Classical spectral analysis: ℓ_2 perturbation theory

In these notes, I'll outline some classical results from matrix perturbation theory which offer a basic toolkit for characterizing the performance of spectral methods. Much of this content follows Chen et al. (2021). We'll consider the signal-plus-noise model

$$\mathbf{A} = \mathbf{A}^\star + \mathbf{E} \tag{1}$$

$\mathbf{A}^\star \in \mathbb{R}^{n\times n}$ is a deterministic, symmetric matrix and $\mathbf{E} = \mathbf{A} - \mathbf{A}^\star$ is a symmetric, mean-zero noise matrix. In Section 1, we'll aim to answer the following questions:

1. How does the *eigenspace* of $\mathbf{A}^\star$ change when perturbed by a symmetric matrix $\mathbf{E}$?

2. How do the *eigenvalues* of $\mathbf{A}^\star$ change when perturbed by a symmetric matrix $\mathbf{E}$?

We begin by describing some notions of distance between two subspaces, and how these notions are related. We will then present arguably the two most important theorems in matrix perturbation theory, the Davis-Kahan $\sin\boldsymbol{\Theta}$ theorem and Weyl's inequality, for controlling the eigenspaces and eigenvalues of a matrix under a random perturbation. We then move on to matrix concentration inequalities for quantifying the size of a perturbation matrix. We present two theorems: the classical matrix Bernstein and a more recent sharp inequality which improves on it by a logarithmic factor. Finally, we consider some applications in graph inference: graph embedding, link prediction and anomaly detection, and clustering under the stochastic block model.

Notation. In what follows, we consider the signal-plus-noise model (1) and let $\lambda_1 \geq \cdots \geq \lambda_n$ denote the eigenvalues of $\mathbf{A}$ and let $u_1, \ldots, u_n$ denote the corresponding set of orthonormal eigenvectors. For a given $1 \leq s \leq r \leq n$, we let $\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_s, \ldots, \lambda_r)$ and let $\mathbf{U} = (u_s, \ldots, u_r)$ denote the principal eigenspace. Let $d = r - s + 1$ and by convention set $\lambda_0 := \infty$ and $\lambda_{n+1} := -\infty$. We define the population quantities $\lambda_1^\star, \ldots, \lambda_n^\star, u_1^\star, \ldots, u_n^\star, \boldsymbol{\Lambda}^\star, \mathbf{U}^\star$ for $\mathbf{A}^\star$ similarly. We suppress the dependence of these matrices on s and r, but make clear when we consider a specific case.

1 Distances and angles between subspaces.

We'll begin by describing some useful metrics for quantifying the proximity of two eigenspaces.

We will represent an d-dimensional subspace $\mathcal{U}$ in $\mathbb{R}^n$ by a matrix $\mathbf{U} \in \mathbb{R}^{n\times d}$ whose columns form an orthonormal basis of $\mathcal{U}$. This representation $\mathbf{U}$ is not unique: for any rotation matrix $\mathbf{O} \in \mathcal{O}^{d\times d}$, the columns of $\mathbf{UO}$ also form an orthonormal basis for $\mathcal{U}$, and so any notion of distance between subspaces must take this rotational ambiguity into account. Despite this, we will use the notation $\mathcal{U}$ and $\mathbf{U}$ interchangeably. In the following, we will use $|||\cdot|||$ to denote a norm of choice, typically the spectral norm or the Fronbenius norm.

