \mathbf{u}_1 is located around some indices $\mathcal{I} \subseteq [n]$ if the magnitude of $(\mathbf{u}_1)_i$ for any $i \in \mathcal{I}$ is much larger than those not in \mathcal{I} .

Assumption 2. Assume that (i) $\langle \mathbf{u}_1, \mathbf{1}_n \rangle^2 = \Omega(1)$ and (ii) there exists a constant $\kappa' \in (0, 1]$ such that

$$\sum_{i=2}^n \lambda_i^{2q+1} \xi_i \geq \kappa' \sum_{i=2}^n |\lambda_i|^{2q+1} \xi_i,$$

where $\xi_i = \mathbb{E} \left[\langle \mathbf{S}^T \mathbf{u}_i, \frac{\mathbf{1}_d}{\sqrt{d}} \rangle^2 \right]$, for all $i \in [n]$.

Lemma 4. Assume that \mathbf{A} satisfies Assumption 2. Let $\mathbf{S} \sim \text{Bernoulli}(p)^{n \times d}$ for a constant $p \in (0, 1)$. There exists a constant $c_{\kappa'} \in (0, 1]$, such that

$$\sum_{i=1}^n \lambda_i^{2q+1} \langle \mathbf{S}^T \mathbf{u}_i, \mathbf{S}^T \mathbf{u}_1 \rangle^2 \geq c_{\kappa'} \sum_{i=1}^n |\lambda_i|^{2q+1} \langle \mathbf{S}^T \mathbf{u}_i, \mathbf{S}^T \mathbf{u}_1 \rangle^2,$$

with probability at least $1 - e^{-\Omega(\sqrt{d}\kappa'^2)}$.

Our last result, Theorem 6, immediately follows from Theorems 4 and 5 and Lemma 4. The proof and the full version are in Appendix B.3.

Theorem 6. Assume that \mathbf{A} satisfies Assumptions 1 and 2. Let $\hat{\mathbf{u}} = \text{RandSum}(\mathbf{A}, q, d, p)$ for any constant $p \in (0, 1)$ and any $q \in \mathbb{N}$, and i_0 be defined as in Definition 2. Then,

$$R(\hat{\mathbf{u}}) = \Omega\left(\left(\max\left\{\frac{d}{d+i_0}, \frac{\langle \mathbf{u}_1, \mathbf{1}_n \rangle^2}{n}\right\}\right)^{\frac{1}{2q+1}}\right)$$

holds with probability at least $1 - e^{-\Omega(\sqrt{d})}$. (For the full dependency on κ , κ' , c_{κ} , and $c_{\kappa'}$, see Appendix B.3.)

6 EXPERIMENTS

In this section we evaluate the randomized algorithms we analyze in this paper using synthetic and real-world datasets. In Section 6.1, we use synthetic datasets to benchmark the RSVD algorithm with respect to the R measure, and study the effect of its parameters. In Section 6.2, we employ RSVD and RandSum as subroutines of spectral approaches for specific knowledge-discovery tasks on real-world datasets.

Settings. We use LanczosMethod, provided by the ARPACK library (Lehoucq et al., 1998), for computing λ_1 , which is required for measuring R . We fix $q = 1$ while varying $d \in \{1, 5, 10, 25, 50\}$ to study the effect of d , and fix $d = 10$ while varying $q \in \{1, 2, 4, 8, 16\}$ to study the effect of q . Each setting is repeated 100 times and the average is reported. All experiments are performed on an Intel Core i5 machine at 1.8 GHz with 8 GB RAM. All methods are implemented in Python 3.7.4.⁶

⁶The code is available at the github repo <https://bit.ly/34dI4N1>.

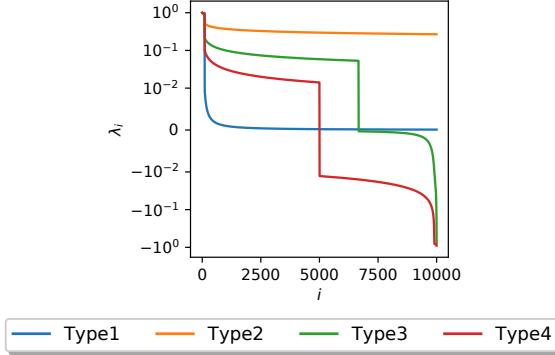


Figure 1: Different types of eigenvalue distributions.

6.1 Evaluation with synthetic data

We consider different types of eigenvalue distributions, also illustrated in Figure 1. The size of the input matrix is set to $n = 10\,000$ and $i_0 = 100$ (see Definition 2). For all types of synthetic matrices we set $\lambda_i = i^{-0.01}$, for $i < i_0$, and the rest of the eigenvalues $\{\lambda_i\}_{i \geq i_0}^n$ are specified as follows:

- Type 1: $\lambda_i = i^{-1}$ for $i \geq i_0$.
- Type 2: $\lambda_i = i^{-\frac{1}{7}}$ for $i \geq i_0$.
- Type 3: $\lambda_i = \begin{cases} i^{-\frac{1}{3}} & \text{if } i \in [i_0, \frac{2n}{3}], \\ -(i - \frac{2n}{3})^{-1} & \text{if } i > \frac{2n}{3}. \end{cases}$
- Type 4: $\lambda_i = \begin{cases} i^{-\frac{1}{2}} & \text{if } i \in [i_0, \frac{n}{2}], \\ -\frac{9}{10}(i - \frac{n}{2})^{-\frac{1}{2}} & \text{if } i \in (\frac{n}{2}, n - i_0), \\ -\frac{9}{10}i^{-0.01} & \text{if } i \geq n - i_0. \end{cases}$

For the value of κ in Assumption 1, we compute κ with $q = 1$ and get: $\kappa = 1$ for Type 1 and Type 2, $\kappa = 0.99$ for Type 3, and $\kappa = 0.22$ for Type 4. For each type of eigenvalue distribution, we generate a random $n \times n$ input matrix by sampling the eigenvectors uniformly from the space of orthogonal matrices.

Figure 2 shows the value of R for the vector $\hat{\mathbf{u}}$ computed by $\text{RSVD}(\mathbf{A}, \mathcal{N}(0, 1)^{n \times d}, q, d)$, and the speedup in running time against `LanczosMethod`.

For matrices of Type 1, it is expected that RSVD performs the best as the eigenvalues of such matrices have the fastest decay and $\kappa = 1$.

For matrices of Type 2, we notice that $R(\hat{\mathbf{u}})$ is very close to 1 when $q \geq 4$. This result is better than what our analysis predicts, since by Theorem 3 it holds that $R(\hat{\mathbf{u}}) = \Omega(1)$ with high probability after $q = 7$ (since the decay rate of Type 2 is $1/7$).

For matrices of Type 3, despite being indefinite, the magnitude of the negative eigenvalues is almost negligible ($\kappa = 0.99$). By Theorem 4 and Lemma 2, $R(\hat{\mathbf{u}})$ is

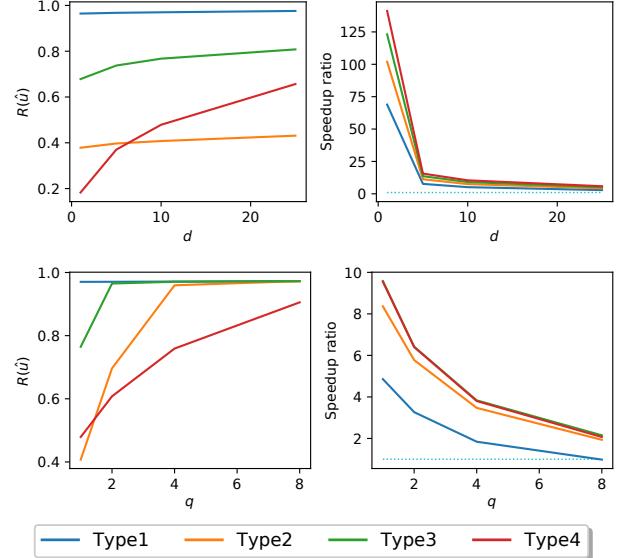


Figure 2: The value of $R(\hat{\mathbf{u}})$ for $\hat{\mathbf{u}}$ computed by $\text{RSVD}(\mathbf{A}, \mathcal{N}(0, 1)^{n \times d}, q, d)$. Top row shows dependence with d . Bottom row shows dependence with q . The speedup is measured against `LanczosMethod`.

nearly identical to its counterpart \bar{R} (see (9)), so it is expected that RSVD performs better on data of Type 3 than on data of Type 2, as the eigenvalue-distribution decay rate is faster.

For matrices of Type 4, although the eigenvalues decay faster than those of Type 3 matrices, the magnitudes of the negative eigenvalues are much larger ($\kappa = 0.22$). By Theorem 4 and Lemma 2, $R(\hat{\mathbf{u}})$ is upper-bounded by a factor of κ when increasing q , and the results indeed show that the performance of RSVD is worse for Type 4 matrices, compared to Type 3 ($\kappa = 0.99$).

6.2 Applications on real-world data

We use publicly-available networks from the SNAP collection (Leskovec and Krevl, 2014). Statistics of the datasets are listed in Tables 1 and 2.

6.2.1 Detection of 2 conflicting groups

The problem of 2-conflicting group detection aims to find two optimal groups that maximize the polarity objective $P(\mathbf{x}) = \mathbf{x}^T \mathbf{A} \mathbf{x} / \mathbf{x}^T \mathbf{x}$, where \mathbf{A} is the signed adjacency matrix and $\mathbf{x} \in \mathcal{T} = \{0, \pm 1\}^n \setminus \{\mathbf{0}_n\}$. Bonchi et al. (2019) propose a tight $\mathcal{O}(n^{1/2})$ -approximation algorithm based on the leading eigenvector \mathbf{u}_1 . In Appendix D we show that applying their approach on the approximated top-eigenvector $\hat{\mathbf{u}}$ yields an $\mathcal{O}(n^{1/2}/R(\hat{\mathbf{u}}))$ -approx algorithm.

Datasets. The statistics of datasets we use for this

Table 1: Datasets for conflicting group detection.

	WikiVot	Referendum	Slashdot	WikiCon
$ V $	7 115	10 884	82 140	116 717
$ E $	100 693	251 406	500 481	2 026 646
(γ, i_0)	(4.6, 15)	(4.5, 16)	(5.3, 17)	(2.8, 22)
κ	0.397	0.620	0.204	0.034
$\cos \theta(\mathbf{u}_1, \mathbf{1}_n)$	0.378	0.399	0.194	0.193

Table 2: Datasets for community detection.

	FBArtist	Gnutella31	YouTube	RoadCA
$ V $	50 515	62 586	1 134 890	1 965 206
$ E $	819 306	147 892	2 987 624	2 766 607

experiment are listed in Table 1. We observe that all datasets have rapidly-decaying singular values. To measure the parameters γ and i_0 (see Definition 2), due to memory limitations, we compute the top 1 000 eigenvalues (in magnitude) of its signed adjacency matrix by `LanczosMethod`, and fit the parameters (γ, i_0) by an MLE-based method (Clauset et al., 2009). Moreover, we test the validity of Assumption 1 by computing κ with $q = 1$, and also computing $\langle \mathbf{u}_1, \mathbf{1}_n \rangle$.

Results. Figure 3 illustrates the results obtained by applying the spectral algorithm of Bonchi et al. (2019) on the top-eigenvector $\hat{\mathbf{u}}$ returned by RSVD and RandSum. Due to the value of κ , the result is that, as expected, both algorithms perform the best on Referendum. Due to the value of $\cos \theta(\mathbf{u}_1, \mathbf{1}_n)$, the superiority of RandSum over RSVD is, as expected, more pronounced on WikiVot and Referendum than on Slashdot and WikiCon.

6.2.2 Detection of 2 communities

For the task of detecting two communities in a graph, Newman (2006) proposed an efficient algorithm by maximizing the modularity score $Q(\mathbf{x}) = \mathbf{x}^T \mathbf{M} \mathbf{x} / 4|E|$, where $\mathbf{M}_{i,j} = \mathbf{A}_{i,j} - \deg(i) \deg(j)/2|E|$, \mathbf{A} is the adjacency matrix of the input graph, and the two communities are determined by the sign of the top eigenvector of \mathbf{M} .

Datasets. The datasets used for evaluating this task are listed in Table 2. As the modularity matrix \mathbf{M} is dense and the networks are large, `LanczosMethod` runs out-of-memory on our machine when trying to compute the top eigenvalues, and hence, unlike Table 1, the number κ and the parameters (γ, i_0) are not displayed in Table 2.

Results. Figure 4 shows the results by applying the spectral algorithm of Newman (2006) on the top-eigenvector $\hat{\mathbf{u}}$ returned by RSVD and RandSum. Notice that on this task, RandSum has no advantage over

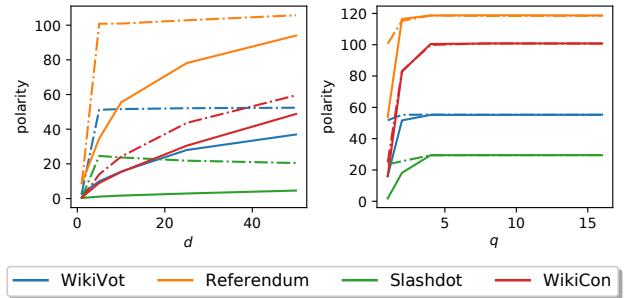


Figure 3: Results on the task of detecting 2 conflicting groups. Results for RSVD (resp. RandSum) are plotted with a solid (resp. dashed) line.

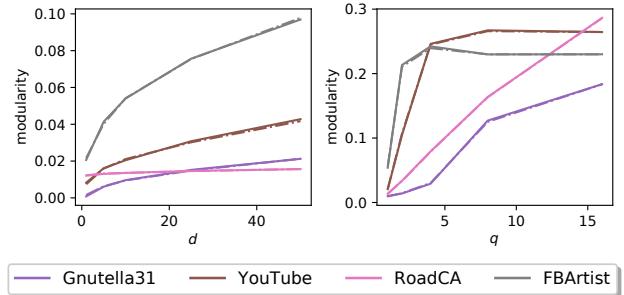


Figure 4: Results on the task of detecting 2 communities. Results for RSVD (in solid line) and RandSum (in dotted line) are nearly the same.

RSVD since $\mathbf{M}\mathbf{1}_n = 0$, and thus $\langle \mathbf{u}_1, \mathbf{1}_n \rangle = \mathbf{0}_n$ if $\lambda_1 \geq 0$. When fixing $d = 10$ and increasing q , the modularity scores converge much faster on FBArtist and YouTube than on Gnutella31 and RoadCA, suggesting that it could be hard to discover community structures in Gnutella31 and RoadCA. This is an expected result. For Gnutella (Gnutella31) the design of the network prevents the formation of large communities so as to enable reliable communication. For the road network of California (RoadCA) the reason is the grid-like structure of the network (Leskovec et al., 2009).

7 CONCLUSION

In this paper, we study the problem of approximating the leading eigenvector of a matrix with limited number of passes. The problem is of interest in many applications. We provide a tight theoretical analysis of the popular randomized SVD method, with respect to the metric $R(\hat{\mathbf{u}}) = \lambda_1^{-1} \hat{\mathbf{u}}^T \mathbf{A} \hat{\mathbf{u}} / \hat{\mathbf{u}}^T \hat{\mathbf{u}}$. Our results substantially improve the analysis of randomized SVD in the regime of $o(\ln n)$ passes and recover the analysis of prior works in the regime of $\Omega(\ln n)$ passes. A new technique is introduced to transform the problem of maximizing $R(\hat{\mathbf{u}})$ into a well-studied problem in the

literature of high-dimensional probability.

Our work opens several interesting directions. First, it is an open problem to characterize the fundamental limit of maximizing $R(\hat{\mathbf{u}})$ for any algorithm with fixed number of pass and $\mathcal{O}(n)$ space. Second, our results may be extended in different ways. For example, we may relax the requirement on the input matrix from symmetric to stochastic, so as to analyze approximations of PageRank (Page et al., 1999). Or, we may extend `RandSum` to use any non-centered sub-gaussian distribution for drawing \mathbf{S}_2 , and we conjecture this yields similar results. Another direction is to extend our analysis to top- k eigenvectors; since there are already several methods for computing top- k eigenvectors (Halko et al., 2011b; Mackey, 2008; Allen-Zhu and Li, 2016), the most challenging part is to define the proper metric to maximize, as a generalization of $R(\hat{\mathbf{u}})$.

Acknowledgements

We thank the anonymous reviewers for their insightful feedback. This research is supported by the ERC Advanced Grant REBOUND (834862), the EC H2020 RIA project SoBigData++ (871042), and the Wallenberg AI, Autonomous Systems and Software Program (WASP) funded by the Knut and Alice Wallenberg Foundation.

