Proceedings of Machine Learning Research vol 195:1–66, 2023 36th Annual Conference on Learning Theory

Private Covariance Approximation and Eigenvalue-Gap Bounds for Complex Gaussian Perturbations

Oren Mangoubi *Worcester Polytechnic Institute*

Nisheeth K. Vishnoi *Yale University*

Editors: Gergely Neu and Lorenzo Rosasco

Abstract

We consider the problem of approximating a $d \times d$ covariance matrix M with a rank-k matrix under (ε, δ)-differential privacy. We present and analyze a complex variant of the Gaussian mechanism and show that the Frobenius norm of the difference between the matrix output by this mechanism and the best rank-k approximation to M is bounded by roughly $\tilde{O}(\sqrt{kd})$, whenever there is an appropriately large gap between the k'th and the $k+1$'th eigenvalues of M. This improves on previous work that requires that the gap between every pair of top-k eigenvalues of M is at least $\sqrt{d}$ for a similar bound. Our analysis leverages the fact that the eigenvalues of complex matrix Brownian motion repel more than in the real case, and uses Dyson's stochastic differential equations governing the evolution of its eigenvalues to show that the eigenvalues of the matrix M perturbed by complex Gaussian noise have large gaps with high probability. Our results contribute to the analysis of low-rank approximations under average-case perturbations and to an understanding of eigenvalue gaps for random matrices, which may be of independent interest.

1. Introduction

Given a matrix $M \in \mathbb{R}^{d\times d}$, consider the following basic problem of finding a rank-k matrix X that is closest to M in Frobenius norm (Bhatia, 2013; Blum et al., 2020): $\min_{X:\, \mathrm{rank}(X)\leq k} \|M - X\|_F$. Of interest is the case when M is the covariance matrix of a data matrix: Given a matrix $A \in \mathbb{R}^{m\times d}$, consisting of m individuals with d-dimensional features, $M = A^\top A$. Such an M is positive semi-definite (PSD) and has non-negative eigenvalues $\sigma_1 \geq \cdots \geq \sigma_d \geq 0$. The solution to the rank-k optimization problem above is well-known (Bhatia, 2013): It is given by $M_k := V\Gamma_k V^\top$ where $\Gamma_k := \mathrm{diag}(\sigma_1, \ldots, \sigma_k, 0, \ldots, 0)$ and V is a matrix whose columns are the eigenvectors of M.

In several modern applications of this low-rank approximation problem, the rows of A encode sensitive features of individuals and the release of a low-rank approximation to M may reveal these features; see Bennett and Lanning (2007). In such contexts, differential privacy (DP) has been employed to quantify the extent to which an algorithm preserves privacy (Dwork et al., 2006b) and, in particular, algorithms for low-rank covariance matrix approximation under differential privacy have been widely studied (Blum et al., 2005; Dwork et al., 2006b; Kapralov and Talwar, 2013; Blocki et al., 2012; Dwork et al., 2014; Upadhyay, 2018; Sheffet, 2019; Mangoubi et al., 2022; Mangoubi and Vishnoi, 2022). In the low-rank approximation problem, a randomized mechanism $\mathcal{A}$ is said to be (ε, δ)-differentially private for privacy parameters $\varepsilon, \delta \geq 0$ if for all "neighboring" matrices $M, M' \in \mathbb{R}^{d\times d}$, and any measurable subset S of the range of $\mathcal{A}$, we have

$$\mathbb{P}(\mathcal{A}(M) \in S) \leq e^\varepsilon \mathbb{P}(\mathcal{A}(M') \in S) + \delta. \tag{1}$$

M and M' are said to be neighbors if their corresponding data matrices $A, A' \in \mathbb{R}^{m\times d}$ differ by at most one row ($M' = M - uu^\top + vv^\top$ where, for each row vector, $\|u\|_2, \|v\|_2 \leq 1$). To measure the "utility" of the mechanism, the Frobenius-norm distance $\|\mathcal{A}(M) - M_k\|_F$ is often used; see Chaudhuri et al. (2012); Dwork et al. (2014); Amin et al. (2019); Mangoubi and Vishnoi (2022) for a discussion on the rationale for this norm. This leads to the problem of designing an (ε, δ)-differentially private mechanism that, given a covariance matrix M with eigenvalues $\sigma_1 \geq \cdots \geq \sigma_d \geq 0$, outputs a rank-$k$ matrix Y that minimizes $\|Y - M_k\|_F$.

Dwork et al. (2014) present the (real) Gaussian mechanism which ensures (ε, δ)-DP by adding a real symmetric matrix E with i.i.d. Gaussian entries from $N(0, \sqrt{\log 1/\delta}/\varepsilon)$ to M and then outputting the Frobenius-norm minimizing rank-k approximation to $M + E$. Roughly speaking, they show that the output Y of their mechanism satisfies $\|M - Y\|_F - \|M - M_k\|_F = \tilde{O}(k\sqrt{d})$ w.h.p.

Mangoubi and Vishnoi (2022) also consider the Gaussian mechanism and prove that, if the top-k eigenvalue gaps of M satisfy $\sigma_i - \sigma_{i+1} \geq \tilde{\Omega}(\sqrt{d})$ for every $i \leq k$, then the utility bound (for a stronger metric $\|Y - M_k\|_F$) can be improved by a factor of $\sqrt{k}$ to, roughly, $\|Y - M_k\|_F \leq \tilde{O}(\sqrt{kd})$ in expectation whenever the k'th eigenvalue gap of M is at least $\sigma_k - \sigma_{k+1} \geq \Omega(\sigma_k)$. The assumption of a large k'th eigenvalue gap is common in the matrix approximation literature, as a large gap $\sigma_k - \sigma_{k+1}$ in the eigenvalues of the covariance matrix M for some $k < d$ can motivate the problem of finding a rank-k approximation because it suggests the presence of a useful rank-k "signal" M_k which one wishes to extract from the background "noise" in the data (see e.g. Dwork et al. (2014)). Such a gap is also necessary for good rank-k approximations to exist under the stronger utility metric $\|Y - M_k\|_F$. However, in many datasets where one has a large k'th eigenvalue gap for some $k < d$, the other gaps in the top-k eigenvalues are oftentimes small or even zero (for instance, this happens whenever two or more of the features in a dataset are highly correlated). Thus, Mangoubi and Vishnoi (2022) left as an open problem to investigate if the assumption that *all* the top-k eigenvalues of M have large gaps can be removed.

Our contributions. We show that a *complex* Gaussian mechanism (Algorithm 1) can give a utility bound $\|Y - M_k\|_F \leq \tilde{O}(\sqrt{kd})$ *without* assuming that the gaps in all of the top-k eigenvalues of M are at least $\tilde{\Omega}(\sqrt{d})$; see Theorem 1. As in Mangoubi and Vishnoi (2022), we view the addition of Gaussian noise $B(t)$ as a stochastic process $M + B(t)$, whose eigenvalues $\gamma_i(t)$ and eigenvectors evolve according to the stochastic differential equations (SDE) discovered by Dyson (1962); see (3). This leads to a bound on the utility which includes terms of the form $(\lambda_i - \lambda_{i+1})^2/(\gamma_i(t) - \gamma_{i+1}(t))^2$ and $(\lambda_i - \lambda_{i+1})/(\gamma_i(t) - \gamma_{i+1}(t))^2$ integrated over time, where, roughly speaking, $\lambda_1, \ldots, \lambda_k$ are the eigenvalues of the rank-k approximation and the $\gamma_i(t)$ are the eigenvalues of the matrix $M + B(t)$. If one does not assume that the initial gaps are $\tilde{\Omega}(\sqrt{d})$, the gaps $\gamma_i(t) - \gamma_{i+1}(t)$ may become very small at some times t, causing the terms in the utility bound to become very large. To bypass this, the following novel steps are employed here: 1) Rather than analyzing the utility by considering the output eigenvalues λ_i to be fixed numbers, we instead set the top-k output eigenvalues λ_i to be *dynamically* changing over time and equal to $\gamma_i(t)$, making the gaps in the numerators small at exactly those times when the denominators are small. 2) We then leverage the fact that our mechanism adds *complex* Gaussian noise, which implies that $\gamma_i(t)$s evolve by repelling each other with a stronger "force" than when only real noise is added, to show that the gaps between the eigenvalues satisfy a high-probability lower bound of $\mathbb{P}(\gamma_i(t) - \gamma_{i+1}(t) \leq s/\sqrt{td}) \leq \tilde{O}(s^3)$; see Lemma 4. Our bound improves, in the setting where the random matrix is Gaussian, on previous eigenvalue gap bounds of Nguyen et al. (2017) where the bound on the probability decays as $O(s^2)$, which is insufficient for our application. We prove Lemma 4 by first showing that one can reduce the problem of bounding the gaps $\gamma_i(t) - \gamma_{i+1}(t)$ to the special case when the initial eigenvalues are all zero; see Lemma 3.