Distance metrics between subspaces. The following are some widely used metrics:

1. *Distance with optimal rotation.* A natural way to account for rotational ambiguity is to optimally align the basis vectors before computing the distance, motivating the distance

$$\mathsf{dist}_{|||\cdot|||}(\mathbf{U}, \mathbf{U}^\star) := \min_{\mathbf{O}\in\mathcal{O}^{d\times d}} |||\mathbf{UO} - \mathbf{U}^\star|||.$$

2. *Distance between projection matrices.* The projection matrix onto the subspace $\mathcal{U}$ is given by $\mathbf{U}\mathbf{U}^\top$ and is uniquely defined, since $\mathbf{U}\mathbf{U}^\top = \mathbf{U}\mathbf{O}\mathbf{O}^\top\mathbf{U}^\top$ for any $\mathbf{O} \in \mathcal{O}^{d\times d}$. This motivates the distance
$$\mathsf{dist}_{\mathsf{p},\|\cdot\|}(\mathbf{U}, \mathbf{U}^\star) := \left\|\left\|\mathbf{U}\mathbf{U}^\top - \mathbf{U}^\star\mathbf{U}^{\star\top}\right\|\right\|$$
where again $|||\cdot|||$ is a norm of choice.

3. *Principal angles.* Let $\sigma_1, \ldots, \sigma_d \in [0, 1]$ be the singular values of $\mathbf{U}^\top\mathbf{U}^\star$, arranged in descending order, and define the principal angles between two subspaces as
$$\theta_i := \arccos(\sigma_i), \qquad \text{for all } 1 \le i \le d.$$
When $d = 1$, this coincides with the conventional notion of angle between two unit vectors. One might then measure the distance between two subspaces with
$$\mathsf{dist}_{\mathsf{sin},|||\cdot|||}(\mathbf{U}, \mathbf{U}^\star) := |||\sin\mathbf{\Theta}|||$$
where $\sin\mathbf{\Theta} := \mathrm{diag}(\sin\theta_1, \ldots, \sin\theta_d)$.

Connections between the distance metrics. It turns out that all three notions of distance are intimately related. We begin with the relation between $\mathsf{dist}_{\mathsf{p},|||\cdot|||}$ and $\mathsf{dist}_{\mathsf{sin},|||\cdot|||}$ which is elucidated by the following lemma.

Lemma 1 (Stewart and Sun (1990))**.** *If $2r \le n$, then the non-zero singular values of $\mathbf{U}\mathbf{U}^\top - \mathbf{U}^\star\mathbf{U}^{\star\top}$ are $\sin(\theta_1), \ldots, \sin(\theta_d)$, each occurring twice.*

From this we derive the following relationship.

Lemma 2. *For any $1 \le d \le n$,*
$$\|\mathbf{U}\mathbf{U}^\top - \mathbf{U}^\star\mathbf{U}^{\star\top}\| = \|\sin\mathbf{\Theta}\|;$$
$$\frac{1}{\sqrt{2}}\|\mathbf{U}\mathbf{U}^\top - \mathbf{U}^\star\mathbf{U}^{\star\top}\|_{\mathsf{F}} = \|\sin\mathbf{\Theta}\|_{\mathsf{F}}.$$

The next lemma shows the near equivalence of $\mathsf{dist}_{\mathsf{p},|||\cdot|||}$ and $\mathsf{dist}_{|||\cdot|||}$.

Lemma 3. *For any $1 \le d \le n$,*
$$\|\mathbf{U}\mathbf{U}^\top - \mathbf{U}^\star\mathbf{U}^{\star\top}\| \le \min_{\mathbf{O}\in\mathcal{O}^{d\times d}} \|\mathbf{U}\mathbf{O} - \mathbf{U}^\star\| \le \sqrt{2}\|\mathbf{U}\mathbf{U}^\top - \mathbf{U}^\star\mathbf{U}^{\star\top}\|;$$
$$\frac{1}{\sqrt{2}}\|\mathbf{U}\mathbf{U}^\top - \mathbf{U}^\star\mathbf{U}^{\star\top}\|_{\mathsf{F}} \le \min_{\mathbf{O}\in\mathcal{O}^{d\times d}} \|\mathbf{U}\mathbf{O} - \mathbf{U}^\star\|_{\mathsf{F}} \le \|\mathbf{U}\mathbf{U}^\top - \mathbf{U}^\star\mathbf{U}^{\star\top}\|_{\mathsf{F}}.$$

The distance which is easiest to bound will depend on the problem and tools which we choose to use. The good news is that Lemmas 2 and 3 allow us to easily move between them, For the rest of the session, we will concentrate on $\mathsf{dist}_{|||\cdot|||}(\mathbf{U}, \mathbf{U}^\star) := \min_{\mathbf{O}\in\mathcal{O}^{d\times d}} |||\mathbf{U}\mathbf{O} - \mathbf{U}^\star|||$.