References

- Amirali Abdullah, Alexandr Andoni, Ravindran Kannan, and Robert Krauthgamer. Spectral approaches to nearest neighbor search. In *Proc. of FOCS*. IEEE, 2014.
- Dimitris Achlioptas. Database-friendly random projections. In *Proc. of PODS*, 2001.
- Zeyuan Allen-Zhu and Yuanzhi Li. Lazysvd: Even faster svd decomposition yet without agonizing pain. In *Proc. of NeurIPS*, 2016.
- Aris Anagnostopoulos, Luca Becchetti, Adriano Fazzzone, Cristina Menghini, and Chris Schwiegelshohn. Spectral relaxations and fair densest subgraphs. In *Proc. of CIKM*, 2020.
- Francesco Bonchi, Edoardo Galimberti, Aristides Gionis, Bruno Ordozgoiti, and Giancarlo Ruffo. Discovering polarized communities in signed networks. In *Proc. of CIKM*, 2019.
- Deepayan Chakrabarti and Christos Faloutsos. Graph mining: Laws, generators, and algorithms. *ACM computing surveys (CSUR)*, 2006.
- Xixian Chen, Irwin King, and Michael R Lyu. Frosh: Faster online sketching hashing. In *Proc. of UAI*, 2017.
- Aaron Clauset, Cosma Rohilla Shalizi, and Mark EJ Newman. Power-law distributions in empirical data. *SIAM review*, 2009.
- IBM Research Corporation. libskylark: Sketching-based distributed matrix computations for machine learning. <https://github.com/xdata-skylark/libskylark>, 2021. [Online; accessed 27-April-2021].
- Petros Drineas, Ilse CF Ipsen, Eugenia-Maria Kontopoulou, and Malik Magdon-Ismail. Structural convergence results for approximation of dominant subspaces from block krylov spaces. *SIAM Journal on Matrix Analysis and Applications*, 2018.
- Nicole Eikmeier and David F Gleich. Revisiting power-law distributions in spectra of real world networks. In *Proc of SIGKDD*, 2017.
- N Benjamin Erichson, Sergey Voronin, Steven L Brunton, and J Nathan Kutz. Randomized matrix decompositions using r. *Journal of Statistical Software*, 2019.
- Mina Ghashami, Edo Liberty, Jeff M Phillips, and David P Woodruff. Frequent directions: Simple and deterministic matrix sketching. *SIAM Journal on Scientific Computing*, 2016.
- Allan Gut. Multivariate random variables. In *An Intermediate Course in Probability*. Springer, 2009.
- Nathan Halko, Per-Gunnar Martinsson, Yoel Shkolnisky, and Mark Tygert. An algorithm for the principal component analysis of large data sets. *SIAM Journal on Scientific Computing*, 2011a.
- Nathan Halko, Per-Gunnar Martinsson, and Joel A Tropp. Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions. *SIAM review*, 2011b.
- Moritz Hardt and Eric Price. The noisy power method: a meta algorithm with applications. In *Proc. of NeurIPS*, 2014.
- Samuel B Hopkins, Tselil Schramm, Jonathan Shi, and David Steurer. Fast spectral algorithms from sum-of-squares proofs: tensor decomposition and planted sparse vectors. In *Proc. of STOC*, 2016.
- Roger A Horn, Roger A Horn, and Charles R Johnson. *Topics in matrix analysis*. Cambridge university press, 1994.
- Zengfeng Huang. Near optimal frequent directions for sketching dense and sparse matrices. In *Proc. of ICML*. PMLR, 2018.

- Ian T Jolliffe. Principal components in regression analysis. In *Principal component analysis*. Springer, 1986.
- Ravindran Kannan and Santosh Vempala. *Spectral algorithms*. Now Publishers Inc., 2009.
- Jacek Kuczyński and Henryk Woźniakowski. Estimating the largest eigenvalue by the power and lanczos algorithms with a random start. *SIAM Journal on Matrix Analysis and Applications*, 1992.
- Richard B Lehoucq, Danny C Sorensen, and Chao Yang. *ARPACK users' guide: solution of large-scale eigenvalue problems with implicitly restarted Arnoldi methods*. SIAM, 1998.
- Jure Leskovec and Andrej Krevl. SNAP Datasets: Stanford large network dataset collection. <http://snap.stanford.edu/data>, 2014.
- Jure Leskovec, Kevin J Lang, Anirban Dasgupta, and Michael W Mahoney. Community structure in large networks: Natural cluster sizes and the absence of large well-defined clusters. *Internet Mathematics*, 2009.
- Antoine Liutkus. randomized singular value decomposition in matlab central file exchange. 47835-randomized-singular-value-decomposition, 2021. [Online; accessed 27-April-2021].
- Lester Mackey. Deflation methods for sparse pca. In *Proc. of NIPS*, 2008.
- Michael W Mahoney et al. Randomized algorithms for matrices and data. *Foundations and Trends® in Machine Learning*, 2011.
- Per-Gunnar Martinsson and Joel A Tropp. Randomized numerical linear algebra: Foundations and algorithms. *Acta Numerica*, 2020.
- Cameron Musco and Christopher Musco. Randomized block krylov methods for stronger and faster approximate singular value decomposition. In *Proc. of NeurIPS*, 2015.
- Cameron Musco and David P Woodruff. Sublinear time low-rank approximation of positive semidefinite matrices. In *Proc. of FOCS*. IEEE, 2017.
- Mark EJ Newman. Modularity and community structure in networks. *Proc. of NAS*, 2006.
- Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. The pagerank citation ranking: Bringing order to the web. Technical report, Stanford InfoLab, 1999.
- Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. Scikit-learn: Machine learning in python. *JMLR*, 2011.
- Radim Řehůrek and Petr Sojka. Software framework for topic modelling with large corpora. In *Proc. of LREC 2010 Workshop*, 2010. https://radimrehurek.com/gensim/models/lsimodel.html#gensim.models.lsimodel.stochastic_svd.
- Mark Rudelson and Roman Vershynin. Smallest singular value of a random rectangular matrix. *Communications on Pure and Applied Mathematics*, 2009.
- Mark Rudelson and Roman Vershynin. Non-asymptotic theory of random matrices: extreme singular values. In *Proc. of International Congress of Mathematicians*, 2010.
- Arlei Silva, Ambuj Singh, and Ananthram Swami. Spectral algorithms for temporal graph cuts. In *Proc. of WWW*, 2018.
- Max Simchowitz, Ahmed El Alaoui, and Benjamin Recht. Tight query complexity lower bounds for pca via finite sample deformed wigner law. In *Proc. of STOC*, 2018.
- G. W. Stewart and Ji Guang Sun. *Matrix Perturbation Theory*. Academic Press, 1990.
- Pascal Terray and Françoise Pinsard. statpack: Eig procedures. https://terray.locean-ipsl.upmc.fr/statpack2.1/manuals/Module_Eig_Procedures.html, 2021. [Online; accessed 17-September-2021].
- Ruo-Chun Tzeng, Bruno Ordozgoiti, and Aristides Giannis. Discovering conflicting groups in signed networks. In *Proc. of NeurIPS*, 2020.
- Roman Vershynin. *High-dimensional probability: An introduction with applications in data science*. Cambridge university press, 2018.
- David P Woodruff et al. Sketching as a tool for numerical linear algebra. *Foundations and Trends® in Theoretical Computer Science*, 2014.
- Peng Xu, Bryan He, Christopher De Sa, Ioannis Mitliagkas, and Chris Re. Accelerated stochastic power iteration. In *Proc. of AISTATS*. PMLR, 2018.
- Yi Yu, Tengyao Wang, and Richard J Samworth. A useful variant of the davis–kahan theorem for statisticians. *Biometrika*, 2015.

Supplementary Material: Improved analysis of randomized SVD for top-eigenvector approximation

A Proofs of RSVD

A.1 Large deviation of projection length for Gaussian random matrix

This subsection is devoted to proving Lemma 1 restated below.

Lemma 1. *Let $\mathbf{v} \in \mathbb{R}^n$ be a nonzero vector and $\mathbf{S} \sim \mathcal{N}(0, 1)^{n \times d}$ where $n, d \in \mathbb{N}$ and $n \geq d$. Then,*

$$\cos^2 \theta(\mathbf{v}, \mathbf{S}) = \Theta\left(\frac{d}{n}\right)$$

with probability at least $1 - e^{-\Omega(d)}$.

This lemma stems from the observations that $\frac{\sigma_1(\mathbf{S}^T \mathbf{v})}{\sigma_1(\mathbf{S})} \leq \cos \theta(\mathbf{v}, \mathbf{S}) \leq \frac{\sigma_1(\mathbf{S}^T \mathbf{v})}{\sigma_d(\mathbf{S})}$ and the distribution of $\frac{\mathbf{S}^T \mathbf{v}}{\|\mathbf{v}\|_2}$ is exactly $\mathcal{N}(0, 1)^{d \times 1}$. The proof relies on the union bound of concentration inequalities on the extreme singular values of Gaussian random matrix, Lemma 5, and Lemma 6. Similar inequalities shown in the previous works, e.g. Hardt and Price (2014), also rely on this observation.

Lemma 5 (Theorem 4.4.5 (Vershynin, 2018)). *Let \mathbf{S} be a $n \times d$ random matrix whose entries are i.i.d. zero-mean subgaussian r.v.'s.*

$$\text{For all } t > 0, \quad \mathbb{P}\left[\sigma_1(\mathbf{S}) \geq c(\sqrt{n} + \sqrt{d} + t)\right] \leq 2e^{-t^2},$$

where $c > 0$ depends linearly only on $\|\mathbf{S}_{1,1}\|_{\psi_2}$ (see Definition 4 of ψ_2 -norm in Appendix C.1).

Lemma 6 (Theorem 1.1 (Rudelson and Vershynin, 2009)). *Let \mathbf{S} be a $n \times d$ random matrix whose entries are i.i.d. zero-mean subgaussian r.v.'s and $n \geq d$.*

$$\text{For all } \delta > 0, \quad \mathbb{P}\left[\sigma_d(\mathbf{S}) \leq \delta(\sqrt{n} - \sqrt{d-1})\right] \leq (c_1 \delta)^{n-d+1} + e^{-c_2 n},$$

where $c_1, c_2 > 0$ have polynomial dependence on $\|\mathbf{S}_{1,1}\|_{\psi_2}$ (see Definition 4 of ψ_2 -norm in Appendix C.1).

Proof of Lemma 1: For the simplicity of presentation, we assume $\|\mathbf{v}\|_2 = 1$ as $\cos \theta(\cdot, \cdot)$ is scale-invariant.

(i) $\cos \theta(\mathbf{v}, \mathbf{S}) = \Omega(\sqrt{d/n})$:

Recall that $\cos \theta(\mathbf{v}, \mathbf{S}) = \max_{\mathbf{a} \in \mathbb{S}^{d-1}} \frac{\langle \mathbf{v}, \mathbf{S}\mathbf{a} \rangle}{\|\mathbf{S}\mathbf{a}\|_2}$. Let $\mathbf{a} = \mathbf{S}^T \mathbf{v} / \|\mathbf{S}^T \mathbf{v}\|$. We get

$$\cos \theta(\mathbf{v}, \mathbf{S}) \geq \frac{\langle \mathbf{v}, \mathbf{S}\mathbf{S}^T \mathbf{v} \rangle}{\|\mathbf{S}\mathbf{S}^T \mathbf{v}\|_2} = \frac{\|\mathbf{S}^T \mathbf{v}\|_2^2}{\|\mathbf{S}\mathbf{S}^T \mathbf{v}\|_2} \geq \frac{\|\mathbf{S}^T \mathbf{v}\|_2}{\sigma_1(\mathbf{S})} = \frac{\sigma_1(\mathbf{S}^T \mathbf{v})}{\sigma_1(\mathbf{S})},$$

where the second inequality directly follows from the definitions of the largest singular value. Because $\mathbf{S}^T \mathbf{v} \sim \mathcal{N}(0, 1)^{d \times 1}$, invoking Lemma 6 with $\delta = e^{-1}$ yields that $\mathbb{P}\left[\sigma_1(\mathbf{S}^T \mathbf{v}) \geq \sqrt{d}/e\right] \geq 1 - e^{-\Omega(d)}$. Meanwhile, Lemma 5

with $t = \sqrt{n} - \sqrt{d}$ implies that $\mathbb{P}[\sigma_1(\mathbf{S}) \leq 2c\sqrt{n}] \geq 1 - e^{-\Omega(n)}$. We hence conclude (i) by applying the union bound.

(ii) $\cos \theta(\mathbf{v}, \mathbf{S}) = \mathcal{O}(\sqrt{d/n})$:

Due to $\sigma_d(\mathbf{S}) \leq \|\mathbf{S}\|_2$ and $\langle \mathbf{v}, \mathbf{S}\mathbf{a} \rangle \leq \|\mathbf{S}^T \mathbf{v}\|_2 \|\mathbf{a}\|_2 = \sigma_1(\mathbf{S}^T \mathbf{v})$, for all $\mathbf{a} \in \mathbb{S}^{d-1}$,

$$\cos \theta(\mathbf{v}, \mathbf{S}) = \max_{\mathbf{a} \in \mathbb{S}^{d-1}} \frac{\langle \mathbf{v}, \mathbf{S}\mathbf{a} \rangle}{\|\mathbf{S}\mathbf{a}\|_2} \leq \frac{\sigma_1(\mathbf{S}^T \mathbf{v})}{\sigma_d(\mathbf{S})}.$$

For the denominator, Lemma 6 with $\delta = e^{-1}$ is applied to permit that $\mathbb{P}[\sigma_d(\mathbf{S}) \geq \frac{\sqrt{n}-\sqrt{d-1}}{e}] \geq 1 - e^{-\Omega(n-d+1)} - e^{-\Omega(n)}$. For the numerator, as $\mathbf{S}^T \mathbf{v} \sim \mathcal{N}(0, 1)^{d \times 1}$, Lemma 5 with $t = \sqrt{d}$ shows that $\mathbb{P}[\sigma_1(\mathbf{S}^T \mathbf{v}) \leq 2\sqrt{d}] \geq 1 - e^{-\Omega(d)}$. Thus, (ii) holds by applying the union bound. \square

A.2 RSVD with positive semidefinite matrices

Lemma 7. Let $\mathbf{x} = (\mathbf{x}_1, \dots, \mathbf{x}_n)$ and $\mathbf{y} = (\mathbf{y}_1, \dots, \mathbf{y}_n)$ be two vectors in \mathbb{R}^n satisfying (i) there exists $i \in [n]$ s.t. $\mathbf{x}_i \mathbf{y}_i \neq 0$, and (ii) there exists $j \in [n]$ s.t. $\mathbf{y}_j \neq 0$. Then for all $q \in \mathbb{N}$,

$$\frac{\sum_{i=1}^n |\mathbf{x}_i|^{2q+1} \mathbf{y}_i^2}{\sum_{i=1}^n |\mathbf{x}_i|^{2q} \mathbf{y}_i^2} \geq \left(\frac{\sum_{i=1}^n |\mathbf{x}_i|^{2q} \mathbf{y}_i^2}{\sum_{i=1}^n \mathbf{y}_i^2} \right)^{\frac{1}{2q}}.$$

Proof For any n -dimensional vectors $\mathbf{a} = (\mathbf{a}_1, \dots, \mathbf{a}_n)$, $\mathbf{v} = (\mathbf{b}_1, \dots, \mathbf{b}_n)$ satisfying that (i)' there exists $i \in [n]$ s.t. $\mathbf{a}_i \mathbf{b}_i \neq 0$, and (ii)' there exists $j \in [n]$ s.t. $\mathbf{b}_j \neq 0$, Hölder's inequality implies that

$$\left(\frac{\sum_{i=1}^n |\mathbf{a}_i|^r}{\sum_{i=1}^n |\mathbf{a}_i \mathbf{b}_i|} \right)^{\frac{1}{r}} \geq \left(\frac{\sum_{i=1}^n |\mathbf{a}_i \mathbf{b}_i|}{\sum_{i=1}^n |\mathbf{b}_i|^s} \right)^{\frac{1}{s}}, \quad (10)$$

where $r, s \in [1, \infty]$ with $1/r + 1/s = 1$. Let $\mathbf{a}_i = |\mathbf{x}_i|^{2q} \mathbf{y}_i^{2/r}$ and $\mathbf{b}_i = |\mathbf{y}_i|^{2/s}$, for all $i \in [n]$, then (i) and (ii) imply (i)' and (ii)' respectively. Hence, (10) with $r = (2q+1)/2q$, $s = 2q+1$ gives us that

$$\left(\frac{\sum_{i=1}^n \left(|\mathbf{x}_i|^{2q} \mathbf{y}_i^{\frac{4q}{2q+1}} \right)^{\frac{2q+1}{2q}}}{\sum_{i=1}^n |\mathbf{x}_i|^{2q} \mathbf{y}_i^2} \right)^{\frac{2q}{2q+1}} \geq \left(\frac{\sum_{i=1}^n |\mathbf{x}_i|^{2q} \mathbf{y}_i^2}{\sum_{i=1}^n \left(\mathbf{y}_i^{\frac{2}{2q+1}} \right)^{2q+1}} \right)^{\frac{1}{2q+1}}.$$

We conclude this lemma by rearranging the above inequality. \square

Theorem 1. Let \mathbf{A} be a positive semidefinite matrix with $\lambda_1 > 0$ and $\hat{\mathbf{u}} = \text{RSVD}(\mathbf{A}, \mathcal{N}(0, 1)^{n \times d}, q, d)$ for any $q \in \mathbb{N}$. Then,

$$R(\hat{\mathbf{u}}) = \left(\Omega \left(\frac{d}{n} \right) \right)^{\frac{1}{2q+1}}$$

holds with probability at least $1 - e^{-\Omega(d)}$.