2. Main results

For an $S \in \mathbb{C}^{d\times d}$, denote by S^* its conjugate transpose. S is Hermitian if $S = S^*$. For any Hermitian matrix $S \in \mathbb{C}^{d\times d}$ with spectral decomposition $S = U\Lambda U^*$ where $\Lambda := \mathrm{diag}(\lambda_1, \ldots, \lambda_d)$ is a diagonal matrix containing the eigenvalues $\lambda_1 \geq \cdots \geq \lambda_d$ of S and $U := [u_1, \ldots, u_d]$ a unitary matrix containing the eigenvectors $u_1, \ldots, u_d$ of S, denote by $\Lambda_k := \mathrm{diag}(\lambda_1, \ldots, \lambda_k, 0, \ldots, 0)$ and by $S_k := U\Lambda_k U^*$ the best rank-k approximation of S. Denote by $U_k := [u_1, \ldots, u_k]$ the $d \times k$ matrix of the top-k eigenvectors of U.

2.1. Private covariance approximation and complex random perturbations

Our first result (Theorem 1) analyzes the complex Gaussian mechanism (Algorithm 1) and provides an upper bound on the expected Frobenius distance utility of this mechanism for the problem of rank-k covariance approximation. In the following, the $\tilde{O}$ notation hides factors of $(\log d)^{\log\log d}$.

Algorithm 1: Complex Gaussian Mechanism

Input: $\varepsilon, \delta > 0$, $d, k \in \mathbb{N}$. A real symmetric PSD matrix $M \in \mathbb{R}^{d\times d}$
Output: A real symmetric matrix $Y \in \mathbb{R}^{d\times d}$

Sample matrices $W_1, W_2 \in \mathbb{R}^{d\times d}$ with i.i.d. $N(0,1)$ entries
Set $G := (W_1 + \mathrm{i}W_2) + (W_1 + \mathrm{i}W_2)^*$
Set $\hat{M} := M + \sqrt{T}G$, where $T := \frac{2\log\frac{1.25}{\delta}}{\varepsilon^2}$
Compute the diagonalization $\hat{M} = \hat{V}\hat{\Sigma}\hat{V}^*$ with eigenvalues $\hat{\sigma}_1 \geq \cdots \geq \hat{\sigma}_d$
Set $\hat{M}_k := \hat{V}\hat{\Sigma}_k\hat{V}^*$, where $\hat{\Sigma}_k := \mathrm{diag}(\hat{\sigma}_1, \ldots, \hat{\sigma}_k, 0, \ldots, 0)$
Output Y to be the real part of $\hat{M}_k$

Theorem 1 ((ε, δ)-DP rank-k covariance approximation) *Given $\varepsilon, \delta > 0$, there is an (ε, δ)-differentially private algorithm (Algorithm 1) that, on input $k > 0$ and a real symmetric matrix $M \in \mathbb{R}^{d\times d}$ with eigenvalues $\sigma_1 \geq \cdots \geq \sigma_d \geq 0$ satisfying $\sigma_k - \sigma_{k+1} \geq 4\frac{\sqrt{2\log\frac{1.25}{\delta}}}{\varepsilon}\sqrt{d}$ and $\sigma_1 \leq d^{50}$, outputs a rank-k matrix $Y \in \mathbb{R}^{d\times d}$ such that*

$$\sqrt{\mathbb{E}[\|M_k - Y\|_F^2]} \leq \tilde{O}\left(\sqrt{kd}\frac{\sigma_k}{\sigma_k - \sigma_{k+1}} \times \frac{\sqrt{\log\frac{1}{\delta}}}{\varepsilon}\right).$$

M_k is the Frobenius-norm minimizing rank-k approximation to M.

The proof of Theorem 1 appears in Section B. We note that the requirement in Theorem 1 that $\sigma_1 \leq d^{50}$ is an artifact of the proof, and can be replaced with $\sigma_1 \leq d^C$ for any large universal constant $C > 0$. Theorem 1 only requires a lower bound $\sigma_k - \sigma_{k+1} \geq \tilde{\Omega}\left(\frac{\sqrt{d\log\frac{1}{\delta}}}{\varepsilon}\right)$ on the k'th eigenvalue gap. Thus, it improves on the main result of Mangoubi and Vishnoi (2022) (their Corollary 2.3) as it no longer relies on their assumption (Assumption 2.1) that all the gaps in the top-k eigenvalues of M are at least $\sigma_i - \sigma_{i+1} \geq \tilde{\Omega}\left(\frac{\sqrt{d\log\frac{1}{\delta}}}{\varepsilon}\right)$ for all $i \leq k$.

Moreover, for matrices M whose k'th eigengap satisfies $\sigma_k - \sigma_{k+1} = \Omega(\sigma_k)$, Theorem 1 improves by a factor of $\sqrt{k}$ on the bound in Theorem 7 of Dwork et al. (2014) which says the output Y of their mechanism satisfies $\|Y - M\|_F - \|M_k - M\|_F = \tilde{O}(k\sqrt{d})$ w.h.p. This is because an upper bound on $\|Y - M_k\|_F$ implies an upper bound on their utility measure by the triangle inequality. The reason why Dwork et al. (2014) is independent of the gap $\sigma_k - \sigma_{k+1}$ while our bound depends on the ratio $\frac{\sigma_k}{\sigma_k - \sigma_{k+1}}$ is due to the fact that if, e.g., $\sigma_k - \sigma_{k+1} = 0$ an arbitrarily small Gaussian perturbation to M would lead to a perturbation of $\|\hat{V}_k - V_k\|_2 = \Omega(1)$ w.h.p., where $\hat{V}_k$ and V_k are the matrices containing the top-k eigenvectors of $\hat{M}$ and M respectively. Roughly speaking, this, in turn, would lead to a perturbation of at least $\|Y - M_k\|_F \geq \|\sigma_k \hat{V}_k \hat{V}_k^* - \sigma_k V_k V_k^*\|_2 \geq \Omega(\sigma_k)$. The techniques used in the proof of Theorem 1 can also be used to improve this Frobenius utility to $\tilde{O}(\sqrt{kd})$ *without* assuming the eigengap condition; see Section H. Finally, the expectation bound in Theorem 1 immediately implies a high probability bound with polynomial decay in the probability via Chebyshev's inequality; see also the discussion at the end of Section 3.2.

The privacy guarantee in Theorem 1 follows directly from prior works on the (real) Gaussian mechanism (see Section B) The utility bound in Theorem 1 follows from the following "average-case" matrix perturbation bound for complex Gaussian random perturbations.

Theorem 2 (Frobenius bound for complex Gaussian perturbations) *Suppose we are given $k > 0$, $T > 0$, and a Hermitian matrix $M \in \mathbb{C}^{d\times d}$ with eigenvalues $\sigma_1 \geq \cdots \geq \sigma_d \geq 0$. Let $\hat{M} := M + \sqrt{T}[(W_1 + \mathrm{i}W_2) + (W_1 + \mathrm{i}W_2)^*]$ where $W_1, W_2 \in \mathbb{R}^{d\times d}$ have entries which are independent $N(0,1)$ random variables. Denote, respectively, by $\sigma_1 \geq \cdots \geq \sigma_d$ and $\hat{\sigma}_1 \geq \ldots \geq \hat{\sigma}_d \geq 0$ the eigenvalues of M and $\hat{M}$, and by V and $\hat{V}$ the matrices whose columns are the corresponding eigenvectors of M and $\hat{M}$. Moreover, let $M_k := V\Gamma_k V^*$ and $\hat{M}_k := \hat{V}\hat{\Gamma}_k\hat{V}^*$ be the rank-k approximations of M and $\hat{M}$, where $\Gamma_k := \mathrm{diag}(\sigma_1, \ldots, \sigma_k, 0, \ldots, 0)$ and $\hat{\Gamma}_k := \mathrm{diag}(\hat{\sigma}_1, \ldots, \hat{\sigma}_k, 0, \ldots, 0)$. Suppose that $\sigma_k - \sigma_{k+1} \geq 4\sqrt{Td}$ and that $\sigma_1 \leq d^{50}$. Then we have*

$$\sqrt{\mathbb{E}\left[\left\|\hat{M}_k - M_k\right\|_F^2\right]} \leq \tilde{O}\left(\sqrt{kd}\frac{\sigma_k}{\sigma_k - \sigma_{k+1}} \times \sqrt{T}\right).$$

The proof of Theorem 2 is presented in Section C. The requirement $\sigma_1 \leq d^{50}$ can be replaced with $\sigma_1 \leq d^C$ for any large universal constant $C > 0$.

One can also compare the bound in this theorem to those obtained by deploying deterministic eigenvector perturbation bounds such as those of Davis and Kahan (1970), which say roughly that given any Hermitian matrices M, E, one has

$$\|\hat{V}_k \hat{V}_k^* - V_k V_k^*\|_2 \leq \tfrac{\|E\|_2}{\sigma_k - \sigma_{k+1}}, \tag{2}$$

where V_k and $\hat{V}_k$ are, respectively, the top-k eigenvectors of the input matrix M and the perturbed matrix $\hat{M} := M + E$. Applying (2), together with concentration bounds which say that the spectral norm of a random matrix G with i.i.d. $N(0,1)$ entries satisfies $\|G\|_2 = O(\sqrt{d})$ w.h.p. (e.g. Theorem 4.4.5 of Vershynin (2018)), one can obtain a bound on the Frobenius distance of $\|\hat{M}_k - M_k\|_F \leq O((k^{1.5}\sqrt{d} + \frac{\sigma_k}{\sigma_k - \sigma_{k+1}}\sqrt{k}\sqrt{d})\sqrt{T})$ w.h.p. when $\sigma_k - \sigma_{k+1} \geq \Omega(\sqrt{Td})$ (see e.g. Appendix C of Mangoubi and Vishnoi (2022) for details). Theorem 2 improves (in expectation) on this bound by a factor of k when e.g. $\sigma_k - \sigma_{k+1} = \Omega(\sigma_k)$.