2 Perturbation of eigenspaces and eigenvalues.

Perturbation of eigenspaces: the Davis-Kahan sinΘ theorem. One of the most commonly-used theorems in matrix perturbation theory is the Davis-Kahan $\sin \boldsymbol{\Theta}$ theorem (Davis and Kahan, 1970). It bounds change in the eigenspace in terms of the size of the perturbation and the associated eigengap. The original theorem is a deep operator-theoretic result — here, we present a simplified variant of Yu et al. (2015) which is particularly applicable to statistical applications. For simplicity of notation, we consider only the *principal eigenspaces*, although the result generalises to any eigenspace.

Theorem 4 (Davis-Kahan sin$\boldsymbol{\Theta}$ theorem). *For fixed* $1 \leq s \leq r \leq n$*, assume that* $\min\{\lambda^\star_{r-1} - \lambda^\star_r, \lambda^\star_s - \lambda^\star_{s+1}\} > 0$*, then*

$$\|\sin \boldsymbol{\Theta}\| \leq \frac{2 \min\left\{\sqrt{d}\|\mathbf{E}\|, \|\mathbf{E}\|_{\mathsf{F}}\right\}}{\min\{\lambda^\star_{r-1} - \lambda^\star_r, \lambda^\star_s - \lambda^\star_{s+1}\}}$$

The implication of Theorem 4 is that the eigenspaces of real symmetric matrices are stable under small perturbations.

Perturbation of eigenvalues: Weyl's inequality. Weyl's inequality bounds the amount by which an eigenvalue can change by the size of the perturbation.

Theorem 5 (Weyl's inequality). *For* $1 \leq r \leq n$*,*

$$|\lambda_i - \lambda^\star_i| \leq \|\mathbf{E}\|.$$

The implication of Theorem 5 is that the eigenvalues of real symmetric matrices are stable under small perturbations.

3 Concentration of perturbation matrices.

The results from the previous section give bounds on eigenvalues and eigenspaces in terms the size (i.e. some norm) of the perturbation. In this section, we will see how to obtain bounds on the spectral norm of the perturbation $\mathbf{E}$.

The matrix Bernstein inequality. The first concentration inequality we present is the classical Matrix Bernstein inequality.

Theorem 6 (Matrix Bernstein inequality). *Let* $\mathbf{X}_1, \ldots, \mathbf{X}_m \in \mathbb{R}^{n \times n}$ *be independent, symmetric random matrices satisfying* $\mathbb{E}(\mathbf{X}_i) = 0$ *and* $\|\mathbf{X}_i\| \leq M$ *for* $1 \leq i \leq m$*, and let* $\mathbf{X} = \sum_{i=1}^m \mathbf{X}_i$*. For all* $t \geq 0$*,*

$$\mathbb{P}\left(\|\sum_{i=1}^m \mathbf{X}_m\| \geq t\right) \leq 2m \exp\left(\frac{-t^2/2}{v + Mt/3}\right)$$

where $v = \|\sum_{i=1}^m \mathbb{E}(\mathbf{X}_i^2)\|$ *is the matrix variance statistic.*

We can use the Matrix Bernstein inequality to derive a concentration inequality for a symmetric random matrix with independent entries.