Proof Thanks to Lemma 1, the proof follows if the following inequality holds almost surely

$$R(\hat{\mathbf{u}})^{2q+1} \geq \max_{\mathbf{a} \in \mathbb{S}^{d-1}} \frac{\langle \mathbf{S}^T \mathbf{u}_1, \mathbf{a} \rangle^2}{\sum_{i=1}^n \langle \mathbf{S}^T \mathbf{u}_i, \mathbf{a} \rangle^2} = \cos^2 \theta(\mathbf{u}_1, \mathbf{S}), \quad (11)$$

where the equation is due to Definition 1. We show (11) by Lemma 7 and the alternating form of $R(\hat{\mathbf{u}})$ follows by (4) in Section 4.2,

$$R(\hat{\mathbf{u}}) = \max_{\mathbf{a} \in \mathbb{S}^{d-1}} R_{\mathbf{a}} = \max_{\mathbf{a} \in \mathbb{S}^{d-1}} \frac{\sum_{i=1}^n \alpha_i^{2q+1} \langle \mathbf{S}^T \mathbf{u}_i, \mathbf{a} \rangle^2}{\sum_{i=1}^n \alpha_i^{2q} \langle \mathbf{S}^T \mathbf{u}_i, \mathbf{a} \rangle^2}. \quad (12)$$

Let $\mathbf{x}_i = \alpha_i$ and $\mathbf{y}_i = \langle \mathbf{S}^T \mathbf{u}_i, \mathbf{a} \rangle$, for all $i \in [n]$, because $\langle \mathbf{S}^T \mathbf{u}_1, \mathbf{a} \rangle \neq 0$ a.e., the conditions of Lemma 7, (i) and (ii)., hold a.e.. Therefore, it holds almost surely that

$$\begin{aligned} R_{\mathbf{a}} &= \frac{\sum_{i=1}^n \alpha_i^{2q+1} \langle \mathbf{S}^T \mathbf{u}_i, \mathbf{a} \rangle^2}{\sum_{i=1}^n \alpha_i^{2q} \langle \mathbf{S}^T \mathbf{u}_i, \mathbf{a} \rangle^2} \geq \left(\frac{\sum_{i=1}^n \alpha_i^{2q} \langle \mathbf{S}^T \mathbf{u}_i, \mathbf{a} \rangle^2}{\sum_{i=1}^n \langle \mathbf{S}^T \mathbf{u}_i, \mathbf{a} \rangle^2} \right)^{\frac{1}{2q}} \\ &= \left(\frac{\sum_{i=1}^n \alpha_i^{2q} \langle \mathbf{S}^T \mathbf{u}_i, \mathbf{a} \rangle^2}{\sum_{i=1}^n \alpha_i^{2q+1} \langle \mathbf{S}^T \mathbf{u}_i, \mathbf{a} \rangle^2} \frac{\sum_{i=1}^n \alpha_i^{2q+1} \langle \mathbf{S}^T \mathbf{u}_i, \mathbf{a} \rangle^2}{\sum_{i=1}^n \langle \mathbf{S}^T \mathbf{u}_i, \mathbf{a} \rangle^2} \right)^{\frac{1}{2q}} = \left(R_{\mathbf{a}}^{-1} \frac{\sum_{i=1}^n \alpha_i^{2q+1} \langle \mathbf{S}^T \mathbf{u}_i, \mathbf{a} \rangle^2}{\sum_{i=1}^n \langle \mathbf{S}^T \mathbf{u}_i, \mathbf{a} \rangle^2} \right)^{\frac{1}{2q}}, \end{aligned}$$

where the last equation follows from (4) in Section 4.2 again. Rearranging the above inequality, we get that

$$R_{\mathbf{a}}^{2q+1} \geq \frac{\sum_{i=1}^n \alpha_i^{2q+1} \langle \mathbf{S}^T \mathbf{u}_i, \mathbf{a} \rangle^2}{\sum_{i=1}^n \langle \mathbf{S}^T \mathbf{u}_i, \mathbf{a} \rangle^2} \geq \frac{\langle \mathbf{S}^T \mathbf{u}_1, \mathbf{a} \rangle^2}{\sum_{i=1}^n \langle \mathbf{S}^T \mathbf{u}_i, \mathbf{a} \rangle^2}, \text{ a.e.,} \quad (13)$$

where the second inequality is leveraged the fact that $\sum_{i \neq 1} \alpha_i^{2q+1} \langle \mathbf{S}^T \mathbf{u}_i, \mathbf{a} \rangle^2 \geq 0$. (13) and the definition $R(\hat{\mathbf{u}}) = \max_{\mathbf{a} \in \mathbb{S}^{d-1}} R_{\mathbf{a}}$ imply (11) as desired and hence the proof completes. \square

Theorem 2. For any $q \in \mathbb{N}$, there exists a positive semi-definite matrix \mathbf{A} with $\lambda_1 > 0$, so that for $\hat{\mathbf{u}} = \text{RSVD}(\mathbf{A}, \mathcal{N}(0, 1)^{n \times d}, q, d)$, it holds

$$R(\hat{\mathbf{u}}) = \mathcal{O}\left(\left(\frac{d}{n}\right)^{\frac{1}{2q+1}}\right),$$

with probability at least $1 - e^{-\Omega(d)}$.

Proof Let \mathbf{A} be a diagonal matrix with $\mathbf{A}_{1,1} = 1$ and $\mathbf{A}_{i,i} = (d/n)^{\frac{1}{2q+1}}$, for all $i \neq 1$. Apparently, $\mathbf{A} = \mathbf{e}_1^T \mathbf{e}_1 + \sum_{i=2}^n \alpha \mathbf{e}_i^T \mathbf{e}_i$, where $\alpha = (d/n)^{\frac{1}{2q+1}}$ and $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ is the canonical basis in \mathbb{R}^n . As discussed in Section 4, $R(\hat{\mathbf{u}}) = \max_{\mathbf{a} \in \mathbb{S}^{d-1}} R_{\mathbf{a}}$ and the alternating expression of $R_{\mathbf{a}}$, (4) in Section 4.2, can be rewritten as

$$\text{for all } \mathbf{a} \in \mathbb{S}^{d-1}, \quad R_{\mathbf{a}} = \frac{\langle \mathbf{S}^T \mathbf{e}_1, \mathbf{a} \rangle^2}{\langle \mathbf{S}^T \mathbf{e}_1, \mathbf{a} \rangle^2 + \sum_{i=2}^n \alpha^{2q} \langle \mathbf{S}^T \mathbf{e}_i, \mathbf{a} \rangle^2} + \frac{\sum_{i=2}^n \alpha^{2q+1} \langle \mathbf{S}^T \mathbf{e}_i, \mathbf{a} \rangle^2}{\langle \mathbf{S}^T \mathbf{e}_1, \mathbf{a} \rangle^2 + \sum_{i=2}^n \alpha^{2q} \langle \mathbf{S}^T \mathbf{e}_i, \mathbf{a} \rangle^2}. \quad (14)$$

On the one hand, as $1 > (d/n)^{\frac{2q}{2q+1}} = \alpha^{2q}$, the first term in (14) is upper bounded as:

$$\frac{\langle \mathbf{S}^T \mathbf{e}_1, \mathbf{a} \rangle^2}{\langle \mathbf{S}^T \mathbf{e}_1, \mathbf{a} \rangle^2 + \sum_{i=2}^n \alpha^{2q} \langle \mathbf{S}^T \mathbf{e}_i, \mathbf{a} \rangle^2} \leq \frac{\langle \mathbf{S}^T \mathbf{e}_1, \mathbf{a} \rangle^2}{\sum_{i=1}^n \alpha^{2q} \langle \mathbf{S}^T \mathbf{e}_i, \mathbf{a} \rangle^2} \leq \alpha^{-2q} \cos^2 \theta(\mathbf{e}_1, \mathbf{S}), \quad (15)$$

where the second inequality follows directly from the definition of $\cos^2 \theta(\mathbf{e}_1, \mathbf{S})$. On the other hand, the second term in (14) is upper bounded as:

$$\frac{\sum_{i=2}^n \alpha^{2q+1} \langle \mathbf{S}^T \mathbf{e}_i, \mathbf{a} \rangle^2}{\langle \mathbf{S}^T \mathbf{e}_1, \mathbf{a} \rangle^2 + \sum_{i=2}^n \alpha^{2q} \langle \mathbf{S}^T \mathbf{e}_i, \mathbf{a} \rangle^2} \leq \frac{\sum_{i=2}^n \alpha^{2q+1} \langle \mathbf{S}^T \mathbf{e}_i, \mathbf{a} \rangle^2}{\sum_{i=2}^n \alpha^{2q} \langle \mathbf{S}^T \mathbf{e}_i, \mathbf{a} \rangle^2} = \alpha. \quad (16)$$

By substituting (15) and (16) into (14), we derive that $R_{\mathbf{a}} \leq \alpha^{-2q} \cos^2 \theta(\mathbf{e}_1, \mathbf{S}) + \alpha$, for all $\mathbf{a} \in \mathbb{S}^{d-1}$, which provides an upper bound of $R(\hat{\mathbf{u}})$. Finally, invoking Lemma 1, which states that $\cos^2 \theta(\mathbf{e}_1, \mathbf{S}) = \Theta(d/n)$ with high probability, and recalling that $\alpha = (d/n)^{\frac{1}{2q+1}}$ yields the conclusion. \square

Theorem 3. Let \mathbf{A} be a positive semi-definite matrix, $\hat{\mathbf{u}} = \text{RSVD}(\mathbf{A}, \mathcal{N}(0, 1)^{n \times d}, q, d)$ for any $q \in \mathbb{N}$, and i_0 be defined as in Definition 2 in Section 4.2. Then

$$R(\hat{\mathbf{u}}) = \Omega\left(\left(\frac{d}{d+i_0}\right)^{\frac{1}{2q+1}}\right)$$

holds with probability at least $1 - e^{-\Omega(d)}$.

Proof If $i_0 = n$, then we subsume the result by Theorem 1 directly. Hence we assume that $i_0 < n$ below.

By applying Corollary 1 in Appendix C.2 with $\delta = \frac{1}{3}$, $\mathbf{x} = \mathbf{u}_1$, we have probability $1 - e^{-\Omega(d)}$ that

$$\frac{2d}{3} \leq \|\mathbf{S}^T \mathbf{u}_1\|_2^2 \leq \frac{4d}{3}, \quad (17)$$

which directly implies that $\mathbf{S}^T \mathbf{u}_1 \neq 0$. In the following, we consider (i). $\sum_{i=1}^{i_0} \alpha_i^{2q} \langle \mathbf{S}^T \mathbf{u}_i, \mathbf{S}^T \mathbf{u}_1 \rangle^2 > \sum_{i=i_0+1}^n \alpha_i^{2q} \langle \mathbf{S}^T \mathbf{u}_i, \mathbf{S}^T \mathbf{u}_1 \rangle^2$; (ii). otherwise. Then, we show the claimed lower bound in (i). and (ii). separately by invoking Lemma 12 in Appendix C.3, which gives the bounds for the weighted sum with high probability.

$$(i). \sum_{i=1}^{i_0} \alpha_i^{2q} \langle \mathbf{S}^T \mathbf{u}_i, \mathbf{S}^T \mathbf{u}_1 \rangle^2 > \sum_{i=i_0+1}^n \alpha_i^{2q} \langle \mathbf{S}^T \mathbf{u}_i, \mathbf{S}^T \mathbf{u}_1 \rangle^2.$$

Roughly speaking in (i), the top i_0 terms dominate, hence one can expect the similar proof for Theorem 1 without the last $n - i_0$ terms will help us reason. The alternating form of $R(\hat{\mathbf{u}})$ follows by (4) in Section 4.1,

$$R(\hat{\mathbf{u}}) = \max_{\mathbf{a} \in \mathbb{S}^{d-1}} \frac{\sum_{i=1}^n \alpha_i^{2q+1} \langle \mathbf{S}^T \mathbf{u}_i, \mathbf{a} \rangle^2}{\sum_{i=1}^n \alpha_i^{2q} \langle \mathbf{S}^T \mathbf{u}_i, \mathbf{a} \rangle^2} \geq \frac{\sum_{i=1}^n \alpha_i^{2q+1} \langle \mathbf{S}^T \mathbf{u}_i, \mathbf{S}^T \mathbf{u}_1 \rangle^2}{\sum_{i=1}^n \alpha_i^{2q} \langle \mathbf{S}^T \mathbf{u}_i, \mathbf{S}^T \mathbf{u}_1 \rangle^2} > \frac{\sum_{i=1}^{i_0} \alpha_i^{2q+1} \langle \mathbf{S}^T \mathbf{u}_i, \mathbf{S}^T \mathbf{u}_1 \rangle^2}{2 \sum_{i=1}^{i_0} \alpha_i^{2q} \langle \mathbf{S}^T \mathbf{u}_i, \mathbf{S}^T \mathbf{u}_1 \rangle^2}, \quad (18)$$

where the first inequality comes from the fact that $\mathbf{S}^T \mathbf{u}_1 / \|\mathbf{S}^T \mathbf{u}_1\|_2 \in \mathbb{S}^{d-1}$ and the last one uses that $\sum_{i=i_0+1}^n \alpha_i^{2q+1} \langle \mathbf{S}^T \mathbf{u}_i, \mathbf{S}^T \mathbf{u}_1 \rangle^2 \geq 0$ and (i). From (18), we repeat the deduction of (6) in Section 4.1, by viewing $R_{\mathbf{a}}$ as $2R(\hat{\mathbf{u}})$, $\alpha_i = \alpha_i$ for $i = 1, \dots, i_0$, $\alpha_i = 0$ for $i > i_0$, and $\langle \mathbf{S}^T \mathbf{u}_i, \mathbf{S}^T \mathbf{a} \rangle^2 = \langle \mathbf{S}^T \mathbf{u}_i, \mathbf{S}^T \mathbf{u}_1 \rangle^2$ to conclude that (an alternative way is to use Lemma 7 as shown Appendix A.2)

$$\frac{\sum_{i=1}^{i_0} \alpha_i^{2q+1} \langle \mathbf{S}^T \mathbf{u}_i, \mathbf{S}^T \mathbf{u}_1 \rangle^2}{\sum_{i=1}^{i_0} \alpha_i^{2q} \langle \mathbf{S}^T \mathbf{u}_i, \mathbf{S}^T \mathbf{u}_1 \rangle^2} > \left((2R(\hat{\mathbf{u}}))^{-1} \frac{\sum_{i=1}^{i_0} \alpha_i^{2q+1} \langle \mathbf{S}^T \mathbf{u}_i, \mathbf{S}^T \mathbf{u}_1 \rangle^2}{\sum_{i=1}^{i_0} \langle \mathbf{S}^T \mathbf{u}_i, \mathbf{S}^T \mathbf{u}_1 \rangle^2} \right)^{\frac{1}{2q}}. \quad (19)$$

Rearranging the inequalities (18) and (19), we get

$$(2R(\hat{\mathbf{u}}))^{2q+1} > \frac{\langle \mathbf{S}^T \mathbf{u}_1, \mathbf{S}^T \mathbf{u}_1 \rangle^2}{\sum_{i=1}^{i_0} \langle \mathbf{S}^T \mathbf{u}_i, \mathbf{S}^T \mathbf{u}_1 \rangle^2} \geq \frac{4d^2}{16d^2 + 9 \sum_{1 \leq i \leq i_0} \langle \mathbf{S}^T \mathbf{u}_i, \mathbf{S}^T \mathbf{u}_1 \rangle^2}, \quad (20)$$

where the last inequality is a consequence of (17). By applying Lemma 12 in Appendix C.3 with $\epsilon = \frac{1}{3}$, $\delta = \frac{1}{3}$, $\beta_1 = \dots = \beta_{i_0} = 1$ and $\beta_{i_0+1} = \dots = \beta_n = 0$, then we have probability $1 - e^{-\Omega(d)}$ that $\sum_{1 \leq i \leq i_0} \langle \mathbf{S}^T \mathbf{u}_i, \mathbf{S}^T \mathbf{u}_1 \rangle^2 \leq \frac{16di_0}{9}$. Together with (20), the proof is derived by the union bound.

$$(ii). \sum_{i=1}^{i_0} \alpha_i^{2q} \langle \mathbf{S}^T \mathbf{u}_i, \mathbf{S}^T \mathbf{u}_1 \rangle^2 \leq \sum_{i=i_0+1}^n \alpha_i^{2q} \langle \mathbf{S}^T \mathbf{u}_i, \mathbf{S}^T \mathbf{u}_1 \rangle^2.$$

As $\mathbf{S}^T \mathbf{u}_1 / \|\mathbf{S}^T \mathbf{u}_1\|_2 \in \mathbb{S}^{d-1}$, (4) in Section 4.2 yields that

$$R(\hat{\mathbf{u}}) \geq \frac{\sum_{i=1}^n \alpha_i^{2q+1} \langle \mathbf{S}^T \mathbf{u}_i, \mathbf{S}^T \mathbf{u}_1 \rangle^2}{\sum_{i=1}^n \alpha_i^{2q} \langle \mathbf{S}^T \mathbf{u}_i, \mathbf{S}^T \mathbf{u}_1 \rangle^2} \geq \frac{\langle \mathbf{S}^T \mathbf{u}_1, \mathbf{S}^T \mathbf{u}_1 \rangle^2}{2 \sum_{i=i_0+1}^n \alpha_i^{2q} \langle \mathbf{S}^T \mathbf{u}_i, \mathbf{S}^T \mathbf{u}_1 \rangle^2} \geq \frac{2d^2}{9 \sum_{i=i_0+1}^n \alpha_i^{2q} \langle \mathbf{S}^T \mathbf{u}_i, \mathbf{S}^T \mathbf{u}_1 \rangle^2},$$

where the second inequality is due to (ii); the last is a result of (17). By Lemma 12 with $\delta = d$, $\epsilon = \frac{1}{2}$, $\beta_2 = \dots = \beta_{i_0} = 0$, and $\beta_i = \alpha_i^{2q}$ for all $i = i_0 + 1, \dots, n$, we have

$$\mathbb{P} \left[\sum_{i=i_0+1}^n \alpha_i^{2q} \langle \mathbf{S}^T \mathbf{u}_i, \mathbf{S}^T \mathbf{u}_1 \rangle^2 \leq \frac{3d(d+1)}{2} \sum_{i=i_0+1}^n \alpha_i^{2q} \right] \leq 1 - e^{-\Omega(d)}.$$

By Definition 2, since $\gamma > 1/q$,

$$\sum_{i=i_0+1}^n \alpha_i^{2q} \leq C \int_1^\infty x^{-2q\gamma} dx < C \int_1^\infty x^{-2} dx = C.$$

Hence, the union bound yields $R(\hat{\mathbf{u}}) = \Omega(1)$ with probability at least $1 - e^{-\Omega(d)}$. \square

A.3 RSVD with indefinite matrices

Assumption 1 is restated here for convenience.

Assumption 1. Assume there exists a constant $\kappa \in (0, 1]$ such that $\sum_{i=2}^n \lambda_i^{2q+1} \geq \kappa \sum_{i=2}^n |\lambda_i|^{2q+1}$.