O'Rourke et al. (2018) provide eigenvector perturbation bounds for matrices $\hat{M} := M + E$ when the input matrix M is a deterministic low-rank matrix of rank $r \geq k$ and the matrix E is a random matrix. In particular, their Theorem 18 improves w.h.p. on the deterministic bound (2) for certain inputs M of sufficiently low rank and random matrices E. If one directly applies their bound to the setting when E is a Hermitian Gaussian random matrix (e.g., by plugging in their bound in place of (2) in Appendix C of Mangoubi and Vishnoi (2022)), one obtains a bound on the quantity $\|\hat{M}_k - M_k\|_F$. Theorem 2 improves (in expectation) on the resulting bound by a factor of $k^{1.5}$ whenever e.g. $\sigma_k - \sigma_{k+1} \geq \Omega(\sqrt{d})$.

While we do not know if our bound in Theorem 2 is tight for every input matrix M, we do verify that it is tight for every $k \leq d$, up to factors of $(\log d)^{\log \log d}$ hidden in the $\tilde{O}$ notation (see Section G for details). Theorem 2 may be of independent interest to other applications where matrix perturbation bounds are used.

2.2. Eigenvalue gaps

To prove Theorem 2, we view the addition of complex Gaussian noise to the matrix M as a matrix-valued Brownian motion. Towards this end, let $W(t) \in \mathbb{C}^{d \times d}$ be a matrix where the real part and complex part of each entry is an independent standard Brownian motion with distribution $N(0, tI_d)$ at time t, and let $B(t) := W(t) + W(t)^*$. Define the Hermitian-matrix valued stochastic process $\Phi(t)$ as follows:

$$\Phi(t) := M + B(t) \qquad \forall t \geq 0. \tag{3}$$

At every time $t > 0$, the eigenvalues $\gamma_1(t), \ldots, \gamma_d(t)$ of $\Phi(t)$ are real-valued and distinct w.p. 1, and (3) induces a stochastic process on the eigenvalues and eigenvectors. The evolution of the eigenvalues can be expressed by the following stochastic differential equations (SDE) (Dyson, 1962):

$$\mathrm{d}\gamma_i(t) = \mathrm{d}B_{ii}(t) + \beta \sum_{j \neq i} \frac{1}{\gamma_i(t) - \gamma_j(t)} \mathrm{d}t \qquad \forall i \in [d], t > 0, \tag{4}$$

where the parameter $\beta = 2$ for the complex case ($\beta = 1$ for the real matrix Brownian motion) (Figure 1). It is well known that, with probability 1, a solution to (4) exists and is unique when coupled to the underlying Brownian motion $B(t)$. Moreover, the paths traversed by the eigenvalues are continuous on all $t \in [0, \infty)$ and do not intersect at any time $t > 0$ (4) (see e.g. Anderson et al. (2010); Inukai (2006); Rogers and Shi (1993)).

The corresponding eigenvector process $u_1(t), \ldots, u_d(t)$, referred to as the Dyson vector flow, is also a "diffusion" and, conditional on the eigenvalue process (4), is given by the following SDEs:

$$\mathrm{d}u_i(t) = \sum_{j \neq i} \frac{\mathrm{d}B_{ij}(t)}{\gamma_i(t) - \gamma_j(t)} u_j(t) - \frac{\beta}{2} \sum_{j \neq i} \frac{\mathrm{d}t}{(\gamma_i(t) - \gamma_j(t))^2} u_i(t) \qquad \forall i \in [d], t > 0. \tag{5}$$

We use (5) to track the utility over time. Letting $\Phi(t) = U(t)\Gamma(t)U(t)^*$ be a spectral decomposition of the Hermitian matrix $\Phi(t)$ at every time t where $\Gamma(t)$ is a diagonal matrix of eigenvalues at time t and $U(t)$ a unitary matrix of eigenvectors. We now define the rank-k matrix $\Psi(t)$ to be the Hermitian matrix with any eigenvalues $\lambda_1(t) \geq \cdots \geq \lambda_d(t)$, where $\lambda_i(t) = \gamma_i(t)$ for $i \leq k$ and $\lambda_i(t) = 0$ for $i > k$, and with eigenvectors $U(t)$: $\Psi(t) := U(t)\Lambda(t)U(t)^*$ for all $t \in [0, T]$, where $\Lambda(t) := \mathrm{diag}(\lambda_1(t), \ldots, \lambda_d(t))$.

Using (5) to obtain SDEs for $\Psi(t)$, and integrating these SDEs, we obtain a formula for the utility $\mathbb{E}[\|\hat{M}_k - M_k\|_F^2] = \mathbb{E}[\|\Psi(T) - \Psi(0)\|_F^2]$ (Lemma 19): $\mathbb{E}[\|\hat{M}_k - M_k\|_F^2] =$

$$O\left(\sum_{i=1}^d \int_0^T \sum_{j \neq i} \mathbb{E}\left[\frac{(\lambda_i(t) - \lambda_j(t))^2}{(\gamma_i(t) - \gamma_j(t))^2}\right] + T\mathbb{E}\left[\left(\sum_{j \neq i} \frac{\lambda_i(t) - \lambda_j(t)}{(\gamma_i(t) - \gamma_j(t))^2}\right)^2\right] \mathrm{d}t\right). \tag{6}$$

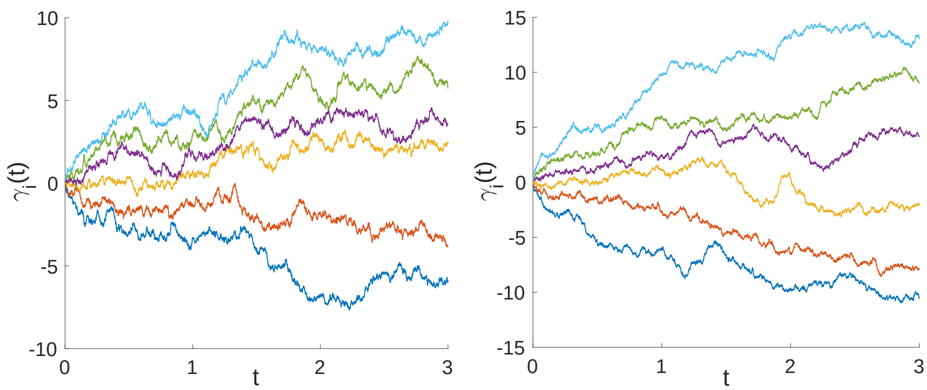

Figure 1: One run of a simulation of the eigenvalues $\gamma_1(t) \geq \cdots \geq \gamma_d(t)$ of Dyson Brownian, in the real case (left) and the complex case (right) with initial condition $\gamma_1(0) = \cdots = \gamma_d = 0$, for $d = 6$. In the complex case, eigenvalue repulsion is stronger and the gaps between the eigenvalues are not as small as in the real case.

After simplifying (6), we are left with terms which are a time-integral of $\mathbb{E}\left[\frac{1}{(\gamma_i(t)-\gamma_j(t))^2}\right]$. To bound these terms, we wish to show that at any time t the gaps $\gamma_i(t) - \gamma_j(t)$ of Dyson Brownian motion are large with high probability.

We first show, in the following lemma, that one can reduce this task to the problem of bounding the gaps of a Dyson Brownian motion initialized at the 0 vector. In the following, we define $\mathcal{W}_d := \{(x_1, \ldots, x_d) \in \mathbb{R}^d : x_1 \geq \cdots \geq x_d\}$.

Lemma 3 (Eigenvalue-gap comparison Lemma) *Let $\xi(t) = (\xi_1(t), \ldots, \xi_d(t))$ and $\gamma(t) = (\gamma_1(t), \ldots, \gamma_d(t))$ be two solutions of (4) coupled to the same underlying Brownian motion $B(t)$, starting respectively from initial conditions $\xi(0), \gamma(0)$. Assume that $\xi_i(0) - \xi_{i+1}(0) \leq \gamma_i(0) - \gamma_{i+1}(0)$ for all $1 \leq i < d$. Then, with probability 1, $\xi_i(t) - \xi_{i+1}(t) \leq \gamma_i(t) - \gamma_{i+1}(t)$ for all $t > 0$ and all $1 \leq i < d$.*

The proof of Lemma 3 appears in Section D and an overview appears in Section 3. Anderson et al. (2010) show a different eigenvalue comparison theorem (their Lemma 4.3.6) which says that if ξ and γ are two coupled Dyson Brownian motions with initial conditions satisfying $\xi_i(0) \leq \gamma_i(0)$ for all $i \in [d]$, then with probability 1, $\xi_i(t) \leq \gamma_i(t)$ at every $t \geq 0$. However, this does not imply the gaps of $\gamma_i(t)$ are at least as large as the corresponding gaps of $\xi_i(t)$ since we could have that $\gamma_i(t) - \gamma_{i+1}(t) < \xi_i(t) - \xi_{i+1}(t)$ even if $\xi_i(t) \leq \gamma_i(t)$ for all i; see also Erdős et al. (2011); Landon and Yau (2017); Lee et al. (2016) for results about the eigenvalues of Dyson Brownian motion and their gaps from non-zero initial conditions.