Corollary 7. *Let $\mathbf{X} \in \mathbb{R}^{n\times n}$ be a symmetric random matrix whose entries satisfy $\mathbb{E}(\mathbf{X}_{ij}) = 0$ and $|\mathbf{X}_{ij}| \le M$ almost surely for all $1 \le i, j \le n$. Define $v_{ij} = \mathbb{E}(\mathbf{X}_{ij}^2)$ and*

$$v := \max_i \sum_j v_{ij}.$$

Then for any $t \ge 0$,

$$\mathbb{P}\left(\|\mathbf{X}\| \ge t\right) \le 2n^2 \exp\left(\frac{-t^2/2}{v + Mt/3}\right).$$

Proof. Let $\mathbf{X}^{(i,j)}$ be a matrix with $\mathbf{X}_{ij}$ in position (i,j) and (j,i) and zeros elsewhere for $i \le j$. Clearly, $\mathbf{X}^{(i,j)}, i \le j$ are independent and $\mathbf{X} = \sum_{i\le j} \mathbf{X}^{(i,j)}$. It is easily verified that the matrix $\mathbb{E}[(\mathbf{A}^{(i,j)})^2]$ is a diagonal matrix with v_{ij} in positions (i,i) and (j,j) and zeros elsewhere, so

$$\sum_{i\le j} \mathbb{E}\left[(\mathbf{X}^{(i,j)})^2\right] = \operatorname{diag}\left(\sum_j v_{1j}, \ldots, \sum_j v_{nj}\right).$$

Therefore $\| \sum_{i\le j} \mathbb{E}\left[(\mathbf{X}^{(i,j)})^2\right] \| = \max_i \sum_j v_{ij}$, and applying Theorem 6 concludes the proof. □

Sharp bounds for random matrices with independent entries. For matrices with independent entries, the Matrix Bernstein inequality is loose by a logarithmic factor. We next present a result which appeared in Bandeira and Van Handel (2016) (Corollary 3.12) and is sharp.

Theorem 8 (Bandeira and Van Handel (2016), Corollary 3.12)**.** *Let $\mathbf{X} \in \mathbb{R}^{n\times n}$ be a symmetric random matrix whose entries satisfy $\mathbb{E}(\mathbf{X}_{ij}) = 0$ and $|\mathbf{X}_{ij}| \le M$ almost surely for all $1 \le i, j \le n$. Define $v_{ij} = \mathbb{E}(\mathbf{X}_{ij}^2)$ and*

$$v := \max_i \sum_j v_{ij}.$$

Then the exists a universal constant $c > 0$ such that for any $t \ge 0$,

$$\mathbb{P}\left(\|\mathbf{X}\| \ge 4\sqrt{v} + t\right) \le n \exp\left(\frac{-t^2}{cM^2}\right).$$

4 Applications: graph inference.

In this section, we assume that $\mathbf{A}$ is a symmetric adjacency matrix whose upper-triangluar entries are independent Bernoulli random variables with means given by a low-rank matrix $\mathbf{A}^\star$. Specifically, we assume that $\mathbf{A}^\star$ has rank d, so that $\lambda^\star_{d+1} = \cdots = \lambda^\star_n = 0$. For simplicity, we assume that $\lambda^\star_d > 0$ and we allow self-loops, although these assumptions are easily relaxed.

4.1 Concentration of the adjacency matrix.

In this section, we apply the results on matrix concentration to the adjacency matrix of a random graph. We let $\Delta^\star = \max_i \sum_j \mathbf{A}^\star_{ij}$ denote the maximum expected degree and note that $v_{ij} =: \mathbb{E}(\mathbf{E}_{ij}^2) = \mathbf{A}^\star_{ij}(1 - \mathbf{A}^\star_{ij}) \le \mathbf{A}^\star_{ij}$ since $\mathbf{A}^\star_{ij} \le 1$. We therefore have that

$$v =: \max_i \sum v_{ij} \le \max_i \sum_j \mathbf{A}^\star_{ij} = \Delta^\star.$$

In addition, we have that $|\mathbf{E}_{ij}| \le 2|\mathbf{A}_{ij}| \le 2$ almost surely for all $1 \le i, j \le n$.