Lemma 2. Assume that matrix \mathbf{A} satisfies Assumption 1 and $\mathbf{S} \sim \mathcal{N}(0, 1)^{n \times d}$. There exists a constant $c_\kappa \in (0, 1]$ such that

$$\mathbb{P} \left[\sum_{i=1}^n \lambda_i^{2q+1} \langle \mathbf{S}^T \mathbf{u}_i, \mathbf{S}^T \mathbf{u}_1 \rangle^2 \geq c_\kappa \sum_{i=1}^n |\lambda_i|^{2q+1} \langle \mathbf{S}^T \mathbf{u}_i, \mathbf{S}^T \mathbf{u}_1 \rangle^2 \right] \geq 1 - e^{-\Omega(\sqrt{d}\kappa^2)}.$$

Proof Recall $\alpha_i = \lambda_i / \lambda_1$ for all $i \in [n]$ and introduce $\mathcal{I}_+ = \{i \in [n] \setminus \{1\} : \alpha_i > 0\}$ and $\mathcal{I}_- = \{i \in [n] : \alpha_i < 0\}$. It is natural to assume $\mathcal{I}_- \neq \emptyset$, as otherwise this lemma trivially holds. Also, for simplicity $\kappa \in (0, 1]$ is assumed to be the number such that $\sum_{i=2}^n \alpha_i^{2q+1} = \kappa \sum_{i=2}^n |\alpha_i|^{2q+1}$ (this number can be found always).

Apparently in both sums of interest, $\sum_{i=1}^n \alpha_i^{2q+1} \langle \mathbf{S}^T \mathbf{u}_i, \mathbf{S}^T \mathbf{u}_1 \rangle^2$ and $\sum_{i=1}^n |\alpha_i|^{2q+1} \langle \mathbf{S}^T \mathbf{u}_i, \mathbf{S}^T \mathbf{u}_1 \rangle^2$, the largest single term is the first one, $\langle \mathbf{S}^T \mathbf{u}_1, \mathbf{S}^T \mathbf{u}_1 \rangle^2$, so our initial step is to derive its high probability bound. Applying Corollary 1 in Appendix C.2 with $\mathbf{x} = \mathbf{u}_1$ and $\delta = 1 - \sqrt{1 - \kappa/4}$ (resp. $\delta = \sqrt{1 + \kappa/4} - 1$) for the lower-tail (resp. upper-tail) yields

$$\mathbb{P} \left[d^2 \left(1 - \frac{\kappa}{4} \right) \leq \langle \mathbf{S}^T \mathbf{u}_1, \mathbf{S}^T \mathbf{u}_1 \rangle^2 \leq d^2 \left(1 + \frac{\kappa}{4} \right) \right] \geq 1 - e^{-\Omega(d\kappa^2)}. \quad (21)$$

However, as the other terms highly depend on the decay rate of eigenvalues, to derive a high probability bound, we need to carefully choose the parameters when applying concentration inequalities. In what follows, we define $s_+ = \sum_{i \in \mathcal{I}_+} \alpha_i^{2q+1}$ and $s_- = \sum_{i \in \mathcal{I}_-} |\alpha_i|^{2q+1}$, and then prove in two cases: either (i). $s_- = \Omega(\sqrt{d})$ or (ii). $s_- = o(\sqrt{d})$.

(i). $s_- = \Omega(\sqrt{d})$. Applying the lower-tail (resp. upper tail) of Lemma 12 in Appendix C.2 with $\delta = \epsilon = 1 - \sqrt{1 - \kappa/2}$ (resp. $\delta = \epsilon = \sqrt{1 + \kappa/2} - 1$), $\beta_i = \alpha_i$ for $i \in \mathcal{I}_+$, and $\beta_i = 0$ otherwise, we get

$$\mathbb{P} \left[d \left(1 - \frac{\kappa}{2} \right) s_+ \leq \sum_{i \in \mathcal{I}_+} \alpha_i^{2q+1} \langle \mathbf{S}^T \mathbf{u}_i, \mathbf{S}^T \mathbf{u}_1 \rangle^2 \leq d \left(1 + \frac{\kappa}{2} \right) s_+ \right] \geq 1 - e^{-\Omega(\sqrt{d}\kappa^2)}. \quad (22)$$

In addition, using Lemma 12 with $\delta = \epsilon = \sqrt{1 + \kappa/2} - 1$ in Appendix C.2, $\beta_i = |\alpha_i|$ for $i \in \mathcal{I}_-$, and $\beta_i = 0$ otherwise, we derive

$$\mathbb{P} \left[\sum_{i \in \mathcal{I}_-} |\alpha_i|^{2q+1} \langle \mathbf{S}^T \mathbf{u}_i, \mathbf{S}^T \mathbf{u}_1 \rangle^2 \leq d \left(1 + \frac{\kappa}{2} \right) s_- \right] \geq 1 - e^{-\Omega(\sqrt{d}\kappa^2)}. \quad (23)$$

Now, we prove our assertion. The lower-tails in (21)(22) and the upper-tail in (23) implies

$$\begin{aligned} \sum_{i=1}^n \alpha_i^{2q+1} \langle \mathbf{S} \mathbf{u}_i, \mathbf{S} \mathbf{u}_1 \rangle^2 &\geq d \left(d \left(1 - \frac{\kappa}{4} \right) + \left(1 - \frac{\kappa}{2} \right) s_+ - \left(1 + \frac{\kappa}{2} \right) s_- \right) \\ &\stackrel{(a)}{=} d \left(\frac{d(4 - \kappa)}{4} + \frac{\kappa(s_+ + s_-)}{2} \right) \\ &\stackrel{(b)}{\geq} \frac{\kappa}{3} \left(d \left(d \left(1 + \frac{\kappa}{4} \right) + \left(1 + \frac{\kappa}{2} \right) (s_+ + s_-) \right) \right) \stackrel{(c)}{\geq} \frac{\kappa}{3} \sum_{i=1}^n |\alpha_i|^{2q+1} \langle \mathbf{S} \mathbf{u}_i, \mathbf{S} \mathbf{u}_1 \rangle^2, \end{aligned}$$

where (a) is due to $(1 - \kappa)s_+ = (1 + \kappa)s_-$ (rearranged from $\sum_{i=2}^n \alpha_i^{2q+1} = \kappa \sum_{i=2}^n |\alpha_i|^{2q+1}$), (b) is easily checked by comparing the coefficients, and (c) follows from the upper-tails in (21)(22)(23). Therefore, a union bound completes the proof with $c_\kappa = \frac{\kappa}{3}$ in this case.

(ii). $s_- = o(\sqrt{d})$. This is equivalent to say that there exists a constant $c > 0$ such that $s_- \leq c\sqrt{d}$. Notice that from $\sum_{i=1}^n \alpha_i^{2q+1} \langle \mathbf{S}^T \mathbf{u}_i, \mathbf{S}^T \mathbf{u}_1 \rangle^2$ to $\sum_{i=1}^n |\alpha_i|^{2q+1} \langle \mathbf{S}^T \mathbf{u}_i, \mathbf{S}^T \mathbf{u}_1 \rangle^2$, only the term with index $i \in \mathcal{I}_-$ changes its sign and we show our assertion in the sense that the terms with indices in \mathcal{I}_- do not affect too much with high probability.

Invoking Lemma 12 in Appendix C.3 with $\delta = \frac{\kappa\sqrt{d}}{8c}$, $\epsilon = \frac{\delta}{1+\delta}$, $\beta_i = |\alpha_i|$ for $i \in \mathcal{I}_-$, and $\beta_i = 0$ otherwise , we get

$$\mathbb{P} \left[\sum_{i \in \mathcal{I}_-} |\alpha_i|^{2q+1} \langle \mathbf{S}^T \mathbf{u}_i, \mathbf{S}^T \mathbf{u}_1 \rangle^2 \leq d \left(1 + \frac{\kappa\sqrt{d}}{4c} \right) s_- \right] \geq 1 - e^{-\Omega(\sqrt{d}\kappa^2)}. \quad (24)$$

On the one hand, the lower-tail in (21) and the upper-tail in (24) yield that with high probability

$$\sum_{i=1}^n \alpha_i^{2q+1} \langle \mathbf{S} \mathbf{u}_i, \mathbf{S} \mathbf{u}_1 \rangle^2 \geq d^2 \left(1 - \frac{\kappa}{4} \right) - d \left(1 + \frac{\kappa\sqrt{d}}{4c} \right) s_- + \sum_{i \in \mathcal{I}_+} \alpha_i^{2q+1} \langle \mathbf{S} \mathbf{u}_i, \mathbf{S} \mathbf{u}_1 \rangle^2. \quad (25)$$

On the other hand, the upper-tails in (21) and (24) imply that with high probability

$$\sum_{i=1}^n |\alpha_i|^{2q+1} \langle \mathbf{S} \mathbf{u}_i, \mathbf{S} \mathbf{u}_1 \rangle^2 \leq d^2 \left(1 + \frac{\kappa}{4} \right) + d \left(1 + \frac{\kappa\sqrt{d}}{4c} \right) s_- + \sum_{i \in \mathcal{I}_+} \alpha_i^{2q+1} \langle \mathbf{S} \mathbf{u}_i, \mathbf{S} \mathbf{u}_1 \rangle^2. \quad (26)$$

Finally with a union bound on (25) and (26) , we have probability at least $1 - e^{-\Omega(\sqrt{d}\kappa^2)}$ that for any $d \geq \left(\frac{14c}{10-7\kappa} \right)^2 = \Theta(1)$,

$$\begin{aligned} \sum_{i=1}^n \alpha_i^{2q+1} \langle \mathbf{S} \mathbf{u}_i, \mathbf{S} \mathbf{u}_1 \rangle^2 - \frac{\sum_{i=1}^n |\alpha_i|^{2q+1} \langle \mathbf{S} \mathbf{u}_i, \mathbf{S} \mathbf{u}_1 \rangle^2}{6} &\geq d^2 \left(1 - \frac{\kappa}{4} \right) - d \left(1 + \frac{\kappa\sqrt{d}}{4c} \right) s_- - \frac{d^2 \left(1 + \frac{\kappa}{4} \right) + d \left(1 + \frac{\kappa\sqrt{d}}{4c} \right) s_-}{6} \\ &\geq \frac{(10-7\kappa)d^2 - 14cd\sqrt{d}}{12} \geq 0, \end{aligned}$$

where the second inequality stems from $s_- \leq c\sqrt{d}$. Hence, the proof is completed with $c_\kappa = \frac{1}{6}$ in this case. \square

Theorem 4. Assume \mathbf{A} satisfies Assumption 1. Let $\hat{\mathbf{u}} = \text{RSVD}(\mathbf{A}, \mathcal{N}(0, 1)^{n \times d}, q, d)$ for any $q \in \mathbb{N}$. Then,

$$R(\hat{\mathbf{u}}) = \Omega \left(c_\kappa \left(\frac{d}{d+i_0} \right)^{\frac{1}{2q+1}} \right)$$

with probability at least $1 - e^{-\Omega(\sqrt{d}\kappa^2)}$.

Proof Evaluating $R_{\mathbf{a}}$ defined in (4) in Section 4.1 on $\mathbf{a} = \mathbf{S}^T \mathbf{u}_1 / \|\mathbf{S}^T \mathbf{u}_1\|_2$ and by Lemma 2 there exists a constant $c_\kappa \in (0, 1]$ such that

$$R_{\mathbf{a}} = \frac{\sum_{i=1}^n \alpha_i^{2q+1} \langle \mathbf{S}^T \mathbf{u}_i, \mathbf{S}^T \mathbf{u}_1 \rangle^2}{\sum_{i=1}^n \alpha_i^{2q} \langle \mathbf{S}^T \mathbf{u}_i, \mathbf{S}^T \mathbf{u}_1 \rangle^2} \geq c_\kappa \frac{\sum_{i=1}^n |\alpha_i|^{2q+1} \langle \mathbf{S}^T \mathbf{u}_i, \mathbf{S}^T \mathbf{u}_1 \rangle^2}{\sum_{i=1}^n \alpha_i^{2q} \langle \mathbf{S}^T \mathbf{u}_i, \mathbf{S}^T \mathbf{u}_1 \rangle^2} = c_\kappa \bar{R}_{\mathbf{a}},$$

where $\bar{R}_{\mathbf{a}}$ is introduced in (9) in Section 4.3, with probability at least $1 - e^{-\Omega(\sqrt{d}\kappa^2)}$. Repeating the arguments in the proof of Theorem 3 in Appendix A.2 with replacing $R(\hat{\mathbf{u}})$ by $\bar{R}_{\mathbf{a}}$ yields:

$$\bar{R}_{\mathbf{a}} = \Omega \left(c_\kappa \left(\frac{d}{d+i_0} \right)^{\frac{1}{2q+1}} \right)$$

with probability at least $1 - e^{-\Omega(\sqrt{d}\kappa^2)}$, and hence the desired result follows by the union bound. \square

B Proofs of RandSum

B.1 Large deviation of projection length for Bernoulli random matrix

This subsection is used to prove Lemma 3, which serves as an intermediate step for Theorem 5, restated below. The proof relies on a simple but powerful concept, ε -net. As its usefulness, the definition and related theorems can be found in literature of random matrix. Here we shortly define it and state its important property below Lemma 3. Interested reader are referred to the reference therein.

Definition 3 (ε -net, Definition 4.2.1 in (Vershynin, 2018)). *Let $(\mathbb{S}^{d-1}, \|\cdot\|_2)$ be a metric space and $\varepsilon > 0$. A subset $\mathcal{N}_\varepsilon \subseteq \mathbb{S}^{d-1}$ is called ε -net if*

$$\forall x, y \in \mathcal{N}_\varepsilon, \|x - y\|_2 \leq \varepsilon.$$

Lemma 8 (Corollary 4.2.13 in (Vershynin, 2018)). *For any $\varepsilon \in (0, 1)$, the size of \mathcal{N}_ε is bounded by*

$$|\mathcal{N}_\varepsilon| \leq 3^d \varepsilon^{-d}.$$

We are ready to prove Lemma 3 restated below.

Lemma 3. *Let $\mathbf{v} \in \mathbb{S}^{n-1}$, $d \leq n/3$, and $\mathbf{S} \sim \text{Bernoulli}(p)^{n \times d}$ for a constant $p \in (0, 1)$. Then,*

$$\cos^2 \theta(\mathbf{v}, \mathbf{S}) = \Omega\left(\frac{\max\{1, \langle \mathbf{v}, \mathbf{1}_n \rangle^2\}}{n}\right)$$

holds with probability at least $1 - e^{-\Omega(d)}$.

Proof As it is easy to see that \mathbf{S} is a nonzero matrix with probability $1 - e^{-nd}$, the following deduction will be made under $\|\mathbf{S}\|_2 > 0$.

By the second inequality in Corollary 2 in Appendix C.2 with $\mathbf{x} = \mathbf{v}$ and $\delta = 1/2$, we deduce that

$$\mathbb{P}\left[\|\mathbf{S}^T \mathbf{v}\|_2 \geq \sqrt{\frac{dp(1-p+p\langle \mathbf{v}, \mathbf{1}_n \rangle^2)}{2}}\right] \geq 1 - e^{-\Omega(d)}. \quad (27)$$

Recall that $\cos \theta(\mathbf{v}, \mathbf{S}) = \max_{\mathbf{a} \in \mathbb{S}^{d-1}} \frac{\langle \mathbf{v}, \mathbf{S}\mathbf{a} \rangle}{\|\mathbf{S}\mathbf{a}\|_2}$, (27) allows us to substitute $\mathbf{a} = \mathbf{S}^T \mathbf{v} / \|\mathbf{S}^T \mathbf{v}\|$ and have

$$\cos \theta(\mathbf{v}, \mathbf{S}) \geq \frac{\|\mathbf{S}^T \mathbf{v}\|_2^2}{\|\mathbf{S}\mathbf{S}^T \mathbf{v}\|_2} \geq \frac{\|\mathbf{S}^T \mathbf{v}\|_2}{\|\mathbf{S}\|_2} \geq \frac{\sqrt{dp(1-p+p\langle \mathbf{v}, \mathbf{1}_n \rangle^2)}}{\sqrt{2} \|\mathbf{S}\|_2},$$

where the second inequality is due to submultiplicativity of $\|\cdot\|_2$, namely $\|\mathbf{S}\mathbf{S}^T \mathbf{v}\|_2 \leq \|\mathbf{S}\|_2 \|\mathbf{S}^T \mathbf{v}\|_2$, and the last one is a consequence of (27). It remains to show that $\|\mathbf{S}\|_2 \leq \mathcal{O}(\sqrt{nd})$ w.h.p., then the proof is done. For this goal, we use the ε -net technique, introduced in the beginning of this subsection, and give a bound in two steps:

(i). Let \mathcal{N}_ε be an ε -net defined on $(\mathbb{S}^{d-1}, \|\cdot\|_2)$ for some $\varepsilon \in (0, 1)$ to be determined later. We claim that

$$\|\mathbf{S}\|_2 \leq \frac{1}{1 - \varepsilon} \sup_{\mathbf{x} \in \mathcal{N}_\varepsilon} \|\mathbf{S}\mathbf{x}\|_2. \quad (28)$$

Let $\mathbf{w}^* \in \operatorname{argmax}_{\mathbf{x} \in \mathbb{S}^{d-1}} \|\mathbf{S}\mathbf{x}\|_2$, and since there exists $\mathbf{x}^* \in \mathcal{N}_\varepsilon$ satisfying $\|\mathbf{w}^* - \mathbf{x}^*\|_2 \leq \varepsilon$, by submultiplicativity and triangle inequality, we get

$$\varepsilon \|\mathbf{S}\|_2 \geq \|\mathbf{S}(\mathbf{w}^* - \mathbf{x}^*)\|_2 \geq \|\mathbf{S}\|_2 - \|\mathbf{S}\mathbf{x}^*\|_2 \geq \|\mathbf{S}\|_2 - \sup_{\mathbf{x} \in \mathcal{N}_\varepsilon} \|\mathbf{S}\mathbf{x}\|_2,$$

and rearranging the terms yields (28).