Lemma 3 implies that it is enough to show the following high probability lower bound on the gaps of the eigenvalues of the GUE random matrix.

Lemma 4 (Eigenvalue gaps of Gaussian Unitary Ensemble (GUE)) *Let $A := G + G^*$ where G is a matrix with i.i.d. complex standard Gaussian entries, and denote by $\eta_1, \ldots, \eta_d$ the eigenvalues of A. Then*

$$\mathbb{P}\left(\eta_i - \eta_{i+1} \leq s\frac{1}{\mathfrak{b}\sqrt{d}}\right) \leq s^3 + \frac{1}{d^{1000}}$$

for all $s > 0$, and for all $1 \leq i < d$, where $\mathfrak{b} = (\log d)^{L \log\log d}$ and L is a universal constant.

The proof of Lemma 4 is presented in Section E and an overview appears in Section 3. We note that the term $\frac{1}{d^{1000}}$ in Lemma 4 can be replaced by $\frac{1}{d^C}$ for any universal constant $C > 0$. Thus, Lemma 4 says that for any $s > d^{-C}$ (where C can be taken to be any large universal constant), the probability that any gap $\eta_i - \eta_{i+1}$ of the GUE random matrix is $\leq \tilde{O}\left(\frac{s}{\sqrt{d}}\right)$ is $O(s^3)$. The s^3 dependence is important to our analysis of the Frobenius-distance utility in Theorem 2, where we wish to bound the time-average of the second moment of the inverse gaps $\mathbb{E}\left[\frac{1}{(\gamma_i(t)-\gamma_j(t))^2}\right]$. Lemma 4 allows us to bound this term by $O(d)$. We use it to bound the utility in (6) by $O(kd)$, thus implying the bound in Theorem 2.

The distribution of the gaps of the GUE in the limit as $d \to \infty$ was studied e.g. in Dyson and Mehta (1963); Tao (2013); Arous and Bourgade (2013), and was also studied non-asymptotically in e.g. Nguyen et al. (2017). However, to the best of our knowledge, we are not aware of a previous (non-asymptotic in d) lower bound on the gaps of GUE random matrices which scales as small as $O(s^3)$. For instance, Nguyen et al. (2017), which studies eigenvalue gaps of Wigner random matrices with sub-Gaussian tails–a more general class of random matrices which includes as a special case the GUE random matrices–show a bound of $\mathbb{P}\left(\eta_i - \eta_{i+1} \leq \frac{s}{\sqrt{d}}\right) \leq O(s^2)$ for any $s > d^{-C}$ where $C > 0$ is a universal constant (Corollary 2.2 in Nguyen et al. (2017), which they can extend to the complex case). On the other hand, we note that Nguyen et al. (2017) focus on matrix universality results that apply to a larger class of random matrices than the GUE random matrices, and that our bound includes additional factors of $(\log d)^{\log\log d}$ hidden in the $\tilde{O}$ notation.

Finally, we note that many results in the random matrix literature rely on explicit determinantal formulas that are only available for complex-valued random matrices (see e.g. Ratnarajah et al. (2004); Johansson (2005); Leake et al. (2021)). While it is possible to simplify the proof of our eigenvalue gap bounds by viewing the eigenvalues of complex Dyson Brownian motion as a determinantal point process, our proofs avoid determinantal methods to allow our results to generalize to the real case. Indeed, Lemma 3 applies to both the real and complex versions of Dyson Brownian motion and the proofs of Lemma 4 and Corollary 24 can be extended to the real case with minor modifications. The main difference is that, for the real case, we would get a term s^2 (which is $= s^{\beta+1}$ for $\beta = 1$) on the r.h.s. of Lemma 4 in place of the term s^3 which appears in the complex version of these results.

3. Overview of the proof of Theorem 2

We bound the Frobenius-distance utility $\|\hat{M}_k - M_k\|_F$, where $\hat{M} := M + G + G^*$ and G is a matrix of i.i.d. standard complex Gaussians. For simplicity, we assume $T = 1$ in this section.

3.1. Bounding the utility of the Gaussian mechanism with Dyson Brownian motion

3.1.1. The previous approach based on Dyson Brownian motion

To obtain their bounds, Mangoubi and Vishnoi (2022) also view the addition of (real) Gaussian noise as a continuous-time matrix Brownian motion $\Phi(t) := M + B(t)$. The eigenvalues $\gamma_i(t)$ and eigenvectors $u_i(t)$ of $\Phi(t)$ evolve according to the same SDEs (4), (5) as in the complex case, but with parameter $\beta = 1$. To bound the utility of the rank-k approximation $\hat{M}_k$, letting $\Phi(t) = U(t)\Gamma(t)U(t)^\top$ be a spectral decomposition of the symmetric matrix $\Phi(t)$ at every time $t \geq 0$, they define a new rank-k matrix-valued stochastic process $\Theta(t) := U(t)\Sigma_k U(t)^\top$ whose eigenvalues are fixed to be the top-k eigenvalues of the input M at every time t. Using the evolution equations (5) for the eigenvectors $u_i(t)$, they obtain an SDE for the matrix-valued process $\Theta(t)$ and integrate this SDE to get an expression for the expected utility:

$$\mathbb{E}\left[\|\Theta(T)-\Theta(0)\|_F^2\right] = \frac{1}{2}\mathbb{E}\left[\left\|\int_0^T \sum_{i=1}^d \sum_{j\neq i}(\lambda_i-\lambda_j)\frac{\mathrm{d}B_{ij}(t)}{\gamma_i(t)-\gamma_j(t)}(u_i(t)u_j^\top(t)+u_j(t)u_i^\top(t))\right\|_F^2\right] + \mathbb{E}\left[\left\|\int_0^T \sum_{i=1}^d \sum_{j\neq i}(\lambda_i-\lambda_j)\frac{\mathrm{d}t}{(\gamma_i(t)-\gamma_j(t))^2}u_i(t)u_i^\top(t)\right\|_F^2\right], \tag{7}$$

where $\lambda_i := \sigma_i$ for $i \le k$ and $\lambda_i := 0$ for $i > k$. The idea is that, roughly speaking, each differential term $\frac{\mathrm{d}B_{ij}(t)}{\gamma_i(t)-\gamma_j(t)}(u_i(t)u_j^\top(t)+u_j(t)u_i^\top(t))$ adds noise to the matrix independently of the other terms at every time t since the stochastic derivatives of the Brownian motions, $\mathrm{d}B_{ij}(t)$, are independent for every i, j, t and independent of the $u_i(s)$ for all current and past times $s \le t$. This allows the contribution of each of these terms to the (squared) Frobenius norm of the first term on the r.h.s. to add up as a sum of squares. Integrating (7) via Ito's Lemma (restated in our preliminaries as Lemma 6), they obtain an expression for the utility as a sum-of-squares of the ratios of the eigenvalue gaps:

$$\mathbb{E}\left[\|\Theta(T)-\Theta(0)\|_F^2\right] = \sum_{i=1}^d \int_0^T \mathbb{E}\left[\sum_{j\neq i}\frac{(\lambda_i-\lambda_j)^2}{(\gamma_i(t)-\gamma_j(t))^2}\right] + T\mathbb{E}\left[\left(\sum_{j\neq i}\frac{\lambda_i-\lambda_j}{(\gamma_i(t)-\gamma_j(t))^2}\right)^2\right]\mathrm{d}t. \tag{8}$$

To bound the gap terms $\gamma_i(t)-\gamma_j(t)$ in (8) for all $i, j \le k$, $i \neq j$, they use Weyl's inequality (restated here as Lemma 11), a deterministic eigenvalue perturbation bound which says that $\gamma_i(t)-\gamma_j(t) \ge \gamma_i(0)-\gamma_j(0)-\|B(t)\|_2$ for all t. Thus, since $\|B(t)\|_2 = \Theta(\sqrt{d})$ w.h.p. for all $t \in [0,T]$, they must require that all gaps in the top-k eigenvalues of M satisfy $\gamma_i(0)-\gamma_{i+1}(0) = \sigma_i - \sigma_{i+1} \ge \Omega(\sqrt{d})$ for every $i \le k$. Simplifying (8), they show that

$$\mathbb{E}\left[\left\|\hat{M}_k - M_k\right\|_F^2\right] \approx \mathbb{E}\left[\|\Theta(T)-\Theta(0)\|_F^2\right] \le \tilde{O}\left(\sqrt{k}\sqrt{d}\frac{\sigma_k}{\sigma_k-\sigma_{k+1}}\right)$$

under their assumption that $\sigma_i - \sigma_{i+1} \ge \Omega(\sqrt{d})$ for every $i \le k$.