Apply the matrix Bernstein inequality. Applying the Corollary 7 of the Matrix Bernstein inequality we obtain for all $t \geq 0$

$$\mathbb{P}\left(\|\mathbf{E}\| \geq t\right) \leq 2n^2 \exp\left(\frac{-t^2/2}{\Delta^\star + 2t/3}\right).$$

Assuming $\Delta^\star \geq \log(2n^2/\delta)$, setting $t = 2\sqrt{\Delta^\star \log(2n^2/\delta)}$ we obtain that

$$\begin{aligned}
\mathbb{P}\left(\|\mathbf{E}\| \geq 2\sqrt{\Delta^\star \log(2n^2/t)}\right) &\leq 4n^2 \exp\left(\frac{-2\Delta^\star \log(2n^2/t)}{\Delta^\star + 2\sqrt{\Delta^\star \log(2n^2/t)}/3}\right) \\
&\leq 2n^2 \exp\left(\frac{-2\Delta^\star \log(2n^2/t)}{(1+2/3)\Delta^\star}\right) \\
&\leq 2n^2 \exp\left(-\log(2n^2/t)\right) \\
&\leq \delta.
\end{aligned}$$

From this, we have the following lemma.

Lemma 9. *Providing $\Delta^\star \geq \log(2n^2/\delta)$, with probability $1-\delta$,*

$$\|\mathbf{E}\| \leq 2\sqrt{\Delta^\star \log(2n^2/\delta)}.$$

Applying the sharp bound. Apply Theorem 8, we have that there exists a universal constant $c > 0$ such that for all $t \geq 0$,

$$\mathbb{P}\left(\|\mathbf{X}\| \geq 4\sqrt{\Delta^\star} + t\right) \leq n \exp\left(\frac{-t^2}{c^2}\right).$$

Setting $t = c\sqrt{\log(n/\delta)}$, we obtain

$$\mathbb{P}\left(\|\mathbf{X}\| \geq 4\sqrt{\Delta^\star} + c\sqrt{\log(n/\delta)}\right) \leq \delta.$$

If we additionally assume that $\Delta^\star \geq c\log(n/\delta)$ for a universal constant $c > 0$, we obtain the following result which we state as a lemma.

Lemma 10. *There exists a universal constant $c > 0$ such that providing $\Delta^\star \geq c\log(n/\delta)$, with probability $1-\delta$,*

$$\|\mathbf{E}\| \leq 5\sqrt{\Delta^\star}.$$

Comparing Lemmas 9 and 10, it can be seen, the sharp bound improves on the Matrix Bernstein inequality by a logarithmic factor. For the rest of the section, we assume that $\Delta^\star \geq c\log(n/\delta)$.

4.2 Graph embedding.

We will consider the task of recovering an eigenspace of $\mathbf{A}^\star$ from $\mathbf{A}$, often we interested in the principal eigenspace ($s = 1, r = d$), however not always. For spectral clustering, it is often the second eigenvector which is of interest ($s = r = 2$). We will prove the following theorem:

Theorem 11. *With probability $1-\delta$,*

$$\mathsf{dist}(\mathbf{U}, \mathbf{U}^\star)_\mathsf{F} := \min_{\mathcal{O}^{d\times d}} \|\mathbf{U}\mathbf{O} - \mathbf{U}^\star\|_\mathsf{F} \leq \frac{10d\sqrt{2\Delta^\star}}{\lambda_d^\star}. \tag{2}$$

Notably, the important quantities in this bound are the rank r, maximum expected degree $\Delta^\star$ and the eigengap $\lambda_r^\star$.