(ii). Show that

$$\mathbb{P} \left[\sup_{\mathbf{x} \in \mathcal{N}_\varepsilon} \|\mathbf{S}\mathbf{x}\|_2 \leq \left(\frac{3}{2} np(1-p+pd) \right)^{\frac{1}{2}} \right] \geq 1 - 3^d \varepsilon^{-d} e^{-\Omega(n)} \geq 1 - e^{-\Omega(n+d \ln \frac{\varepsilon}{3})}. \quad (29)$$

For each $\mathbf{x} \in \mathcal{N}_\varepsilon$, the first inequality in Corollary 2 in Appendix C.2 with $\mathbf{x} = \mathbf{x}$ and $\delta = \frac{1}{2}$ (here n and d are reversed) implies that we have probability at least $1 - e^{-\Omega(n)}$

$$\|\mathbf{S}\mathbf{x}\|_2 \leq \left(\frac{3}{2} np(1-p+p\langle \mathbf{x}, \mathbf{1}_d \rangle^2) \right)^{\frac{1}{2}} \leq \left(\frac{3}{2} np(1-p+pd) \right)^{\frac{1}{2}},$$

where the last inequality is due to $\langle \mathbf{x}, \mathbf{1}_d \rangle^2 \leq d$. As the size of \mathcal{N}_ε is upper bounded by $3^d \varepsilon^{-d}$ (see Lemma 8 on the top of this subsection), the union bound over all $\mathbf{x} \in \mathcal{N}_\varepsilon$ yields (29).

Finally, setting $\varepsilon = 1/e$ in (28)-(29) and assumption $n - 2d > d$ lead to $\|\mathbf{S}\mathbf{x}\|_2 \leq \mathcal{O}(\sqrt{nd})$ holds with probability at least $1 - e^{-\Omega(n-2d)} > 1 - e^{-\Omega(d)}$. The union bound completes our proof as desired. \square

B.2 RandSum with positive semidefinite matrices

Theorem 5. Let \mathbf{A} be a positive semi-definite matrix with $\lambda_1 > 0$ and $\hat{\mathbf{u}} = \text{RandSum}(\mathbf{A}, q, d, p)$ for any constant $p \in (0, 1)$, any $q \in \mathbb{N}$, and $d \geq 2$. Then,

$$R(\hat{\mathbf{u}}) = \left(\Omega \left(\frac{\max\{d, \langle \mathbf{u}_1, \mathbf{1}_n \rangle^2\}}{n} \right) \right)^{\frac{1}{2q+1}}$$

with probability at least $1 - e^{-\Omega(d)}$.

Proof Define $\mathcal{A}_1 = \left\{ \begin{bmatrix} \mathbf{a}_1 \\ \mathbf{0}_{\lfloor \frac{d}{2} \rfloor} \end{bmatrix} : \mathbf{a}_1 \in \mathbb{S}^{\lceil \frac{d}{2} \rceil - 1} \right\}$ and $\mathcal{A}_2 = \left\{ \begin{bmatrix} \mathbf{0}_{\lceil \frac{d}{2} \rceil} \\ \mathbf{a}_2 \end{bmatrix} : \mathbf{a}_2 \in \mathbb{S}^{\lfloor \frac{d}{2} \rfloor - 1} \right\}$. Since $R(\hat{\mathbf{u}}) \geq \max_{\mathbf{a} \in \mathbb{S}^{d-1}} R_{\mathbf{a}}$, where $R_{\mathbf{a}}$ is introduce on Section 4.1 and has an expression (4), we can conclude that

$$R(\hat{\mathbf{u}}) \geq \max \left\{ \max_{\mathbf{a} \in \mathcal{A}_1} R_{\mathbf{a}}, \max_{\mathbf{a} \in \mathcal{A}_2} R_{\mathbf{a}} \right\} \geq \max \left\{ \cos^2 \theta(\mathbf{u}_1, \mathbf{S}_1), \cos^2 \theta(\mathbf{u}_1, \mathbf{S}_2) \right\}^{\frac{1}{2q+1}},$$

where the last inequality is an application of (7) in Section 4.1. The proof is completed by Lemma 1 and Lemma 3. \square

B.3 RandSum with indefinite matrices

Assumption 2 is restated here for convenience.

Assumption 2. Assume that (i) $\langle \mathbf{u}_1, \mathbf{1}_n \rangle^2 = \Omega(1)$ and (ii) there exists a constant $\kappa' \in (0, 1]$ such that

$$\sum_{i=2}^n \lambda_i^{2q+1} \xi_i \geq \kappa' \sum_{i=2}^n |\lambda_i|^{2q+1} \xi_i,$$

where $\xi_i = \mathbb{E} \left[\langle \mathbf{S}^T \mathbf{u}_i, \frac{\mathbf{1}_d}{\sqrt{d}} \rangle^2 \right] = p(1 - p + pd \langle \mathbf{u}_i, \mathbf{1}_n \rangle^2)$, $\forall i \in [n]$.

Lemma 4. Assume that \mathbf{A} satisfies Assumption 2. Let $\mathbf{S} \sim \text{Bernoulli}(p)^{n \times d}$ for a constant $p \in (0, 1)$. There exists a constant $c_{\kappa'} \in (0, 1]$ such that

$$\mathbb{P} \left[\sum_{i=1}^n \lambda_i^{2q+1} \langle \mathbf{S}^T \mathbf{u}_i, \mathbf{S}^T \mathbf{u}_1 \rangle^2 \geq c_{\kappa'} \sum_{i=1}^n |\lambda_i|^{2q+1} \langle \mathbf{S}^T \mathbf{u}_i, \mathbf{S}^T \mathbf{u}_1 \rangle^2 \right] \geq 1 - e^{-\Omega(\sqrt{d}\kappa'^2)}.$$

Proof Here we introduce

$$\mu_i = \mathbb{E} \left[\left\| \mathbf{S}_{:,1}^T \mathbf{u}_i \right\|_2^2 \right] = p(1 - p + p\langle \mathbf{u}_i, \mathbf{1}_n \rangle^2), \forall i \in [n].$$

Recall that $\alpha_i = \lambda_i / \lambda_1$ for all $i \in [n]$. From Assumption 2, we have:

- By (i), there exists a constant $\nu \in (0, 1]$ such that $\langle \mathbf{u}_1, \mathbf{1}_n \rangle^2 \geq \nu$. It follows that

$$\xi_1 \geq p^2 d \langle \mathbf{u}_1, \mathbf{1}_n \rangle^2 = pd \cdot p((1-p)\langle \mathbf{u}_1, \mathbf{1}_n \rangle^2 + p\langle \mathbf{u}_1, \mathbf{1}_n \rangle^2) \geq p\nu d \mu_1. \quad (30)$$

- By (ii), there exists $\kappa' \in (0, 1]$ such that $\sum_{i=2}^n \alpha_i^{2q+1} \xi_i = \kappa' \sum_{i=2}^n |\alpha_i|^{2q+1} \xi_i$.

We then partition $[n]$ into three subsets, $[n] = \{1\} \cup \mathcal{I}_+ \cup \mathcal{I}_-$, where $\mathcal{I}_+ = \{i \in [n] \setminus \{1\} : \alpha_i > 0\}$ and $\mathcal{I}_- = \{i \in [n] : \alpha_i < 0\}$. It is natural to assume $\mathcal{I}_- \neq \emptyset$, as otherwise this lemma trivially holds. As similar to what we proceed in the proof of Lemma 2 in Appendix A.3, two important quantities follows from this partition: $s_+ = \sum_{i \in \mathcal{I}_+} \alpha_i^{2q+1} \xi_i$ and $s_- = \sum_{i \in \mathcal{I}_-} |\alpha_i|^{2q+1} \xi_i$.

Firstly, for the term $\langle \mathbf{S}^T \mathbf{u}_1, \mathbf{S}^T \mathbf{u}_1 \rangle^2$, applying Corollary 2 in Appendix C.2 with $\mathbf{x} = \mathbf{u}_1$ and $\delta = 1 - \sqrt{1 - \kappa'/4}$ (resp. $\delta = \sqrt{1 + \kappa'/4} - 1$) for the lower-tail (resp. upper-tail) yields that

$$\mathbb{P} \left[d^2 \mu_1^2 \left(1 - \frac{\kappa'}{4} \right) \leq \langle \mathbf{S}^T \mathbf{u}_1, \mathbf{S}^T \mathbf{u}_1 \rangle^2 \leq d^2 \mu_1^2 \left(1 + \frac{\kappa'}{4} \right) \right] \geq 1 - e^{-\Omega(d\kappa'^2)}. \quad (31)$$

As for the remaining terms, we carefully apply concentration inequalities under two scenarios: either (i). $s_- = \Omega(\sqrt{d})$ or (ii). $s_- = o(\sqrt{d})$.

(i). $s_- = \Omega(\sqrt{d})$. Invoking the lower-tail (resp. upper-tail) of Lemma 13 in C.3 with $\delta = \epsilon = \sqrt{1 + \kappa'/2} - 1$ (resp. $\delta = \epsilon = \sqrt{1 - \kappa'/2} - 1$), $\beta_i = \alpha_i$ for $i \in \mathcal{I}_+$, and $\beta_i = 0$ otherwise, we get

$$\mathbb{P} \left[\xi_1 \left(1 - \frac{\kappa'}{2} \right) s_+ \leq \sum_{i \in \mathcal{I}_+} \alpha_i^{2q+1} \langle \mathbf{S}^T \mathbf{u}_i, \mathbf{S}^T \mathbf{u}_1 \rangle^2 \leq d \mu_1 \left(1 + \frac{\kappa'}{2} \right) s_+ \right] \geq 1 - e^{-\Omega(\sqrt{d}\kappa'^2)}. \quad (32)$$

Again, using Lemma 13 with $\delta = \epsilon = \sqrt{1 + \kappa'/2} - 1$, $\beta_i = |\alpha_i|$ for $i \in \mathcal{I}_-$, and $\beta_i = 0$ otherwise leads to

$$\mathbb{P} \left[\sum_{i \in \mathcal{I}_-} |\alpha_i|^{2q+1} \langle \mathbf{S}^T \mathbf{u}_i, \mathbf{S}^T \mathbf{u}_1 \rangle^2 \leq d \mu_1 \left(1 + \frac{\kappa'}{2} \right) s_- \right] \geq 1 - e^{-\Omega(\sqrt{d}\kappa'^2)}. \quad (33)$$

Now, we prove our assertion. The lower-tails in (31)(32) and upper-tail in (33) imply that

$$\begin{aligned} \sum_{i=1}^n \alpha_i^{2q+1} \langle \mathbf{S} \mathbf{u}_i, \mathbf{S} \mathbf{u}_1 \rangle^2 &\geq d \mu_1 \left(d \mu_1 \left(1 - \frac{\kappa'}{4} \right) + \frac{\xi_1}{d \mu_1} \left(1 - \frac{\kappa'}{2} \right) s_+ - \left(1 + \frac{\kappa'}{2} \right) s_- \right) \\ &\stackrel{(a)}{\geq} p \nu \cdot d \mu_1 \left(d \mu_1 \left(1 - \frac{\kappa'}{4} \right) + \left(1 - \frac{\kappa'}{2} \right) s_+ - \left(1 + \frac{\kappa'}{2} \right) s_- \right) \\ &\stackrel{(b)}{=} p \nu \cdot d \mu_1 \left(\frac{d(4 - \kappa') \mu_1}{4} + \frac{\kappa'(s_+ + s_-)}{2} \right) \\ &\stackrel{(c)}{\geq} \frac{p \nu \kappa'}{3} \left(d \mu_1 \left(\frac{d(4 + \kappa') \mu_1}{4} + \left(1 + \frac{\kappa'}{2} \right) (s_+ + s_-) \right) \right) \stackrel{(d)}{\geq} \frac{p \nu \kappa'}{3} \sum_{i=1}^n |\alpha_i|^{2q+1} \langle \mathbf{S} \mathbf{u}_i, \mathbf{S} \mathbf{u}_1 \rangle^2, \end{aligned}$$

where (a) uses (30), (b) is due to $(1 - \kappa')s_+ = (1 + \kappa')s_-$ (rearranged from $\sum_{i \neq 1} \alpha_i^{2q+1} \xi_i = \kappa' \sum_{i \neq 1} |\alpha_i|^{2q+1} \xi_i$), (c) is easily checked by comparing the coefficients, and (d) follows from the upper-tails in (31)(32)(33). Therefore, a union bound completes the proof with $c_{\kappa'} = \frac{p \nu \kappa'}{3}$ in this case.

(ii). $s_- = o(\sqrt{d})$. There exists a constant $c > 0$ such that $s_- \leq c\sqrt{d}$. Observe that for two summations of interest, $\sum_{i=1}^n \alpha_i^{2q+1} \langle \mathbf{S}^T \mathbf{u}_i, \mathbf{S}^T \mathbf{u}_1 \rangle^2$ and $\sum_{i=1}^n |\alpha_i|^{2q+1} \langle \mathbf{S}^T \mathbf{u}_i, \mathbf{S}^T \mathbf{u}_1 \rangle^2$, only the terms in \mathcal{I}_- change their signs. Our assertion follows in the sense that the terms with indices in \mathcal{I}_- do not affect too much with high probability.

Invoking Lemma 13 in C.3 with $\delta = \frac{\kappa'(\sqrt{d}-1)\mu_1}{4c}$, $\epsilon = \frac{\kappa'\mu_1}{4c(1+\delta)}$, $\beta_i = |\alpha_i|$ for $i \in \mathcal{I}_-$, and $\beta_i = 0$ otherwise, we get

$$\mathbb{P} \left[\sum_{i \in \mathcal{I}_-} |\alpha_i|^{2q+1} \langle \mathbf{S}^T \mathbf{u}_i, \mathbf{S}^T \mathbf{u}_1 \rangle^2 \leq d\mu_1 \left(1 + \frac{\kappa'\sqrt{d}\mu_1}{4c} \right) s_- \right] \geq 1 - e^{-\Omega(\sqrt{d}\kappa'^2)}. \quad (34)$$

On the one hand, the lower-tail in (31) and the upper-tail in (34) yield that

$$\sum_{i=1}^n \alpha_i^{2q+1} \langle \mathbf{S} \mathbf{u}_i, \mathbf{S} \mathbf{u}_1 \rangle^2 \geq d^2 \mu_1^2 \left(1 - \frac{\kappa'}{4} \right) - d\mu_1 \left(1 + \frac{\kappa'\sqrt{d}\mu_1}{4c} \right) s_- + \sum_{i \in \mathcal{I}_+} \alpha_i^{2q+1} \langle \mathbf{S} \mathbf{u}_i, \mathbf{S} \mathbf{u}_1 \rangle^2 \quad (35)$$

On the other hand, the upper-tails in (31)(34) imply that

$$\sum_{i=1}^n |\alpha_i|^{2q+1} \langle \mathbf{S} \mathbf{u}_i, \mathbf{S} \mathbf{u}_1 \rangle^2 \leq d^2 \mu_1^2 \left(1 + \frac{\kappa'}{4} \right) + d\mu_1 \left(1 + \frac{\kappa'\sqrt{d}\mu_1}{4c} \right) s_- + \sum_{i \in \mathcal{I}_+} \alpha_i^{2q+1} \langle \mathbf{S} \mathbf{u}_i, \mathbf{S} \mathbf{u}_1 \rangle^2 \quad (36)$$

As a consequence of a union bound on (35)(36), we have with probability at least $1 - e^{-\Omega(\sqrt{d}\kappa'^2)}$,

$$\begin{aligned} & \sum_{i=1}^n \alpha_i^{2q+1} \langle \mathbf{S} \mathbf{u}_i, \mathbf{S} \mathbf{u}_1 \rangle^2 - \frac{1}{6} \cdot \sum_{i=1}^n |\alpha_i|^{2q+1} \langle \mathbf{S} \mathbf{u}_i, \mathbf{S} \mathbf{u}_1 \rangle^2 \\ & \geq d^2 \mu_1^2 \left(1 - \frac{\kappa'}{4} \right) - d\mu_1 \left(1 + \frac{\kappa'\sqrt{d}\mu_1}{4c} \right) s_- - \frac{d^2 \mu_1^2 \left(1 + \frac{\kappa'}{4} \right) + d\mu_1 \left(1 + \frac{\kappa'\sqrt{d}\mu_1}{4c} \right) s_-}{6} \\ & \geq \frac{1}{12} \left((10 - 7\kappa')d^2 \mu_1^2 - 14cd\sqrt{d}\mu_1 \right) \geq 0, \end{aligned}$$

where the second inequality is due to $s_- \leq c\sqrt{d}$, for any $d \geq \left(\frac{14c}{(10-7\kappa')\mu_1} \right)^2 = \Theta(1)$. Hence, the proof is completed with $c_{\kappa'} = \frac{1}{6}$ in this case. \square

Theorem 6. Assume that \mathbf{A} satisfies Assumptions 1 and 2. Let $\hat{\mathbf{u}} = \text{RandSum}(\mathbf{A}, q, d, p)$ for any constant $p \in (0, 1)$ and any $q \in \mathbb{N}$, and i_0 be defined as in Definition 2 in Section 4.2. Then,

$$R(\hat{\mathbf{u}}) = \Omega \left(\max \left\{ c_\kappa \left(\frac{d}{d+i_0} \right)^{\frac{1}{2q+1}}, c_{\kappa'} \left(\frac{\max\{d, \langle \mathbf{u}_1, \mathbf{1}_n \rangle^2\}}{n} \right)^{\frac{1}{2q+1}} \right\} \right)$$

with probability at least $1 - e^{-\Omega(\sqrt{d} \min(\kappa, \kappa')^2)}$.