3.1.2. OUR APPROACH

To bound the utility of the Gaussian mechanism without any assumptions on the eigenvalue gaps $\sigma_i - \sigma_{i+1}$ for $i \neq k$, we would like to prove bounds on the gaps $\gamma_i(t)-\gamma_j(t)$ which hold even when initial gaps $\gamma_i(0)-\gamma_j(0) = \sigma_i-\sigma_{i+1}$ may not be $\Omega(\sqrt{d})$. Unfortunately, since $\|B(t)\|_2 \ge \Omega(\sqrt{d})$ w.h.p. for $t = \Omega(1)$, we cannot rely on deterministic eigenvalue bounds such as Weyl's inequality, as this would not give any bound on $\gamma_i(t)-\gamma_{i+1}(t)$ unless $\sigma_i-\sigma_{i+1} \ge \Omega(\sqrt{d})$. To bypass this difficulty we would ideally like to obtain probabilistic lower bounds on the eigenvalue gaps $\gamma_i(t)-\gamma_{i+1}(t)$ which hold for any initial conditions on the top-k eigengaps of $\gamma(0)$.

To see what bounds we might hope to show, note that if $\gamma(0) = 0$ then $\gamma(t)$ has the same joint distribution as the eigenvalues $\eta_1, \ldots, \eta_d$ of the rescaled GOE (GUE) matrix $\sqrt{t}(G+G^*)$ where G is a matrix of i.i.d. real (complex) Gaussians. This joint distribution is given by the following formula (Dyson and Mehta, 1963; Ginibre, 1965),

$$f(\eta_1,\ldots,\eta_d) = \frac{1}{R_\beta}\prod_{i<j}|\eta_i-\eta_j|^\beta e^{-\frac{1}{2}\sum_{i=1}^d \eta_i^2}, \tag{9}$$

where $R_\beta := \int \prod_{i<j}|\eta_i-\eta_j|^\beta e^{-\frac{1}{2}\sum_{i=1}^d \eta_i^2}\mathrm{d}\eta_1\cdots\mathrm{d}\eta_d$ is a normalization constant.

From the repulsion factor $|\eta_i - \eta_{i+1}|^\beta$ in the joint distribution of the eigenvalues (9), (and noting that the average eigenvalue gap of the standard GOE/GUE matrix $(G+G^*)$ is $\Theta(\frac{1}{\sqrt{d}})$ w.h.p. since $\|G+G^*\|_2 = \Omega\left(\sqrt{d}\right)$), roughly speaking one might expect that the GOE/GUE eigenvalue gaps satisfy

$$\mathbb{P}\left(\eta_i - \eta_{i+1} \leq \frac{s}{\sqrt{d}}\right) = O\left(\int_0^s z^\beta \mathrm{d}z\right) = O(s^{\beta+1})$$

for all $s \geq 0$, where $\beta = 1$ in the real case and $\beta = 2$ in the complex case. Assuming we can obtain such a bound, we would like to apply these bounds to bound the expectations of the terms on the r.h.s. of (8). The terms on the r.h.s. of (8) with the smallest denominator, and therefore the most challenging to bound, are the terms $\mathbb{E}\left[\frac{(\lambda_i-\lambda_{i+1})^2}{(\gamma_i(t)-\gamma_{i+1}(t))^4}\right]$. Assuming for the moment that we are able to show that

$$\mathbb{P}\left(\gamma_i(t) - \gamma_{i+1}(t) \leq s\frac{\sqrt{t}}{\sqrt{d}}\right) \leq s^{\beta+1}, \tag{10}$$

then we would have the following bound for terms with denominators of order r:

$$\mathbb{E}\left[\frac{1}{(\gamma_i(t)-\gamma_{i+1}(t))^r}\right] = \int_0^\infty \mathbb{P}\left(\gamma_i(t) - \gamma_{i+1}(t) \leq s^{-\frac{1}{r}}\right)\mathrm{d}s \leq \left(\frac{d}{t}\right)^{\frac{r}{2}} \int_0^\infty s^{-\frac{1}{r}(\beta+1)}\mathrm{d}s. \tag{11}$$

For the terms of order $r=2$, the r.h.s. of (11) is $\int_0^\infty s^{-\frac{1}{2}(\beta+1)}\mathrm{d}s = \infty$ in the real case where $\beta = 1$. To bypass this problem, we observe that when the Gaussian noise is complex the integral on the r.h.s. of (11) becomes $\int_0^\infty s^{-\frac{1}{2}(\beta+1)}\mathrm{d}s = O(1)$ since $\beta = 2$ in the complex case. Thus, while in the real case one expects the gaps to be small enough that their inverse second moment $\mathbb{E}\left[\frac{1}{(\gamma_i(t)-\gamma_j(t))^2}\right]$ is infinite, in the complex case the repulsion between eigenvalues allows the gaps to be large enough that the inverse second moment is finite. This motivates replacing the real Gaussian perturbation in the Gaussian mechanism with Complex-valued Gaussian noise (Algorithm 1).

An SDE for a rank-k matrix diffusion with dynamically changing eigenvalues, to track the utility under small initial eigengaps. Unfortunately, for the highest-order terms, of order $r=4$, the r.h.s. of (11) is $\int_0^\infty s^{-\frac{1}{4}(\beta+1)}\mathrm{d}s = \infty$ even in the complex case where $\beta = 2$. To get around this problem we replace the fixed eigenvalues $\lambda_i = \sigma_i$ for $i \leq k$, of the rank-k stochastic process $\Theta(t)$, with eigenvalues $\lambda_i(t)$ which change dynamically over time where at each time $t \geq 0$ we set $\lambda_i(t) = \gamma_i(t)$ for $i \leq k$ and $\lambda_i(t) = 0$ for $i > k$, in the hope that this will lead to cancellations in the highest-order terms. This gives us a new rank-k stochastic process $\Psi(t) := U(t)\Lambda(t)U(t)^*$ with dynamically changing eigenvalues $\Lambda(t) := \mathrm{diag}(\lambda_1(t), \ldots, \lambda_d(t))$. Since $\Psi(T) = \hat{M}_k$ and $\Psi(0) = M_k$, our goal is to bound $\|\hat{M}_k - M_k\|_F = \|\Psi(T) - \Psi(0)\|_F$. Roughly speaking, this would lead to cancellations in the terms on the r.h.s. of (8) at every time $t \geq 0$: the second-order terms would be reduced to constant terms

$$\frac{(\lambda_i(t)-\lambda_j(t))^2}{(\gamma_i(t)-\gamma_j(t))^2} = \frac{(\gamma_i(t)-\gamma_j(t))^2}{(\gamma_i(t)-\gamma_j(t))^2} = 1$$

for $i \neq j$ $i, j \leq k$, and fourth-order terms would be reduced to second-order terms, e.g.,

$$\frac{(\lambda_i(t)-\lambda_{i+1}(t))^2}{(\gamma_i(t)-\gamma_{i+1}(t))^4} = \frac{(\gamma_i(t)-\gamma_{i+1}(t))^2}{(\gamma_i(t)-\gamma_{i+1}(t))^4} = \frac{1}{(\gamma_i(t)-\gamma_{i+1}(t))^2}$$

for $i < k$. This would allow us to obtain a finite bound for the expectation on the r.h.s. of (8).

Towards this end, we first use the equations for the evolution of the eigenvalues (4) and eigenvectors (5) of Dyson Brownian motion to derive an SDE for our new rank-k process $\Psi(t)$ (Lemma 18 and (43)):

$$\mathrm{d}\Psi(t) = \sum_{i=1}^d \lambda_i(t)\mathrm{d}(u_i(t)u_i^*(t)) + (\mathrm{d}\lambda_i(t))(u_i(t)u_i^*(t)) + \mathrm{d}\lambda_i(t)\mathrm{d}(u_i(t)u_i^*(t)), \qquad (12)$$

where

$$\mathrm{d}(u_i(t)u_i^*(t)) = \sum_{j\neq i} \frac{u_i(t)u_j^*(t)\mathrm{d}B_{ij}(t)+u_j(t)u_i^*(t)\mathrm{d}B_{ij}^*(t)}{\gamma_i(t)-\gamma_j(t)} - \frac{(u_i(t)u_i^*(t)-u_j(t)u_j^*(t))\mathrm{d}t}{(\gamma_i(t)-\gamma_j(t))^2},$$

and where $\mathrm{d}\lambda_i(t)$ is given by (4). The last term $\mathrm{d}\lambda_i(t)\mathrm{d}(u_i(t)u_i^*(t))$ in (12) vanishes as it consists only of higher-order differential terms. Applying Itô's lemma to compute the integral $\left\|\int_0^T \mathrm{d}\Psi(t)\right\|_F^2$ for the change in the (squared) Frobenius distance, we get (Lemma 19 and (44) in the Proof of Theorem 2),

$$\mathbb{E}[\|\Psi(T)-\Psi(0)\|_F^2] = \left\|\int_0^T \mathrm{d}\Psi(t)\right\|_F^2 \leq \int_0^T \mathbb{E}\left[\sum_{i=1}^d \sum_{j\neq i} \frac{(\lambda_i(t)-\lambda_j(t))^2}{(\gamma_i(t)-\gamma_j(t))^2}\mathrm{d}t\right]$$
$$+T\int_0^T \mathbb{E}\left[\sum_{i=1}^d \left(\sum_{j\neq i} \frac{\lambda_i(t)-\lambda_j(t)}{(\gamma_i(t)-\gamma_j(t))^2}\right)^2\right]\mathrm{d}t + \int_0^T \sum_{i=1}^k \mathbb{E}\left[\left(\sum_{j\neq i} \frac{1}{\gamma_i(t)-\gamma_j(t)}\right)^2\right]\mathrm{d}t. \qquad (13)$$