Proof. Armed with Lemma 10 and that $\lambda_{d+1}^\star = 0$, we can use the Davis-Kahan $\sin \mathbf{\Theta}$ theorem, along with the inequalities in Lemmas 2 and 3 to obtain a bound on the distance between the eigenspaces of $\mathbf{A}$ and $\mathbf{A}^\star$. We have with probability $1-\delta$,

$$\mathsf{dist}(\mathbf{U}, \mathbf{U}^\star) \leq \sqrt{2}\,\|\sin \mathbf{\Theta}\| \leq \frac{2^{3/2}\sqrt{d}\|\mathbf{E}\|}{\lambda_d^\star} \leq \frac{10\sqrt{2d\Delta^\star}}{\lambda_d^\star}. \tag{3}$$

The Frobenius norm error can be bounded using the standard matrix norm inequality

$$\mathsf{dist}(\mathbf{U}, \mathbf{U}^\star)_{\mathsf{F}} \leq \sqrt{d}\,\mathsf{dist}(\mathbf{U}, \mathbf{U}^\star).$$

□

4.3 Link prediction / anomaly detection.

In applications such as link prediction (locating edges which *should have* occurred, but *didn't*) and anomaly detection (locating edges which *did* occur but *shouldn't have*), it is of interest to directly estimate $\mathbf{A}^\star$ from $\mathbf{A}$. A natural candidate is the low rank eigendecomposition of $\mathbf{A}$,

$$\widehat{\mathbf{A}} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^\top.$$

using $s = 1, r = d$. We will prove the following theorem.

Theorem 12. *With probability* $1-\delta$*,*

$$\|\widehat{\mathbf{A}} - \mathbf{A}^\star\|_{\mathsf{F}} \leq 10\sqrt{d\Delta^\star}.$$

Proof. Using that $\lambda_{d+1}^\star = \cdots, \lambda_n^\star = 0$, we see from Weyl's inequality that

$$|\lambda_i| = |\lambda_i - \lambda_i^\star| \leq \|\mathbf{E}\| \leq 5\sqrt{\Delta^\star}, \qquad \text{for } d+1 \leq i \leq n.$$

From this and the triangle inequality we have that with probability $1-\delta$,

$$\begin{aligned}
\|\widehat{\mathbf{A}} - \mathbf{A}^\star\| &\leq \|\widehat{\mathbf{A}} - \mathbf{A}\| + \|\mathbf{A} - \mathbf{A}^\star\| \\
&= \|\mathbf{E}\| + |\lambda_{d+1}| \leq 2\|\mathbf{E}\| \\
&\leq 10\sqrt{\Delta^\star}.
\end{aligned}$$

To bound the Frobenius norm error, we can use that $\widehat{\mathbf{A}} - \mathbf{A}^\star$ has rank at most $2d$ and apply standard matrix norm inequalities to obtain that with probabilty $1-\delta$,

$$\begin{aligned}
\|\widehat{\mathbf{A}} - \mathbf{A}^\star\|_{\mathsf{F}} &\leq \sqrt{d}\|\widehat{\mathbf{A}} - \mathbf{A}^\star\| \\
&\leq 10\sqrt{d\Delta^\star}
\end{aligned}$$

which concludes the proof. □

4.4 Community recovery under the stochastic block model.

We now consider the task of recovering communities under a random graph model known as the stochastic block model (Holland et al., 1983). In this simplest case, each node belongs to one of two communities and edges occur independently with probability a if two nodes are in the same community and probability $b < a$ if they are in different communities.

Each node is assigned a binary variable $z_1^\star, \ldots, z_n^\star \in \{-1, +1\}$ denoting its community membership. We assume that there are an equal number of nodes in each community ordered so that $z_i^\star = -1$ if $i \leq n/2$ and $z_i^\star = +1$ if $i > n/2$. The population matrix therefore has entries

$$\mathbf{A}_{ij}^\star = \begin{cases} a & \text{if } z_i^\star = z_j^\star, \\ b & \text{if } z_i^\star \neq z_j^\star, \end{cases} \qquad 1 \leq i, j \leq n,$$

has rank 2, and it can be easily verified that its eigenvalues are

$$\lambda_1^\star = \frac{(a+b)n}{2}, \qquad \lambda_2^\star = \frac{(a-b)n}{2},$$

and its eigenvectors are

$$u_1^\star = \frac{1}{\sqrt{n}}(1, \ldots, 1)^\top, \qquad u_2^\star = \frac{1}{\sqrt{n}}(\underbrace{1, \ldots, 1}_{n/2 \text{ times}}, \underbrace{-1, \ldots, -1}_{n/2 \text{ times}})^\top.$$