Proof Let

$$\mathbf{a}_1 = \begin{bmatrix} \frac{\mathbf{S}_1^T \mathbf{u}_1}{\|\mathbf{S}_1^T \mathbf{u}_1\|_2} \\ \mathbf{0}_{\lfloor \frac{d}{2} \rfloor} \end{bmatrix} \quad \text{and} \quad \mathbf{a}_2 = \begin{bmatrix} \mathbf{0}_{\lceil \frac{d}{2} \rceil} \\ \frac{\mathbf{S}_2^T \mathbf{u}_1}{\|\mathbf{S}_2^T \mathbf{u}_1\|_2} \end{bmatrix}.$$

A union bound of Lemma 2 and Lemma 4 implies that there exist constants c_κ and $c_{\kappa'}$ such that

$$R_{\mathbf{a}_1} \geq c_\kappa \bar{R}_{\mathbf{a}_1} \quad \text{and} \quad R_{\mathbf{a}_2} \geq c_{\kappa'} \bar{R}_{\mathbf{a}_2},$$

with probability at least $1 - e^{-\Omega(\sqrt{d}\kappa^2)} - e^{-\Omega(\sqrt{d}\kappa'^2)}$, where $R_{\mathbf{a}}$ and $\bar{R}_{\mathbf{a}}$ are defined in (4) in Section 4.1 and (9) in Section 4.3, respectively. Hence,

$$\mathbb{P} [R(\hat{\mathbf{u}}) \geq \max \{c_\kappa \bar{R}_{\mathbf{a}_1}, c_{\kappa'} \bar{R}_{\mathbf{a}_2}\}] \geq 1 - e^{-\Omega(\sqrt{d} \min(\kappa, \kappa')^2)}.$$

Finally, applying similar argument in the proof of Theorem 3 (see Appendix A.2) to lower bound $\bar{R}_{\mathbf{a}_1}$ and Theorem 5 to lower bound $\bar{R}_{\mathbf{a}_2}$ completes the proof. \square

C Concentration inequalities

Before showing our lemmas on both Gaussian and Bernoulli random variables, there are some necessary definition and standard concentration inequalities to be introduced. For the random variables considered in this work, sub-gaussian and sub-exponential norms are useful to quantify the probabilities of rare events. In C.1, we introduce them for completeness and list the concentration inequalities (Hoeffding, Bernstein, and Hanson-Wright inequalities) used in the following proofs. In C.2, we provide two corollaries yielded by Bernstein inequality for Gaussian and Bernoulli distributions respectively. Finally, our technical lemmas for these two random variables will be shown in C.3.

C.1 Sub-gaussian norm and sub-exponential norm

Definition 4 (Definition 2.5.6 (Vershynin, 2018)). *The sub-gaussian norm $\|\cdot\|_{\psi_2}$ is a norm on the space of sub-gaussian random variables. For any sub-gaussian random variable X ,*

$$\|X\|_{\psi_2} = \inf\{t > 0 : \mathbb{E}[\exp(X^2/t^2)] \leq 2\}.$$

The sum of sub-gaussian random variables is still a sub-gaussian random variable, and its norm can be characterized by the following Proposition.

Proposition 1 (Proposition 2.6.1 (Vershynin, 2018)). *Let X_1, \dots, X_m be a zero-mean sub-gaussian random variables. Then,*

$$\left\| \sum_{i \in [m]} X_i \right\|_{\psi_2}^2 = \mathcal{O} \left(\sum_{i \in [m]} \|X_i\|_{\psi_2}^2 \right),$$

where \mathcal{O} hides an absolute constant.

Definition 5 (Definition 2.7.5 (Vershynin, 2018)). *The sub-exponential norm $\|\cdot\|_{\psi_1}$ is a norm on the space of sub-exponential random variables. For any sub-exponential random variable X ,*

$$\|X\|_{\psi_1} = \inf\{t > 0 : \mathbb{E}[\exp(|X|/t)] \leq 2\}.$$

If X is sub-gaussian random variable, then X is also a sub-exponential random variable. Besides, there is one well-known property for these two norms.

Proposition 2 (Lemma 2.7.6 (Vershynin, 2018)). *Let X be a zero-mean sub-gaussian random variable. Then,*

$$\|X\|_{\psi_2}^2 = \|X^2\|_{\psi_1}.$$

For concreteness, we compute sub-gaussian norms for two basic variables.

Example 1. Here we evaluate the values of $\|\cdot\|_{\psi_2}$ and $\|\cdot\|_{\psi_1}$ for the sub-gaussian random variables which will be used later

- If $X \sim \mathcal{N}(0, \sigma^2)$, for some $\sigma \in \mathbb{R}_+$, then $\|X\|_{\psi_2} = 2\sigma$.
- If $Y \sim \text{Bernoulli}(p)$, for some $p \in (0, 1)$, then $\|Y\|_{\psi_2} = \frac{1}{\sqrt{\ln(1+p^{-1})}}$ and $\|Y\|_{\psi_1} = \frac{1}{\ln(1+p^{-1})}$.

Proof For any $t > \sqrt{2}\sigma$, we observe that

$$\mathbb{E} [\exp(X^2/t^2)] = \frac{1}{\sigma\sqrt{2\pi}} \int_{x \in \mathbb{R}} \exp\left(-\frac{x^2}{2\sigma^2} + \frac{x^2}{t^2}\right) dx = \frac{1}{\sigma\sqrt{\frac{1}{2\sigma^2} - \frac{1}{t^2}}},$$

which is 2 when $t = 2\sigma$, hence $\|X\|_{\psi_2} = 2\sigma$. As for Y , elementary calculus shows that

$$\|Y\|_{\psi_2} = \inf \{t > 0 : p \exp(t^{-2}) + (1-p) \leq 2\} = \inf \left\{ t > 0 : \exp(t^{-2}) \leq \frac{1+p}{p} \right\} = \frac{1}{\sqrt{\ln(1+p^{-1})}},$$

and that

$$\|Y\|_{\psi_1} = \inf \{t > 0 : p \exp(t^{-1}) + (1-p) \leq 2\} = \inf \left\{ t > 0 : \exp(t^{-1}) \leq \frac{1+p}{p} \right\} = \frac{1}{\ln(1+p^{-1})}.$$

□

Here is the list of concentration inequalities we will use later. The first proposition is an immediate result from Definition 4 and 5, the others are standard concentration inequalities characterized by these two norms.

Proposition 3 (Proposition 2.5.2 and Proposition 2.7.1 in (Vershynin, 2018)). *Let X and Y be a sub-gaussian and a sub-exponential random variables, respectively. Then for any $t \geq 0$, we have*

$$\mathbb{P}[|X - \mathbb{E}[X]| \geq t] \leq \exp\left(-\Omega\left(\frac{t^2}{\|X\|_{\psi_2}^2}\right)\right) \quad \text{and} \quad \mathbb{P}[|Y - \mathbb{E}[Y]| \geq t] \leq \exp\left(-\Omega\left(\frac{t}{\|Y\|_{\psi_1}}\right)\right).$$

Lemma 9 (Hoeffding's inequality (Theorem 2.6.3 in (Vershynin, 2018))). *Let $m \in \mathbb{N}$, X_1, \dots, X_m be i.i.d. zero-mean sub-gaussian random variables, and $\mathbf{a} \in \mathbb{R}^m$ be a nonzero vector. Then,*

$$\forall t \geq 0, \quad \mathbb{P}\left[\left|\sum_{i=1}^m \mathbf{a}_i X_i\right| > t\right] \leq \exp\left(-\Omega\left(\frac{t^2}{K \|\mathbf{a}\|_2^2}\right)\right),$$

where $K = \|\mathbf{X}_1\|_{\psi_2}^2$.

Lemma 10 (Bernstein's inequality (Theorem 2.8.2 in (Vershynin, 2018))). *Let $m \in \mathbb{N}$ and $\mathbf{a} = (\mathbf{a}_1, \dots, \mathbf{a}_m) \in \mathbb{R}^m \setminus \{\mathbf{0}_n\}$. Let X_1, \dots, X_m be independent sub-gaussian r.v.'s. Then there exists a universal constant $c > 0$ such that for any $t > 0$,*

$$\mathbb{P}\left[\left|\sum_{i=1}^m \mathbf{a}_i (X_i^2 - \mathbb{E}[X_i^2])\right| \geq t\right] \leq 2 \exp\left(-c \min\left\{\frac{t^2}{K^2 \|\mathbf{a}\|_2^2}, \frac{t}{K \|\mathbf{a}\|_\infty}\right\}\right),$$

where $K = \max_{i \in [m]} \|X_i^2 - \mathbb{E}[X_i^2]\|_{\psi_1}$.

Lemma 11 (Hanson-Wright inequality (Theorem 6.2.1 in (Vershynin, 2018))). *Let $m \in \mathbb{N}$ and $\mathbf{X} = (\mathbf{X}_1, \dots, \mathbf{X}_m)$ be a random vector with i.i.d zero-mean sub-gaussian entries and $\mathbf{M} \in \mathbb{R}^{m \times m} \setminus \{\mathbf{0}_{m \times m}\}$. Then,*

$$\forall t > 0, \quad \mathbb{P}\left[\left|\sum_{i,j \in [m]} \mathbf{M}_{i,j} \mathbf{X}_i \mathbf{X}_j - \mathbb{E}\left[\sum_{i,j \in [m]} \mathbf{M}_{i,j} \mathbf{X}_i \mathbf{X}_j\right]\right| > t\right] \leq \exp\left(-\Omega\left(\min\left\{\frac{t^2}{K^2 \|\mathbf{M}\|_F^2}, \frac{t}{K \|\mathbf{M}\|_2}\right\}\right)\right),$$

where $K = \|\mathbf{X}_1\|_{\psi_2}^2$.

C.2 Useful lemmas derived from Bernstein's inequality

In this subsection, we will use Lemma 10 in C.1 to derive two Bernstein-type concentration inequalities. Corollary 1 (resp. Corollary 2) provides tail bounds on the length $\|\mathbf{S}\mathbf{x}\|_2$ of Gaussian (resp. Bernoulli) random matrix S with linear combination weights \mathbf{x} of its columns.

Corollary 1. Let $\mathbf{x} \in \mathbb{R}^n \setminus \{\mathbf{0}_n\}$ and $\mathbf{S} \sim \mathcal{N}(0, 1)^{n \times d}$. Then, $\forall \delta > 0$,

$$\mathbb{P} \left[\|\mathbf{S}^T \mathbf{x}\|_2^2 \geq d(1 + \delta) \|\mathbf{x}\|_2^2 \right] \leq e^{-\Omega(d \min\{\delta, \delta^2\})}, \quad \text{and} \quad \mathbb{P} \left[\|\mathbf{S}^T \mathbf{x}\|_2^2 \leq d(1 - \delta) \|\mathbf{x}\|_2^2 \right] \leq e^{-\Omega(d \min\{\delta, \delta^2\})}.$$

Proof For each $i = 1, \dots, d$, the i -th column of \mathbf{S} is denoted as $\mathbf{S}_{:,i}$. Because $\langle \mathbf{S}_{:,1}, \frac{\mathbf{x}}{\|\mathbf{x}\|_2} \rangle, \dots, \langle \mathbf{S}_{:,d}, \frac{\mathbf{x}}{\|\mathbf{x}\|_2} \rangle$ are i.i.d. random variable drawn from $\mathcal{N}(0, 1)$, the application of Lemma 10 with $m = d$, $\mathbf{a} = \mathbf{1}_d$, $t = \delta d$, and $X_i = \langle \mathbf{S}_{:,i}, \mathbf{x}/\|\mathbf{x}\|_2 \rangle$ for $i = 1, \dots, d$, implies that there is a universal constant $c > 0$ such that

$$\mathbb{P} \left[\left| \sum_{i=1}^d \langle \mathbf{S}_{:,i}, \frac{\mathbf{x}}{\|\mathbf{x}\|_2} \rangle^2 - d \right| \geq \delta \cdot d \right] \leq 2 \exp \left(-c \min \left\{ \frac{\delta^2 d}{K^2}, \frac{\delta d}{K} \right\} \right) = \exp(-\Omega(d \min\{\delta, \delta^2\})) ,$$

where $K = \|\mathbf{X}_1^2 - \mathbb{E}[\mathbf{X}_1^2]\|_{\psi_1}$. A triangle inequality on ψ_1 norm gives the of K as:

$$K \leq \|\mathbf{X}_1^2\|_{\psi_1} + \|\mathbb{E}[\mathbf{X}_1^2]\|_{\psi_1} \leq \|\mathbf{X}_1\|_{\psi_2}^2 + \frac{1}{\ln 2} \leq 2 + \frac{1}{\ln 2},$$

where the second inequality is a consequence of Proposition 2 and the last one is shown in Example 1. As $\|\mathbf{S}^T \mathbf{x}\|_2^2 = \sum_{i=1}^d \langle \mathbf{S}_{:,i}, \mathbf{x} \rangle^2$, the two claimed inequalities hold by rearranging the above inequality. \square

Corollary 2. Let $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{S} \sim \text{Bernoulli}(p)^{n \times d}$ for a constant $p \in (0, 1)$. Then, $\forall \delta > 0$,

$$\mathbb{P} \left[\|\mathbf{S}^T \mathbf{x}\|_2^2 \geq d(1 + \delta)\mu \right] \leq e^{-\Omega(d \min\{\delta, \delta^2\})} \quad \text{and} \quad \mathbb{P} \left[\|\mathbf{S}^T \mathbf{x}\|_2^2 \leq d(1 - \delta)\mu \right] \leq e^{-\Omega(d \min\{\delta, \delta^2\})},$$

where $\mu = p(1 - p) \|\mathbf{x}\|_2^2 + p^2 \langle \mathbf{x}, \mathbf{1}_n \rangle^2$.

Proof For each $i = 1, \dots, d$, we denote the i -th column of \mathbf{S} as $\mathbf{S}_{:,i}$. Since $\langle \mathbf{S}_{:,1}, \mathbf{x} \rangle, \dots, \langle \mathbf{S}_{:,d}, \mathbf{x} \rangle$ are i.i.d., Lemma 10 with $m = d$, $\mathbf{a} = \mathbf{1}_d$, $t = \delta d \mu$, and $X_i = \langle \mathbf{S}_{:,i}, \mathbf{x} \rangle$ for $i = 1, \dots, d$, implies that there exists a universal constant $c > 0$ such that

$$\mathbb{P} \left[\left| \sum_{i=1}^d \langle \mathbf{S}_{:,i}, \mathbf{x} \rangle^2 - d \mathbb{E}[\langle \mathbf{S}_{:,1}, \mathbf{x} \rangle^2] \right| \geq \delta \cdot d \mu \right] \leq 2 \exp \left(-c \min \left\{ \frac{d \mu^2 \delta^2}{K^2}, \frac{d \mu \delta}{K} \right\} \right),$$

where $K = \|\langle \mathbf{S}_{:,1}, \mathbf{x} \rangle^2 - \mathbb{E}[\langle \mathbf{S}_{:,1}, \mathbf{x} \rangle^2]\|_{\psi_1}$. The proof is done by showing (i). $\mathbb{E}[\langle \mathbf{S}_{:,1}, \mathbf{x} \rangle^2] = \mu$, and (ii). $K = \Theta(\mu)$.

(i). Show $\mathbb{E}[\langle \mathbf{S}_{:,1}, \mathbf{x} \rangle^2] = \mu$: By using linearity of expectation repeatedly, we obtain that

$$\begin{aligned} \mathbb{E}[\langle \mathbf{S}_{:,1}, \mathbf{x} \rangle^2] &= \mathbb{E} \left[\left(\sum_{i=1}^n \mathbf{S}_{i,1} \mathbf{x}_i \right)^2 \right] = \sum_{i=1}^n \mathbb{E}[(\mathbf{S}_{i,1} \mathbf{x}_i)^2] + \sum_{i \neq j} \mathbb{E}[(\mathbf{S}_{i,1} \mathbf{x}_i)(\mathbf{S}_{j,1} \mathbf{x}_j)] \\ &= p \|\mathbf{x}\|_2^2 + p^2 (\langle \mathbf{x}, \mathbf{1}_n \rangle^2 - \|\mathbf{x}\|_2^2) = p(1 - p) \|\mathbf{x}\|_2^2 + p^2 \langle \mathbf{x}, \mathbf{1}_n \rangle^2 = \mu. \end{aligned}$$

(ii). Show $K = \Theta(\mu)$: Let $\mathbf{Z} = \mathbf{S}_{:,1} - p\mathbf{1}_n$. As verified in (i), $\mathbb{E}[\langle \mathbf{S}_{:,1}, \mathbf{x} \rangle^2] = \mu = p(1 - p) \|\mathbf{x}\|_2^2 + p^2 \langle \mathbf{x}, \mathbf{1}_n \rangle^2$, we get

$$\begin{aligned} K &= \|\langle \mathbf{S}_{:,1}, \mathbf{x} \rangle^2 - \mathbb{E}[\langle \mathbf{S}_{:,1}, \mathbf{x} \rangle^2]\|_{\psi_1} = \left\| \langle \mathbf{Z}, \mathbf{x} \rangle^2 + 2p \langle \mathbf{Z}, \mathbf{x} \rangle \langle \mathbf{x}, \mathbf{1}_n \rangle - p(1 - p) \|\mathbf{x}\|_2^2 \right\|_{\psi_1} \\ &\leq \|\langle \mathbf{Z}, \mathbf{x} \rangle^2\|_{\psi_1} + 2p |\langle \mathbf{x}, \mathbf{1}_n \rangle| \|\langle \mathbf{Z}, \mathbf{x} \rangle\|_{\psi_1} + \frac{p(1 - p) \|\mathbf{x}\|_2^2}{\ln 2}. \end{aligned}$$

Since \mathbf{Z} has i.i.d. entries and $p = \Theta(1)$, we evaluate

$$\begin{aligned} \|\langle \mathbf{Z}, \mathbf{x} \rangle^2\|_{\psi_1} &= \|\langle \mathbf{Z}, \mathbf{x} \rangle\|_{\psi_2}^2 = \sum_{i=1}^n \mathbf{x}_i^2 \|\mathbf{Z}_i\|_{\psi_2}^2 = \|x\|_2^2 \|\mathbf{Z}_1\|_{\psi_2}^2 \leq \|x\|_2^2 \left(\|\mathbf{S}_{1,1}\|_{\psi_2} + \|p\|_{\psi_2} \right)^2, \\ \text{and } \|\langle \mathbf{Z}, \mathbf{x} \rangle\|_{\psi_1} &= |\langle \mathbf{x}, \mathbf{1}_n \rangle| \|\mathbf{Z}_1\|_{\psi_1} \leq |\langle \mathbf{x}, \mathbf{1}_n \rangle| \left(\|\mathbf{S}_{1,1}\|_{\psi_1} + \|p\|_{\psi_1} \right). \end{aligned}$$

Because $\|\mathbf{S}_{1,1}\|_{\psi_2} = \frac{1}{\sqrt{\ln(1+p^{-1})}} = \Theta(1)$, $\|\mathbf{S}_{1,1}\|_{\psi_1} = \frac{1}{\ln(1+p^{-1})} = \Theta(1)$ (see Example 1 for ψ_1 and ψ_2 norm), $\|p\|_{\psi_2} = \frac{p}{\sqrt{\ln 2}} = \Theta(1)$, and $\|p\|_{\psi_1} = \frac{p}{\ln 2} = \Theta(1)$, combining all yields $K = \Theta(\mu)$. \square

C.3 Techinical Lemmas

Lemma 12. Let $\beta = (\beta_1, \dots, \beta_n) \in [0, 1]^n$ s.t $(\beta_2, \dots, \beta_n) \neq \mathbf{0}_{n-1}$, $\mathbf{U} = [\mathbf{u}_1, \dots, \mathbf{u}_n] \in \mathbb{R}^{n \times n}$ be an orthonormal matrix, and $\mathbf{S} \sim \mathcal{N}(0, 1)^{n \times d}$. Then, for any $\delta > 0$ and $\epsilon \in (0, 1)$,

$$\begin{aligned} \mathbb{P} \left[\sum_{i=2}^n \beta_i \langle \mathbf{S}^T \mathbf{u}_i, \mathbf{S}^T \mathbf{u}_1 \rangle^2 \geq d(1+\epsilon)(1+\delta) \sum_{i=2}^n \beta_i \right] &\leq \exp \left(-\Omega \left(\max \left\{ 1, \sum_{i=2}^n \beta_i \right\} \min \left\{ \delta, \delta^2 \right\} \right) \right) + e^{-\Omega(d\epsilon^2)}, \\ \text{and } \mathbb{P} \left[\sum_{i=2}^n \beta_i \langle \mathbf{S}^T \mathbf{u}_i, \mathbf{S}^T \mathbf{u}_1 \rangle^2 \leq d(1-\epsilon)(1-\delta) \sum_{i=2}^n \beta_i \right] &\leq \exp \left(-\Omega \left(\max \left\{ 1, \sum_{i=2}^n \beta_i \right\} \min \left\{ \delta, \delta^2 \right\} \right) \right) + e^{-\Omega(d\epsilon^2)}. \end{aligned}$$

Proof In the following, we only focus on the upper-tail bound as the others will hold by symmetry.