Plugging in our choice of $\lambda_i(t)$, we get (Equations (45) and (51) in the proof of Theorem 2),

$$\left\|\int_0^T \mathrm{d}\Psi(t)\right\|_F^2 \leq \sum_{i=1}^k \int_0^T \mathbb{E}\left[\left(k + \sum_{j>k} \frac{(\gamma_i(t))^2}{(\gamma_i(t)-\gamma_j(t))^2}\right)\right] + T\mathbb{E}\Bigg[\left(\sum_{j\neq i: j\leq k} \frac{1}{\gamma_i(t)-\gamma_j(t)}\right)^2$$
$$+\left(\sum_{j>k} \frac{\gamma_i(t)}{(\gamma_i(t)-\gamma_j(t))^2}\right)^2\Bigg] + \mathbb{E}\left[\left(\sum_{j\neq i} \frac{1}{\gamma_i(t)-\gamma_j(t)}\right)^2\right]\mathrm{d}t. \qquad (14)$$

If we can prove the conjectured gap bounds (10), we will have from (11) that

$$\mathbb{E}\left[\frac{1}{(\gamma_i(t)-\gamma_j(t))^2}\right] \leq \frac{d}{t(i-j)^2}$$

for all $i \neq j$ and, more generally, that

$$\mathbb{E}\left[\frac{1}{(\gamma_i(t)-\gamma_j(t))(\gamma_\ell(t)-\gamma_r(t))}\right] \leq \frac{d}{t\min((i-j)^2,(\ell-r)^2)}$$

for all $i \neq j$, $\ell \neq r$. Moreover, if we assume a bound only on the k'th eigenvalue gap of M, $\sigma_k - \sigma_{k+1} \geq \Omega(\sqrt{d})$ (without assuming any bounds on the other eigenvalue gaps of M), we have by Weyl's inequality that $\gamma_k(t) - \gamma_{k+1}(t) \geq \sigma_k - \sigma_{k+1} - \|B(t)\|_2 = \Omega(\sigma_i - \sigma_{i+1})$. Plugging these conjectured probabilistic bounds, together with the worst-case Weyl inequality bounds for the k'th gap $\gamma_k(t) - \gamma_{k+1}(t) \geq \Omega(\sigma_i - \sigma_{i+1})$, into (14) gives (Equation (54) in the proof of Theorem 2)

$$\mathbb{E}\left[\left\|\hat{M}_k - M_k\right\|_F^2\right] = \left\|\int_0^T \mathrm{d}\Psi(t)\right\|_F^2 \leq \tilde{O}\left(kd\frac{\sigma_k^2}{(\sigma_k-\sigma_{k+1})^2}\right).$$

3.2. Bounding the eigenvalue gaps of Dyson Brownian motion

To complete the proof of Theorem 2, we still need to show the conjectured bounds in (10) (or at least show a close approximation to these bounds). We do this by proving Lemmas 3 and 4, and present an overview of their proofs in this section. We start by recalling a few important ideas and results from random matrix theory.

3.2.1. USEFUL IDEAS FROM RANDOM MATRIX THEORY

Starting with (Dyson and Mehta, 1963; Ginibre, 1965), many works have made use of the intuition that the eigenvalues of a random matrix tend to repel each other, and can be interpreted in the context of statistical mechanics as a many-body system of charged particles undergoing a Brownian motion. These particles repel each other with an "electrical force" arising from a potential that decays logarithmically with the distance between pairs of particles (see e.g. Rodriguez-Lujan et al. (2014)). The dynamics of these particles are described by the eigenvalue evolution equations (4) discovered by (Dyson and Mehta, 1963), where the diffusion term $\mathrm{d}\gamma_i(t) = \mathrm{d}B_{ii}(t)$ describes the random component of each particle's motion and the terms $\frac{\beta}{\gamma_i(t)-\gamma_j(t)}$ describe the repulsion between particles; the parameter β can be interpreted either as the strength of the electrical force, or equivalently, as the (inverse) temperature of the system.

Dyson and Mehta (1963) showed that from the evolution equations (4) one can obtain the joint distribution of the eigenvalues of Dyson Brownian motion at equilibrium (9). If one initializes the matrix Brownian motion with all eigenvalues at 0, at every time t the matrix Brownian motion is in equilibrium (after scaling by $\frac{1}{\sqrt{t}}$) and equal in distribution to a GOE or GUE matrix scaled by $\sqrt{t}$, and thus (9) gives an explicit formula for the joint distribution of the eigenvalues of the GOE random matrix (for the real case $\beta = 1$) and GUE random matrix (for the complex case $\beta = 2$).

In the limit as $\beta \to \infty$ (with appropriate rescaling), the temperature of the system can be thought of as going to zero, and the solution to the evolution equations (4) converges to a deterministic solution with particles "frozen" at $\gamma_i(t) = \sqrt{t}\omega_i$ with probability 1 for some $\omega_1, \ldots, \omega_d \in \mathbb{R}$. It has long been observed (Ginibre, 1965; Girko, 1985) that the gaps $\omega_i - \omega_{i+1}$ between these particles is, roughly, $\frac{1}{\sqrt{d}}$ in the "bulk" of the spectrum (i.e., the set of eigenvalues with index $cd < i < d - cd$ for any small constant c), while the particles have larger gaps near the edge of the spectrum.

More recently, Erdős et al. (2012) showed (restated here as Lemma 26) that with high probability, the eigenvalues η of the GOE/GUE random matrices are "rigid" in the sense that each eigenvalue η_i falls within a small distance $\tilde{O}(\min(i, d-i+1)^{-\frac{1}{3}} d^{-\frac{1}{6}})$ of the zero-temperature eigenvalue ω_i, where $\min(i, d-i+1)^{-\frac{1}{3}} d^{-\frac{1}{6}}$ is the average eigengap size in the region of the spectrum near ω_i:

$$|\eta_i - \omega_i| \leq O(\min(i, d-i+1)^{-\frac{1}{3}} d^{-\frac{1}{6}} \log(d)^{\log\log d}), \qquad \forall i \in [d], \qquad \text{w.h.p.} \tag{15}$$

3.2.2. OUR RESULTS ON EIGENVALUE GAPS OF DYSON BROWNIAN MOTION

Overview of proof of Lemma 3. Recall that in Lemma 3 we are given two solutions $\gamma(t)$ and $\xi(t)$ with any initial conditions $\gamma(0), \xi(0) \in \mathcal{W}_d$, and a coupling between $\gamma(t)$ and $\xi(t)$ is defined by setting the underlying Brownian motion which generates each process in (4) to be equal at every time $t \geq 0$. To prove the lemma, we must show that whenever the initial gaps of $\gamma(0)$ are $\geq$ the corresponding initial gaps of $\xi(0)$, $\gamma_i(0) - \gamma_{i+1}(0) \geq \xi_i(0) - \xi_{i+1}(0)$, with probability 1 the gaps of $\gamma(t)$ remain $\geq$ the gaps of the coupled process $\xi(t)$ at every time $t \geq 0$.

The idea behind the proof of Lemma 3 is to consider the net "electrostatic pressure" on each gap $\gamma_i(t) - \gamma_{i+1}(t)$– that is, the difference between the sum of the forces from the eigenvalues $\gamma_j(t)$ for $j \notin \{i, i+1\}$ pushing on the gap $\gamma_i(t) - \gamma_{i+1}(t)$ from the outside to compress it, and the force from the repulsion between the eigenvalues $\gamma_i(t)$ and $\gamma_{i+1}(t)$ pushing to expand the gap. More formally, this net pressure is $\mathrm{d}\left(\gamma_i(t) - \gamma_{i+1}(t)\right) = \mathrm{d}\gamma_i(t) - \mathrm{d}\gamma_{i+1}(t)$ and, thus, we can compute it using (4):

$$\mathrm{d}\gamma_i(t) - \mathrm{d}\gamma_{i+1}(t) = \mathrm{d}B_{i,i}(t) + \sum_{j \neq i} \frac{\beta \mathrm{d}t}{\gamma_i(t)-\gamma_j(t)} - \Big(\mathrm{d}B_{i+1,i+1}(t) + \sum_{j \neq i+1} \frac{\beta \mathrm{d}t}{\gamma_{i+1}(t)-\gamma_j(t)}\Big). \tag{16}$$

Ideally, we would like to show that at any time where all the gaps of $\gamma(t)$ are at least as large as all the gaps of $\xi(t)$, we have $\mathrm{d}\gamma_i(t) - \mathrm{d}\gamma_{i+1}(t) \geq \mathrm{d}\xi_i(t) - \mathrm{d}\xi_{i+1}(t)$. This in turn *would* imply that the gaps of $\gamma(t)$ expand faster (or contract slower) than the corresponding gaps of $\xi(t)$, and hence that the gaps of $\gamma(t)$ remain larger than those of $\xi(t)$ at every time $t \geq 0$. Unfortunately, the opposite may be true: if the eigenvalue gap $\gamma_i(t) - \gamma_{i+1}(t)$ is much larger than the gap $\xi_i(t) - \xi_{i+1}(t)$ then the repulsion between $\gamma_i(t)$ and $\gamma_{i+1}(t)$ pushing to expand the gap $\gamma_i(t) - \gamma_{i+1}(t)$ is much smaller than the repulsion pushing to expand the gap $\xi_i(t) - \xi_{i+1}(t)$.