Clearly, the second population eigenvector contains the the information about community memberships. This motivates the community estimate

$$z_i = \operatorname{sgn}(u_{2,i}) = \begin{cases} 1, & \text{if } u_{2,i} > 0, \\ -1 & \text{if } u_{2,i} \leq 0. \end{cases}$$

for $1 \leq i \leq n$. We bound the misclustering rate with the following theorem.

Theorem 13. *With probability $1 - \delta$, the misclustering rate satisfies*

$$\min_{w \in \pm 1} \frac{1}{n} \sum_{i=1}^n \mathbb{I}\{w z_i = z_i^\star\} \leq \frac{20\sqrt{(a+b)}}{\sqrt{n}\min\{a+b, a-b\}}.$$

Proof. Observe that for any $w z_i \neq z_i^\star$, we have that $\operatorname{sgn}(w u_{2,i}) \neq \operatorname{sgn}(u_{2,i}^\star)$ and therefore that $|w u_{2,i} - u_{2,i}^\star| \geq |u_{2,i}^\star| = 1/\sqrt{n}$. Therefore, we have

$$\begin{aligned} \frac{1}{n}\sum_{i=1}^n \mathbb{I}\{w z_i \neq z_i^\star\} &= \frac{1}{n}\sum_{i=1}^n \mathbb{I}\left\{|w u_{2,i} - u_{2,i}^\star| \geq 1/\sqrt{n}\right\} \\ &\leq \frac{1}{n}\sum_{i=1}^n \mathbb{I}\left\{n(w u_{2,i} - u_{2,i}^\star)^2 \geq 1\right\} \\ &\leq \|w u_2 - u_2^\star\|_2^2, \end{aligned}$$

and by (3) we have, providing $\Delta^\star = \frac{n}{2}(a+b) \geq c\log(n/\delta)$, for a universal constant $c > 0$, then there exists $w \in \pm 1$ such that with probability $1 - \delta$,

$$\frac{1}{n}\sum_{i=1}^n \mathbb{I}\{w z_i = z_i^\star\} \leq \|w u_2 - u_2^\star\|_2^2 \leq \frac{10\sqrt{2\Delta^\star}}{\min\{\lambda_1^\star - \lambda_2^\star, \lambda_2^\star\}} = \frac{20\sqrt{(a+b)}}{\sqrt{n}\min\{a+b, a-b\}},$$

since, by direct calculation, $\min\{\lambda_1^\star - \lambda_2^\star, \lambda_2^\star\} = \frac{n}{2}\min\{a+b, a-b\}$. $\square$

References

Bandeira, A. S. and Van Handel, R. (2016). Sharp nonasymptotic bounds on the norm of random matrices with independent entries. *The Annals of Probability*, 44(4):2479–2506.

Chen, Y., Chi, Y., Fan, J., Ma, C., et al. (2021). Spectral methods for data science: A statistical perspective. *Foundations and Trends in Machine Learning*, 14(5):566–806.

Davis, C. and Kahan, W. M. (1970). The rotation of eigenvectors by a perturbation. iii. *SIAM Journal on Numerical Analysis*, 7(1):1–46.

Holland, P. W., Laskey, K. B., and Leinhardt, S. (1983). Stochastic blockmodels: First steps. *Social networks*, 5(2):109–137.

Stewart, G. W. and Sun, J. G. (1990). *Matrix Perturbation Theory.* Academic Press.

Yu, Y., Wang, T., and Samworth, R. J. (2015). A useful variant of the davis–kahan theorem for statisticians. *Biometrika*, 102(2):315–323.