For the simplicity of presentation, we introduce a set $\mathcal{V}_\epsilon = \{\mathbf{v} \in \mathbb{R}^d : 0 < \|\mathbf{v}\|_2^2 \leq d(1+\epsilon)\}$ and the events

$$E = \mathbb{I} \left\{ \sum_{i=2}^n \beta_i \langle \mathbf{S}^T \mathbf{u}_i, \mathbf{S}^T \mathbf{u}_1 \rangle^2 \geq d(1+\epsilon)(1+\delta) \sum_{i=2}^n \beta_i \right\} \text{ and } G(\mathbf{v}) = \mathbb{I} \{ \mathbf{S}^T \mathbf{u}_1 = \mathbf{v} \}, \forall \mathbf{v} \in \mathcal{V}_\epsilon.$$

Using Corollary 1 in C.2 with $\mathbf{x} = \mathbf{u}_1$, $\delta = \epsilon < 1$ and the fact that $\|\mathbf{S}^T \mathbf{u}_1\|_2^2 > 0$ a.e. yield that $\mathbb{P}[\neg(\cup_{\mathbf{v} \in \mathcal{V}_\epsilon} G(\mathbf{v}))] = \mathbb{P}[\|\mathbf{S}^T \mathbf{u}_1\|_2^2 > d(1+\epsilon)] \leq e^{-\Omega(d\epsilon^2)}$, which explicitly says that $\cup_{\mathbf{v} \in \mathcal{V}_\epsilon} G(\mathbf{v})$ happens with high probability. As a consequence, we have

$$\mathbb{P}[E] \leq \mathbb{P}[E \cap \neg(\cup_{\mathbf{v} \in \mathcal{V}_\epsilon} G(\mathbf{v}))] + \int_{\mathbf{v} \in \mathcal{V}_\epsilon} \mathbb{P}[E \cap G(\mathbf{v})] d\mathbb{P}[G(\mathbf{v})] \leq e^{-\Omega(d\epsilon^2)} + \sup_{\mathbf{v} \in \mathcal{V}_\epsilon} \mathbb{P}[E \cap G(\mathbf{v})], \quad (37)$$

where the last inequality follows from $\mathbb{P}[\cup_{\mathbf{v} \in \mathcal{V}_\epsilon} G(\mathbf{v})] \leq 1$ and the upper bound of $\mathbb{P}[\neg(\cup_{\mathbf{v} \in \mathcal{V}_\epsilon} G(\mathbf{v}))]$ proved above. By (37), it is sufficient to show that for any $\mathbf{v} \in \mathcal{V}_\epsilon$,

$$\mathbb{P}[E \cap G(\mathbf{v})] \leq \exp \left(-\Omega \left(\max \left\{ 1, \sum_{i=2}^n \beta_i \right\} \min \left\{ \delta, \delta^2 \right\} \right) \right). \quad (38)$$

Show (38) for any $\mathbf{v} \in \mathcal{V}_\epsilon$

Since $\|\mathbf{v}\|_2^2 \leq d(1+\epsilon)$ for each $\mathbf{v} \in \mathcal{V}_\epsilon$,

$$\begin{aligned} \mathbb{P}[E \cap G(\mathbf{v})] &= \mathbb{P} \left[\sum_{i=2}^n \beta_i \langle \mathbf{S}^T \mathbf{u}_i, \mathbf{v} \rangle^2 \geq \left(\sum_{i=2}^n \beta_i \right) d(1+\delta)(1+\epsilon) \right] \\ &\leq \mathbb{P} \left[\sum_{i=2}^n \beta_i \langle \mathbf{S}^T \mathbf{u}_i, \mathbf{v} \rangle^2 \geq \left(\sum_{i=2}^n \beta_i \right) (1+\delta) \|\mathbf{v}\|_2^2 \right]. \end{aligned} \quad (39)$$

Because for each $i = 2, \dots, n$, $\langle \mathbf{S}^T \mathbf{u}_i, \mathbf{v} \rangle = \sum_{r=1}^n \sum_{s=1}^d \mathbf{S}_{r,s}(\mathbf{u}_i)_r \mathbf{v}_s$ is a random variable from $\mathcal{N}(0, \|\mathbf{v}\|_2^2)$ (a linear combination of normal distributions is a normal distribution again), Example 1 in C.1 shows that

$\|\langle \mathbf{S}^T \mathbf{u}_i, \mathbf{v} \rangle\|_{\psi_2} = 2 \|\mathbf{v}\|_2$. Moreover, the assumption \mathbf{U} is an orthonormal matrix implies that $\{\langle \mathbf{S}^T \mathbf{u}_i, \mathbf{v} \rangle\}_{i=2}^n$ are independent (see Theorem 8.1, Chap 5(Gut, 2009)). By applying Lemma 10 in C.1 with $m = n - 1$, $X_i = \langle \mathbf{S}^T \mathbf{u}_{i+1}, \mathbf{v} \rangle$, $\forall i = 1, \dots, n - 1$, $\mathbf{a} = (\beta_2, \dots, \beta_n)$, and $t = \delta \cdot (\sum_{i=2}^n \beta_i) \|\mathbf{v}\|_2^2$, we give an upper bound of right-hand side of (39) as below (it is already shown that $\mathbb{E}[\langle \mathbf{S}^T \mathbf{u}_i, \mathbf{v} \rangle^2] = \|\mathbf{v}\|_2^2$ before):

$$\begin{aligned} \mathbb{P} \left[\sum_{i=2}^n \beta_i \left(\langle \mathbf{S}^T \mathbf{u}_i, \mathbf{v} \rangle^2 - \|\mathbf{v}\|_2^2 \right) \geq \delta \left(\sum_{i=2}^n \beta_i \right) \|\mathbf{v}\|_2^2 \right] &\leq \mathbb{P} \left[\left| \sum_{i=2}^n \beta_i \left(\langle \mathbf{S}^T \mathbf{u}_i, \mathbf{v} \rangle^2 - \|\mathbf{v}\|_2^2 \right) \right| \geq \delta \left(\sum_{i=2}^n \beta_i \right) \|\mathbf{v}\|_2^2 \right] \\ &\leq 2 \exp \left(-c \min \left\{ \frac{\delta^2 (\sum_{i=2}^n \beta_i)^2 \|\mathbf{v}\|_2^4}{\|\mathbf{v}\|_2^4 \sum_{i=2}^n \beta_i^2}, \frac{\delta \cdot (\sum_{i=2}^n \beta_i) \|\mathbf{v}\|_2^2}{\|\mathbf{v}\|_2^2 \max_{i \neq 1} \beta_i} \right\} \right) \\ &= 2 \exp \left(-c \min \left\{ \frac{(\sum_{i=2}^n \beta_i)^2 \delta^2}{\sum_{i=2}^n \beta_i^2}, \frac{\sum_{i=2}^n \beta_i \delta}{\max_{i \neq 1} \beta_i} \right\} \right). \end{aligned} \quad (40)$$

Combining (39)(40), it remains to show that

$$(i) \frac{(\sum_{i=2}^n \beta_i)^2}{\sum_{i=2}^n \beta_i^2} \geq \max \left\{ 1, \sum_{i=2}^n \beta_i \right\}, \text{ and } (ii) \frac{\sum_{i=2}^n \beta_i}{\max_{i \neq 1} \beta_i} \geq \max \left\{ 1, \sum_{i=2}^n \beta_i \right\}.$$

For (i). As $(\sum_{i=2}^n \beta_i)^2 = \sum_{i=2}^n \beta_i^2 + \sum_{i \neq j} \beta_i \beta_j \geq \sum_{i=2}^n \beta_i^2$, and $\sum_{i=2}^n \beta_i^2 \leq \sum_{i=2}^n \beta_i$, (i) holds by using these two inequalities in numerator and denominator respectively.

For (ii). As $\sum_{i=2}^n \beta_i \geq \max_{i \neq 1} \beta_i$, and $\max_{i \neq 1} \beta_i \leq 1$, (ii) follows by using these two inequalities in numerator and denominator respectively. \square

Lemma 13. Let $\beta = (\beta_1, \dots, \beta_n) \in [0, 1]^n$ s.t. $(\beta_2, \dots, \beta_n) \neq \mathbf{0}_{n-1}$, $\mathbf{U} = [\mathbf{u}_1, \dots, \mathbf{u}_n] \in \mathbb{R}^{n \times n}$ be an orthonormal matrix with $\langle \mathbf{u}_1, \mathbf{1}_n \rangle^2 = \Omega(1)$, and $\mathbf{S} \sim \text{Bernoulli}(p)^{n \times d}$ with some constant $p \in (0, 1)$. Then, for any $\delta > 0$ and $\epsilon \in (0, 1)$, we have probability at least $1 - e^{-\Omega(\max\{1, \sum_{i=2}^n \beta_i \xi_i\} \frac{(1-\epsilon)^2}{(1+\epsilon)^2} \min\{\delta, \delta^2\})} - e^{-\Omega(\min(d, \xi_1) \epsilon^2)}$ that

$$(1 - \delta)(1 - \epsilon) \sum_{i=2}^n \beta_i \xi_i \xi_1 \leq \sum_{i=2}^n \beta_i \langle \mathbf{S}^T \mathbf{u}_i, \mathbf{S}^T \mathbf{u}_1 \rangle^2 \leq d(1 + \delta)(1 + \epsilon) \sum_{i=2}^n \beta_i \xi_i \mu_1,$$

where

$$\mu_i = p(1 - p + p \langle \mathbf{u}_i, \mathbf{1}_n \rangle^2), \quad \text{and} \quad \xi_i = p(1 - p + pd \langle \mathbf{u}_i, \mathbf{1}_n \rangle^2), \quad \forall i \in [n].$$

Proof Similar to the proof C.3 of Lemma 12, we introduce the set

$$\mathcal{V}_\epsilon = \{\mathbf{v} \in \mathbb{R}^d : d\mu_1(1 - \epsilon) \leq \|\mathbf{v}\|_2^2 \leq d\mu_1(1 + \epsilon) \text{ and } d\xi_1(1 - \epsilon) \leq \langle \mathbf{v}, \mathbf{1}_d \rangle_2^2 \leq d\xi_1(1 + \epsilon)\}$$

and the events

$$E = \mathbb{I} \left(\left\{ \sum_{i=2}^n \beta_i \langle \mathbf{S}^T \mathbf{u}_i, \mathbf{S}^T \mathbf{u}_1 \rangle^2 \leq (1 - \delta)(1 - \epsilon) \sum_{i=2}^n \beta_i \xi_i \xi_1 \right\} \cup \left\{ \sum_{i=2}^n \beta_i \langle \mathbf{S}^T \mathbf{u}_i, \mathbf{S}^T \mathbf{u}_1 \rangle^2 \geq d(1 + \delta)(1 + \epsilon) \sum_{i=2}^n \beta_i \xi_i \mu_1 \right\} \right)$$

and $G(\mathbf{v}) = \mathbb{I}\{\mathbf{S}^T \mathbf{u}_1 = \mathbf{v}\}$, $\forall \mathbf{v} \in \mathcal{V}_\epsilon$. The sum rule of probability implies that

$$\mathbb{P}[E] = \mathbb{P}[E \cap \neg(\cup_{\mathbf{v} \in \mathcal{V}_\epsilon} G(\mathbf{v}))] + \int_{\mathbf{v} \in \mathcal{V}_\epsilon} \mathbb{P}[E \cap G(\mathbf{v})] d\mathbb{P}[G(\mathbf{v})] \leq \mathbb{P}[\neg(\cup_{\mathbf{v} \in \mathcal{V}_\epsilon} G(\mathbf{v}))] + \sup_{\mathbf{v} \in \mathcal{V}_\epsilon} \mathbb{P}[E \cap G(\mathbf{v})]. \quad (41)$$

To bound the first term in (41), we claim that

$$\begin{aligned} (i) \quad &\mathbb{P} \left[\left| \|\mathbf{S}^T \mathbf{u}_1\|_2^2 - d\mu_1 \right| \geq \epsilon \cdot d\mu_1 \right] \leq e^{-\Omega(d\epsilon^2)}, \\ (ii) \quad &\mathbb{P} \left[\left| \langle \mathbf{S}^T \mathbf{u}_1, \mathbf{1}_d \rangle^2 - d\xi_1 \right| \geq \epsilon \cdot d\xi_1 \right] \leq e^{-\Omega(\xi_1 \epsilon^2)}, \end{aligned}$$

and then an application of a union bound of (i)(ii) yields $\mathbb{P}[\neg(\cup_{\mathbf{v} \in \mathcal{V}_\epsilon} G(\mathbf{v}))] \leq e^{-\Omega(d\epsilon^2)} + e^{-\Omega(\xi_1\epsilon^2)}$. As for the second term in (41), one consequence of Lemma 14 at the end of this section is that

$$\sup_{\mathbf{v} \in \mathcal{V}_\epsilon} \mathbb{P}[E \cap G(\mathbf{v})] \leq e^{-\Omega\left(\max\{1, \sum_{i=2}^n \beta_i \xi_i\} \min\left\{1, \frac{\langle \mathbf{v}, \mathbf{1}_d \rangle^4}{d^2 \|\mathbf{v}\|_2^4}, \frac{\langle \mathbf{v}, \mathbf{1}_d \rangle^2}{d \|\mathbf{v}\|_2^2}\right\} \min\{\delta, \delta^2\}\right)}$$

in which $\min\left\{1, \frac{\langle \mathbf{v}, \mathbf{1}_d \rangle^4}{d^2 \|\mathbf{v}\|_2^4}, \frac{\langle \mathbf{v}, \mathbf{1}_d \rangle^2}{d \|\mathbf{v}\|_2^2}\right\} = \min\left\{1, \frac{(1-\epsilon)^2}{(1+\epsilon)^2} \frac{\xi_1^2}{d^2 \mu_1^2}\right\} = \Omega\left(\frac{(1-\epsilon)^2}{(1+\epsilon)^2}\right)$. Hence, combining all by union bound gives the desired.

It remains to show (i) and (ii). For convenience, let $\mathbf{Z} = \mathbf{S} - p\mathbf{1}_n\mathbf{1}_d^T$, a zero-mean matrix.

(i). This is a direct result of the first inequality in Corollary 2 in C.2 with $\mathbf{x} = \mathbf{u}_1$ and $\delta = \epsilon$.