To solve this problem, we prove Lemma 3 by a contradiction argument. Towards this end, we first define $\tau := \inf\{t \geq 0 : \xi_i(t) - \xi_{i+1}(t) > \gamma_i(t) - \gamma_{i+1}(t) \text{ for some } i \in [d]\}$ to be the first time where for some i, the size of the i'th gap $\xi_i(t) - \xi_{i+1}(t)$ becomes larger than the i'th gap $\gamma_i(t) - \gamma_{i+1}(t)$ of $\gamma(t)$. We assume (falsely), that $\tau < \infty$ and show that this leads to a contradiction.

Since the initial gaps of $\gamma(0)$ are at least as large as those of $\xi(0)$, and since the trajectories $\gamma(t)$ and $\xi(t)$ are continuous w.p. 1, by the intermediate value theorem there must be an $i \in [d]$ such that

$$\gamma_i(\tau) - \gamma_{i+1}(\tau) = \xi_i(\tau) - \xi_{i+1}(\tau), \tag{17}$$

and the other gaps at time τ satisfy $\gamma_j(\tau) - \gamma_{j+1}(\tau) \geq \xi_j(\tau) - \xi_{j+1}(\tau)$ for $j \in [d]$. Plugging (17) into (16), we obtain the difference in net electrostatic pressure on the i'th gap of γ and ξ at time τ:

$$(\mathrm{d}\gamma_i(\tau) - \mathrm{d}\gamma_{i+1}(\tau)) - (\mathrm{d}\xi_i(\tau) - \mathrm{d}\xi_{i+1}(\tau)) = \sum_{j \neq i, i+1} \frac{\beta \mathrm{d}\tau}{\gamma_i(\tau) - \gamma_j(\tau)} - \frac{\beta \mathrm{d}\tau}{\xi_i(\tau) - \xi_j(\tau)} \geq 0. \tag{18}$$

The Brownian motion terms $\mathrm{d}B$ from (16) cancel as we have coupled the processes γ and ξ by setting their underlying Brownian motions $B(t)$ to be equal. The terms $\frac{1}{\gamma_i(\tau) - \gamma_{i+1}(\tau)}$ and $\frac{1}{\xi_i(\tau) - \xi_{i+1}(\tau)}$ arising from (16) which describe repulsion between the i'th and $i+1$'th eigenvalues cancel by (17). Thus, we are only left with the forces from the other eigenvalues pushing to compress the i'th gap of $\xi(\tau)$ and $\gamma(\tau)$ from the outside, which allows us to then show that since the gaps of $\gamma(\tau)$ are at least as large as the corresponding gaps of $\xi(\tau)$ at time τ, the r.h.s. of (18) is ≥ 0 (Proposition D).

Next, we would like to show that (18) implies that the i'th gap of ξ does *not* become larger than the i'th gap of γ at time τ, leading to a contradiction. Unfortunately, (18) is not sufficient to show this, since, if $(\mathrm{d}\gamma_i(\tau) - \mathrm{d}\gamma_{i+1}(\tau)) - (\mathrm{d}\xi_i(\tau) - \mathrm{d}\xi_{i+1}(\tau)) = 0$ we might have that the *second* derivative of the gaps of γ, $(\mathrm{d}^2\gamma_i(\tau) - \mathrm{d}^2\gamma_{i+1}(\tau))$ is strictly smaller than the second derivative of the gaps of ξ, $(\mathrm{d}^2\xi_i(\tau) - \mathrm{d}^2\xi_{i+1}(\tau))$. To overcome this problem, we observe that, since the gaps of $\gamma(\tau)$ are at least the corresponding gaps of $\xi(\tau)$ at time τ, the only way the r.h.s. of (18) could be 0 is if all the gaps of $\gamma(\tau)$ are equal to the corresponding gaps of $\xi(\tau)$. In this case, the gaps would be equal at *every* time t since solutions of Dyson Brownian motion are unique w.r.t. the underlying Brownian motion which defines our coupling (see e.g. Anderson et al. (2010), restated as Lemma 7). Thus, without loss of generality, we may assume that there is at least one j such that the j'th gap of $\gamma(\tau)$ is strictly greater than the j'th gap of $\xi(\tau)$. This in turn implies the r.h.s. of (18) is *strictly* > 0, and hence the i'th gap of γ becomes strictly *larger* than the i'th gap of ξ in an open neighborhood of the time τ. This contradicts the definition of τ, and hence by contradiction, we have $\tau = \infty$, and therefore the gaps of $\gamma(t)$ are $\geq$ the corresponding gaps of $\xi(t)$ at every time $t \geq 0$.

Overview of proof of Lemma 4. Roughly speaking, Lemma 4 requires us to show the conjectured lower bound of $\mathbb{P}\left(\eta_i - \eta_{i+1} \leq \frac{1}{\sqrt{d}} s\right) \leq s^3$ for any i and $s \geq 0$, when $\eta_1, \ldots, \eta_d$ are the eigenvalues of the GUE random matrix. As a first approach, we would ideally like to integrate the formula for the joint eigenvalue density $f(\eta)$ (9) over the set $A(s) := \left\{\eta \in \mathcal{W}_d : \eta_i - \eta_{i+1} \leq \frac{1}{\sqrt{d}} s\right\}$:

$$\mathbb{P}\left(\eta_i - \eta_{i+1} \leq \frac{1}{\sqrt{d}} s\right) = \frac{1}{R_2} \int_{\{\eta \in \mathcal{W}_d : \eta_i - \eta_{i+1} \leq \frac{1}{\sqrt{d}} s\}} \prod_{\ell < j} |\eta_\ell - \eta_j|^2 e^{-\frac{1}{2} \sum_{\ell=1}^d \eta_\ell^2} \mathrm{d}\eta. \tag{19}$$

Unfortunately, we do not know of a closed-form expression for the d-dimensional integral (19).

To get around this, we observe that, given any η such that $\eta_i - \eta_{i+1} \leq \frac{1}{\sqrt{d}} s$, if we can somehow find a map $\phi : \mathcal{W}_d \to \mathcal{W}_d$ such that the term $|\phi(\eta)[i] - \phi(\eta)[i+1]| \geq \frac{1}{s}|\eta_i - \eta_{i+1}|$, and all other terms ($|\phi(\eta_j) - \phi(\eta_\ell)|$ for $(j,\ell) \neq (i, i+1)$ and $e^{-\frac{1}{2}\sum_{\ell=1}^d \eta_\ell^2}$) in the joint density remain unchanged, then we would have that $f(\phi(\eta)) \geq \frac{1}{s^2} f(\eta)$. Moreover, roughly speaking, one might hope that, since ϕ expands one of the gaps by $\frac{1}{s}$ and leaves all the other gaps unchanged, the map ϕ would be invertible and the Jacobian determinant of such a map would satisfy $\det(J_\phi(\eta)) \geq \frac{1}{s}$ for all $\eta \in \mathcal{W}_d$. This in turn would imply that the r.h.s. of (19) would satisfy

$$\mathbb{P}\left(\eta_i - \eta_{i+1} \leq \frac{1}{\sqrt{d}} s\right) = s^3 \int_{A(s)} f(\eta) \frac{1}{s^3} \mathrm{d}\eta \leq s^3 \int_{A(s)} f(\eta) \frac{f(\phi(\eta))}{f(\eta)} \det(J_\phi(\eta)) \mathrm{d}\eta \leq s^3. \tag{20}$$

The last step holds since ϕ is injective and f is a probability density, implying the integral is ≤ 1.