(ii). To show $\mathbb{P}[\lvert \langle \mathbf{S}^T \mathbf{u}_1, \mathbf{1}_d \rangle^2 - d\xi_1 \rvert \geq \epsilon \cdot d\xi_1] \leq e^{-\Omega(\xi_1\epsilon^2)}$, where $\xi = p(1 - p + pd\langle \mathbf{u}_1, \mathbf{1}_n \rangle^2)$. As the lower tail is proved in a similar to the upper tail, in what follows, we will pay attention on the upper tail only. Firstly, it is easy to verified that

$$\langle \mathbf{S}^T \mathbf{u}_1, \mathbf{1}_d \rangle^2 = \langle \mathbf{Z}^T \mathbf{u}_1, \mathbf{1}_d \rangle^2 + 2pd\langle \mathbf{Z}^T \mathbf{u}_1, \mathbf{1}_d \rangle \langle \mathbf{u}_1, \mathbf{1}_n \rangle + p^2 d^2 \langle \mathbf{u}_1, \mathbf{1}_n \rangle^2. \quad (42)$$

To bound the first (resp. the second) term in (42), we will use Proposition 3 in C.1 for sub-exponential (resp. sub-gaussian) r.v., which is quantified the sub-gaussian norm, denoted by $K = \|\langle \mathbf{Z}^T \mathbf{u}_1, \mathbf{1}_d \rangle\|_{\psi_2}$ (recall that the sub-exponential norm can be obtained by sub-gaussian norm, and vice versa, see Propisition 2 in C.1). By Proposition 1 and Example 1 in C.1, we have

$$K^2 = \mathcal{O}\left(d \sum_{i \in [n]} (\mathbf{u}_1)_i^2 \|\mathbf{Z}_{1,1}\|_{\psi_2}^2\right) = \mathcal{O}\left(d \|\mathbf{Z}_{1,1}\|_{\psi_2}^2\right) = \mathcal{O}\left(d \left(\|\mathbf{S}_{i,j}\|_{\psi_2} + \|p\|_{\psi_2}\right)^2\right) = \mathcal{O}(d).$$

Additionally, one can evaluate $\langle \mathbf{Z}^T \mathbf{u}_1, \mathbf{1}_d \rangle^2 = dp(1 - p)$ by repeatedly use the linearity of expectation and the fact that the entries of \mathbf{S} are i.i.d. drawn from Bernoulli(p). Hence, invoking the concentration inequality for sub-exponential (resp. sub-gaussian) in Proposition 3 in C.1 with $t = \frac{d\xi_1\epsilon}{3}$ (resp. $t = \frac{\epsilon\sqrt{d\xi_1}}{3}$) on $\langle \mathbf{Z}^T \mathbf{u}_1, \mathbf{1}_d \rangle^2$ (resp. $\langle \mathbf{Z}^T \mathbf{u}_1, \mathbf{1}_d \rangle$) yields that

$$\mathbb{P}\left[\lvert \langle \mathbf{Z}^T \mathbf{u}_1, \mathbf{1}_d \rangle^2 - dp(1 - p) \rvert \geq \frac{d\xi_1\epsilon}{3}\right] \leq e^{-\Omega(\frac{d\xi_1\epsilon}{K^2})} \leq e^{-\Omega(\xi_1\epsilon^2)}, \quad (43)$$

$$\mathbb{P}\left[\lvert \langle \mathbf{Z}^T \mathbf{u}_1, \mathbf{1}_d \rangle \rvert \geq \frac{\epsilon\sqrt{d\xi_1}}{3}\right] \leq e^{-\Omega(\frac{\epsilon\sqrt{d\xi_1}}{K^2})} = e^{-\Omega(\xi_1\epsilon^2)}. \quad (44)$$

Plugging these (43) and (44) into (42), a union bound gives us that

$$\begin{aligned} \langle \mathbf{Z}^T \mathbf{u}_1, \mathbf{1}_d \rangle^2 + 2pd\langle \mathbf{Z}^T \mathbf{u}_1, \mathbf{1}_d \rangle \langle \mathbf{u}_1, \mathbf{1}_n \rangle + p^2 d^2 \langle \mathbf{u}_1, \mathbf{1}_n \rangle^2 &\leq dp(1 - p) + p^2 d^2 \langle \mathbf{u}_1, \mathbf{1}_n \rangle^2 + \frac{d\xi_1\epsilon}{3} + 2pd\langle \mathbf{u}_1, \mathbf{1}_n \rangle \frac{\epsilon\sqrt{d\xi_1}}{3} \\ &\leq d\xi_1 + d\xi_1\epsilon, \end{aligned}$$

where the second inequality is yielded by $\xi_1 = p(1 - p + pd\langle \mathbf{u}_1, \mathbf{1}_n \rangle^2)$ and $pd\langle \mathbf{u}_1, \mathbf{1}_n \rangle \leq \sqrt{d\xi_1}$. Then we conclude this lemma with (ii) as desired. \square

Lemma 14. Let $\mathbf{v} \in \mathbb{R}^d \setminus \{\mathbf{0}_d\}$, $(\beta_2, \dots, \beta_n) \in [0, 1]^{n-1} \setminus \{\mathbf{0}_{n-1}\}$, $[\mathbf{u}_1, \dots, \mathbf{u}_n] \in \mathbb{R}^{n \times n}$ be an orthonormal matrix, $\mathbf{S} \sim \text{Bernoulli}(p)^{n \times d}$ with some constant $p \in (0, 1)$, and $\xi_i = p(1 - p + pd\langle \mathbf{u}_i, \mathbf{1}_n \rangle^2), \forall i \in [n]$. Then,

$$\mathbb{P}\left[(1 - \delta)\eta_1 \leq \sum_{i=2}^n \beta_i \langle \mathbf{S}^T \mathbf{u}_i, \mathbf{v} \rangle^2 \leq (1 + \delta)\eta_2\right] \geq 1 - e^{-\Omega\left(\max\{1, \sum_{i=2}^n \beta_i \xi_i\} \min\left\{1, \frac{\langle \mathbf{v}, \mathbf{1}_n \rangle^4}{d^2 \|\mathbf{v}\|_2^4}, \frac{\langle \mathbf{v}, \mathbf{1}_n \rangle^2}{d \|\mathbf{v}\|_2^2}\right\} \min\{\delta, \delta^2\}\right)},$$

where $\eta_1 = \sum_{i=2}^n \beta_i \xi_i \frac{\langle \mathbf{v}, \mathbf{1}_d \rangle^2}{d}$ and $\eta_2 = \sum_{i=2}^n \beta_i \xi_i \|\mathbf{v}\|_2^2$.

Proof An elementary calculation of evaluating the expectation of $\sum_{i=2}^n \beta_i \langle \mathbf{S}^T \mathbf{u}_i, \mathbf{v} \rangle^2$ leads to

$$\mathbb{E} \left[\sum_{i=2}^n \beta_i \langle \mathbf{S}^T \mathbf{u}_i, \mathbf{v} \rangle^2 \right] = p(1-p) \sum_{i=2}^n \beta_i \|\mathbf{v}\|_2^2 + p^2 \sum_{i=2}^n \beta_i \langle \mathbf{u}_i, \mathbf{1}_n \rangle^2 \langle \mathbf{v}, \mathbf{1}_d \rangle^2.$$

After applying Cauchy inequality, $\langle \mathbf{v}, \mathbf{1}_d \rangle^2 \leq d \|\mathbf{v}\|_2^2$, twice, we get that $\eta_1 \leq \mathbb{E} [\sum_{i=2}^n \beta_i \langle \mathbf{S}^T \mathbf{u}_i, \mathbf{v} \rangle^2] \leq \eta_2$. This observation inspires us to give high probability lower bound in term of η_1 and upper bound in term of η_2 respectively.

Define $\mathbf{Z} = \mathbf{S} - p\mathbf{1}_n\mathbf{1}_d^T$, $\mathbf{M} = \sum_{i=2}^n \beta_i \mathbf{u}_i \mathbf{u}_i^T$ and $\mathbf{B} = \mathbf{M} \otimes \mathbf{v} \mathbf{v}^T$ where \otimes is the Kronecker product. With these definition, we can express the weighted sum as:

$$\begin{aligned} \sum_{i=2}^n \beta_i \langle \mathbf{S}^T \mathbf{u}_i, \mathbf{v} \rangle^2 &= \sum_{(i_1, j_1), (i_2, j_2) \in [n] \times [d]} \mathbf{B}_{(i_1, j_1), (i_2, j_2)} \mathbf{S}_{i_1, j_1} \mathbf{S}_{i_2, j_2} = (\text{I}) + (\text{II}), \\ \text{where } (\text{I}) &= \sum_{(i_1, j_1), (i_2, j_2) \in [n] \times [d]} \mathbf{B}_{(i_1, j_1), (i_2, j_2)} \mathbf{Z}_{i_1, j_1} \mathbf{Z}_{i_2, j_2} \\ \text{and } (\text{II}) &= \sum_{(i_1, j_1), (i_2, j_2) \in [n] \times [d]} \mathbf{B}_{(i_1, j_1), (i_2, j_2)} (\mathbf{Z}_{i_1, j_1} + \mathbf{Z}_{i_2, j_2} + p)p. \end{aligned} \quad (45)$$

Such decomposition allows us to bound (I) by Lemma 11 and bound (II) by Lemma 9 in C.1, which require us to evaluate the necessary quantities.

- $\|\mathbf{B}\|_F^2 = \sum_{i_1, i_2 \in [n], j_1, j_2 \in [d]} (\mathbf{M}_{i_1, i_2} \mathbf{v}_{j_1} \mathbf{v}_{j_2})^2 = \|\mathbf{v}\|_2^4 \|\mathbf{M}\|_F = \|\mathbf{v}\|_2^4 \sum_{i=2}^n \beta_i^2$,
where the last equation is due to $\mathbf{M} = \sum_{i=2}^n \beta_i \mathbf{u}_i \mathbf{u}_i^T$ is an eigenvalue decomposition of \mathbf{M} .
- $\|\mathbf{B}\|_2 = \|\mathbf{M}\|_2 \|\mathbf{v} \mathbf{v}^T\|_2 = \max_{i \neq 1} \beta_i \|\mathbf{v}\|_2^2$,
where the first equation is a property of Kronecker product (see e.g. Theorem 4.2.15 in (Horn et al., 1994)).

To bound (I), invoking Lemma 11 with $m = nd$, $\mathbf{M} = \mathbf{B}$, $\mathbf{X}_{(i-1)d+j} = \mathbf{Z}_{i,j}$, $\forall i \in [n], j \in [d]$ and $t = \delta\eta_1/2$ (resp. $t = \delta\eta_2/2$) for the lower- (resp. upper-) tail bounds yields that

$$\mathbb{P} \left[\neg \left\{ -\frac{\delta\eta_1}{2} < (\text{I}) - \mathbb{E}[(\text{II})] < \frac{\delta\eta_2}{2} \right\} \right] \leq \exp \left(-\Omega \left(\min \left\{ \frac{\eta_2^2 \delta^2}{\|\mathbf{B}\|_F^2}, \frac{\eta_2 \delta}{\|\mathbf{B}\|_2} \right\} \right) \right) + \exp \left(-\Omega \left(\min \left\{ \frac{\eta_1^2 \delta^2}{\|\mathbf{B}\|_F^2}, \frac{\eta_1 \delta}{\|\mathbf{B}\|_2} \right\} \right) \right). \quad (46)$$

To bound (II), applying Lemma 9 with $t = \delta\eta_1/4$ (resp. $t = \delta\eta_2/4$) for the lower- (resp. upper-) tail bounds yields

$$\mathbb{P} \left[\neg \left\{ -\frac{\delta\eta_1}{2} < (\text{II}) - \mathbb{E}[(\text{II})] < \frac{\delta\eta_2}{2} \right\} \right] \leq \exp \left(-\Omega \left(\frac{\eta_2^2 \delta^2}{\|\mathbf{B}\|_F^2} \right) \right) + \exp \left(-\Omega \left(\frac{\eta_1^2 \delta^2}{\|\mathbf{B}\|_F^2} \right) \right). \quad (47)$$

In what follows, we will show

- $\min \left\{ \frac{\eta_2^2}{\|\mathbf{B}\|_F^2}, \frac{\eta_2}{\|\mathbf{B}\|_2} \right\} = \Omega(\max \{1, \sum_{i=2}^n \beta_i \xi_i\})$, and
- $\min \left\{ \frac{\eta_1^2}{\|\mathbf{B}\|_F^2}, \frac{\eta_1}{\|\mathbf{B}\|_2} \right\} = \Omega \left(\max \{1, \sum_{i=2}^n \beta_i \xi_i\} \min \left\{ \frac{\langle \mathbf{v}, \mathbf{1}_d \rangle^4}{d^2 \|\mathbf{v}\|_2^4}, \frac{\langle \mathbf{v}, \mathbf{1}_d \rangle^2}{d \|\mathbf{v}\|_2^2} \right\} \right)$.

Then this proof is done by using a union bound of (46) and (47) into (45).

(i). From the definition of η_2 and our above computations, we get $\frac{\eta_2^2}{\|\mathbf{B}\|_F^2} = \Omega \left(\frac{(\sum_{i=2}^n \beta_i \xi_i)^2}{\sum_{i=2}^n \beta_i^2} \right)$ and $\frac{\eta_2}{\|\mathbf{B}\|_2} = \Omega \left(\frac{\sum_{i=2}^n \beta_i \xi_i}{\max_{i \neq 1} \beta_i} \right)$. It is done by the following claims:

$$(a). \frac{(\sum_{i=2}^n \beta_i \xi_i)^2}{\sum_{i=2}^n \beta_i^2} = \Omega \left(\max \left\{ 1, \sum_{i=2}^n \beta_i \xi_i \right\} \right), \text{ and (b). } \frac{\sum_{i=2}^n \beta_i \xi_i}{\max_{i \neq 1} \beta_i} = \Omega \left(\max \left\{ 1, \sum_{i=2}^n \beta_i \xi_i \right\} \right).$$

- stems from $(\sum_{i=2}^n \beta_i \xi_i)^2 \geq p^2(1-p)^2 \sum_{i=2}^n \beta_i^2$, and $p(1-p) \sum_{i=2}^n \beta_i^2 \leq p(1-p) \sum_{i=2}^n \beta_i \leq \sum_{i=2}^n \beta_i \xi_i$.
- holds since $\sum_{i=2}^n \beta_i \xi_i \geq p(1-p) \max_{i \neq 1} \beta_i$, and $\max_{i \neq 1} \beta_i \leq 1$.

(ii). From the definition of η_2 and our above computations, we get

$$\min \left\{ \frac{\eta_1^2}{\|\mathbf{B}\|_F^2}, \frac{\eta_1}{\|\mathbf{B}\|_2} \right\} = \Omega \left(\min \left\{ \frac{(\sum_{i=2}^n \beta_i \xi_i)^2}{\sum_{i=2}^n \beta_i^2}, \frac{\sum_{i=2}^n \beta_i \xi_i}{\max_{i \neq 1} \beta_i} \right\} \min \left\{ \frac{\langle \mathbf{v}, \mathbf{1}_d \rangle^4}{d^2 \|\mathbf{v}\|_2^4}, \frac{\langle \mathbf{v}, \mathbf{1}_d \rangle^2}{d \|\mathbf{v}\|_2^2} \right\} \right).$$

We then deduce (ii). by (a). and (b). and conclude this proof. \square

D Conflicting group detection: approximation ratio

Algorithm 3: RandomEigenSign (\mathbf{v}) by Bonchi et al. (2019)

```

for  $i = 1 \rightarrow n$  do
|  $\mathbf{r}_i = \text{sign}(\mathbf{v}_i) \cdot \text{Bernoulli}(|\mathbf{v}_i|)$ ;
end
return  $\mathbf{r}$ ;

```

Theorem 7. For any $\hat{\mathbf{u}} \in \mathbb{S}^{n-1}$, RandomEigenSign($\hat{\mathbf{u}}$) is an $\mathcal{O}(n^{1/2}/R(\hat{\mathbf{u}}))$ -approx algorithm to 2-conflicting group detection.

Proof The proof strategy is similar to the analysis in (Bonchi et al., 2019).

Let $\mathbf{r} = \text{RandomEigenSign}(\hat{\mathbf{u}})$ and $\mathbf{s} = \text{sign}(\hat{\mathbf{u}})$ where $s_i = 1$ if $\hat{\mathbf{u}}_i > 0$ otherwise $s_i = 0$, $\forall i \in [n]$. We have

$$\begin{aligned} \mathbb{E} \left[\frac{\mathbf{r}^T \mathbf{A} \mathbf{r}}{\mathbf{r}^T \mathbf{r}} \right] &= \sum_k \mathbb{E} \left[\frac{\mathbf{r}^T \mathbf{A} \mathbf{r}}{\mathbf{r}^T \mathbf{r}} \middle| \mathbf{r}^T \mathbf{r} = k \right] \mathbb{P} [\mathbf{r}^T \mathbf{r} = k] = \sum_k \frac{1}{k} \sum_{i,j \in [n]} \mathbf{A}_{i,j} \mathbf{s}_i \mathbf{s}_j \mathbb{P} [\mathbf{r}_i \mathbf{r}_j = \mathbf{s}_i \mathbf{s}_j \mid \mathbf{r}^T \mathbf{r} = k] \mathbb{P} [\mathbf{r}^T \mathbf{r} = k] \\ &\stackrel{(a)}{=} \sum_k \frac{1}{k} \sum_{i,j \in [n]} \mathbf{A}_{i,j} \mathbf{s}_i \mathbf{s}_j \mathbb{P} [\mathbf{r}^T \mathbf{r} = k \mid \mathbf{r}_i \mathbf{r}_j = \mathbf{s}_i \mathbf{s}_j] \mathbb{P} [\mathbf{r}_i \mathbf{r}_j = \mathbf{s}_i \mathbf{s}_j] = \sum_{i,j \in [n]} \mathbf{A}_{i,j} \hat{\mathbf{u}}_i \hat{\mathbf{u}}_j \mathbb{E} \left[\frac{1}{r^T r} \mid \mathbf{r}_i \mathbf{r}_j = \mathbf{s}_i \mathbf{s}_j \right] \\ &\stackrel{(b)}{\geq} \sum_{i,j \in [n]} \mathbf{A}_{i,j} \hat{\mathbf{u}}_i \hat{\mathbf{u}}_j \frac{1}{\mathbb{E} [r^T r \mid \mathbf{r}_i \mathbf{r}_j = \mathbf{s}_i \mathbf{s}_j]}, \end{aligned}$$

where (a) results from applying Bayes' rule, and (b) uses conditional Jensen's inequality. By

$$\mathbb{E} [\mathbf{r}^T \mathbf{r} \mid \mathbf{r}_i \mathbf{r}_j = \mathbf{s}_i \mathbf{s}_j] = 2 + \sum_{\ell \in [n] \setminus \{i,j\}} \mathbb{P} [\mathbf{r}_\ell = \mathbf{s}_\ell] \leq 2 + \sqrt{n-2},$$

(b) and $\hat{\mathbf{u}}^T \mathbf{A} \hat{\mathbf{u}} = R(\hat{\mathbf{u}})$, we get that $\mathbb{E} \left[\frac{\mathbf{r}^T \mathbf{A} \mathbf{r}}{\mathbf{r}^T \mathbf{r}} \right] \geq \frac{\hat{\mathbf{u}}^T \mathbf{A} \hat{\mathbf{u}}}{2 + \sqrt{n-2}} = \frac{R(\hat{\mathbf{u}}) \lambda_1}{2 + \sqrt{n-2}}$. \square