Unfortunately, one can easily see that there does not exist a map ϕ which expands the i'th gap term $|\eta_i - \eta_{i+1}|$ in the joint eigenvalue density (9) by $\frac{1}{s}$, but leaves all other terms unchanged. This is because, to expand $|\eta_i - \eta_{i+1}|$ but leave the other gap terms unchanged, one would, e.g., have to translate the other eigenvalues η_j for $j \leq i$ aside by an amount $(\frac{1}{s} - 1)|\eta_i - \eta_{i+1}|$. To circumvent this problem, we instead consider a different map $\phi : \mathcal{W}_d \to \mathcal{W}_d$ which, roughly speaking, expands the eigenvalue gap $\eta_i - \eta_{i+1}$ by a factor of $\frac{1}{s}$, leaves all other gaps unchanged, and translates the eigenvalues of η_j for $j \leq i$ to the left by an amount $\frac{1}{s}(\eta_i - \eta_{i+1})$ to make room for the expanded eigenvalue gap (see equations (121)-(123) for the full definition of ϕ). Since $\eta_i = \Theta(\sqrt{d})$ w.h.p., when e.g. $|\eta_i - \eta_{i+1}| \geq \Theta(\frac{1}{\sqrt{d}})$ this would decrease the exponential term in the joint density by a factor of

$$e^{-\frac{1}{2}\sum_{j=1}^i (\phi(\eta)[i] - \eta_i)^2} \approx e^{-\frac{1}{2}\sum_{j=1}^i (\phi(\eta)[i] - \eta_i)\sqrt{d}} \geq e^{-\frac{1}{2}\sum_{j=1}^i \frac{\sqrt{d}}{\sqrt{d}}} = e^i.$$

For $i \leq O(1)$, this is not an issue as then one has $e^i = O(1)$ and, hence,

$$\frac{f(\phi(\eta))}{f(\eta)} \geq \Omega\left(\frac{1}{s^2}\right)$$

(Lemma 35). Roughly speaking this fact, together with a bound on the Jacobian determinant of ϕ (Proposition 33 which says $\det(J_\phi(\eta)) \geq \frac{1}{s}$) and since ϕ is injective (Proposition 32), allows us to use the above map ϕ to show that (20) holds whenever the i'th eigenvalue gap is near the edge of the spectrum ($i \leq \tilde{O}(1)$).

To bound $\gamma_i - \gamma_{i+1}$ for $i \geq \tilde{\Omega}(1)$, which are not near the edge of the spectrum, we will use the rigidity property of the GUE eigenvalues (15) ((Erdős et al., 2012); restated here as Lemma 26). Roughly, this rigidity property says that none of the eigenvalues η_i fall more than a distance $\mathfrak{b} = O(\log(d)^{\log\log d}) = \tilde{O}(1)$ from their "zero-temperature" locations ω_i. Hence, $\eta_j \in [a, b]$ for all $i - \mathfrak{b} \leq j \leq i + \mathfrak{b}$, where $a := \eta_{i-\mathfrak{b}} \geq \omega_i - \mathfrak{b}^2\sqrt{d}$ and $b := \eta_{i+\mathfrak{b}} \leq \omega_i + \mathfrak{b}^2\sqrt{d}$ w.h.p.

To apply this rigidity property, we define a new map $g : \mathcal{W}_d \to \mathcal{W}_d$ where $g(\eta)$ leaves all eigenvalues η_j outside $[a, b]$ fixed. And $g(\eta)$ expands the i'th eigengap by a factor of $\frac{1}{s}$: $g(\eta)[i] - g(\eta)[i+1] \geq \frac{1}{s}(\eta_i - \eta_{i+1})$. To "make room" for the expansion of the i'th gap without changing the locations of the eigenvalues outside $[a, b]$, it shrinks the eigengaps inside $[a, b]$ by a factor of $1 - \alpha$

where $\alpha := (\frac{1}{s}-1)\frac{\eta_i-\eta_{i+1}}{b-a} \leq \mathfrak{b}^{-3}$ whenever $\eta \in A(\frac{s}{\mathfrak{b}})$ because $\eta_i - \eta_{i+1} \leq s\frac{1}{\mathfrak{b}\sqrt{d}}$ if $\eta \in A(\frac{s}{\mathfrak{b}})$ (See (107)-(110) for the definition of g). Thus, roughly, for all $\eta \in A(\frac{s}{\mathfrak{b}})$,

$$\frac{f(g(\eta))}{f(\eta)} = \prod_{j\neq \ell} \frac{|g(\eta)[j]-g(\eta)[\ell]|^2}{|\eta_j-\eta_\ell|^2} e^{-\frac{1}{2}\sum_{j=i-\mathfrak{b}}^{i+\mathfrak{b}} \eta_j^2 - g(\eta)[j]^2} \geq \frac{1}{s^2}(1-\alpha)^{2\mathfrak{b}^2} e^{-\frac{1}{2}\sum_{j=i-\mathfrak{b}}^{i+\mathfrak{b}} \frac{1}{\mathfrak{b}}} \geq \frac{1}{s^2}. \quad (21)$$

The first inequality holds since the product has $O(\mathfrak{b}^2)$ "repulsion" terms $\frac{|g(\eta)[j]-g(\eta)[\ell]|^2}{|\eta_j-\eta_\ell|^2} \geq (1-\alpha)^2$ where $\ell, j \in [i-\mathfrak{b}, i+\mathfrak{b}]$ and one term $\frac{|g(\eta)[i]-g(\eta)[i+1]|^2}{|\eta_i-\eta_{i+1}|^2} \geq \frac{1}{s^2}$. Replacing g with ϕ and $A(s)$ with $A(\frac{s}{\mathfrak{b}})$ in (20), and plugging in (21) we get, roughly, that for all $s \geq 0$ and all $i \in [d]$,

$$\mathbb{P}\left(\eta_i - \eta_{i+1} \leq \frac{1}{\mathfrak{b}\sqrt{d}}s\right) = \int_{A(\frac{s}{\mathfrak{b}})} f(\eta)\mathrm{d}\eta \leq s^3 \int_{A(\frac{s}{\mathfrak{b}})} f(\eta) \times \frac{f(g(\eta))}{f(\eta)} \det(J_g(\eta))\mathrm{d}\eta \leq s^3. \quad (22)$$

This completes the proof overview of Lemma 4. To conclude, we note that one can convert the expectation bound in Theorem 2 into the following high-probability bound using the approach suggested in Appendix E of Mangoubi and Vishnoi (2022):

$$P(\|\hat{M}_k - M_k\|_F > s\sqrt{k}\sqrt{d}) \leq \tilde{O}(e^{-s})$$

for $s > 0$. However, here there is an additional difficulty since the eigenvalue gaps of Dyson Brownian motion only satisfy high probability lower bounds with polynomial decay, but not exponential decay. To overcome this, we believe one can use an approach similar to the proof of Lemma 4 to show that the joint density of the GUE eigenvalues (9) implies that whenever roughly $|i-j| > \tilde{\Omega}(1)$, the gaps $\eta_i - \eta_{i+1}$ and $\eta_j - \eta_{j+1}$ are (nearly) independent random variables in the sense that roughly

$$\mathbb{P}((\gamma_i(t)-\gamma_{i+1}(t) < x)\times(\gamma_j(t)-\gamma_{j+1}(t) < y)) \approx \mathbb{P}((\gamma_i(t)-\gamma_{i+1}(t) < x)\times\mathbb{P}(\gamma_j(t)-\gamma_{j+1}(t) < y)$$

for $x, y > 0$. It then follows that with a probability that decays exponentially in d, all but $\tilde{O}(1)$ of the GUE eigenvalue gaps satisfy $\eta_i - \eta_{i+1} \geq \tilde{\Omega}\left(\frac{1}{\sqrt{d}}\right)$.

4. Conclusion

We present and analyze a complex variant of the Gaussian mechanism for rank-k covariance matrix approximation under (ε, δ)-differential privacy. Our analysis leverages the fact that the eigenvalues of complex matrix Brownian motion repel more than in the real case, and uses Dyson's stochastic differential equations governing the evolution of its eigenvalues to show that, after any time $t > 0$, the eigenvalues of the matrix M perturbed by complex Gaussian noise have large gaps of size $\tilde{\Omega}\left(s\frac{\sqrt{t}}{\sqrt{d}}\right)$ with high probability $1-O(s^3)$. We believe the decay rate $1-O(s^3)$ in our eigenvalue gap bound is tight, as its derivative has the same exponent $\beta = 2$ which appears in the joint eigenvalue density formula (9) for the complex GUE random matrix.

We suspect our techniques can also be useful for other matrix approximation problems. For instance, one may consider the more general problem where one is given a covariance matrix M and a function $f : \mathbb{R}^{d\times d} \to \mathbb{R}^{d\times d}$ (e.g., the matrix exponential), and the goal is to find a rank-k symmetric matrix Y which minimizes $\|Y - f(M)\|_F$ under (ε, δ)-differential privacy. While the bounds presented in Theorem 1 are nearly tight, it may be possible to obtain stronger bounds when the input matrix M has additional structure; we leave this as a future direction.

Acknowledgments. OM was supported in part by an NSF CCF-2104528 award and a Google Research Scholar award. NV was supported in part by an NSF CCF-2112665 award.

References

Kareem Amin, Travis Dick, Alex Kulesza, Andres Munoz, and Sergei Vassilvitskii. Differentially private covariance estimation. *Advances in Neural Information Processing Systems*, 32, 2019.

Greg W Anderson, Alice Guionnet, and Ofer Zeitouni. *An introduction to random matrices*. Number 118. Cambridge university press, 2010.

Gérard Ben Arous and Paul Bourgade. Extreme gaps between eigenvalues of random matrices. *The Annals of Probability*, 41(4):2648–2681, 2013.

James Bennett and Stan Lanning. The Netflix Prize. In *Proceedings of KDD cup and workshop*, volume 2007, page 35. New York, NY, USA., 2007.

Rajendra Bhatia. *Matrix analysis*, volume 169. Springer Science & Business Media, 2013.

Jeremiah Blocki, Avrim Blum, Anupam Datta, and Or Sheffet. The Johnson-Lindenstrauss transform itself preserves differential privacy. In *2012 IEEE 53rd Annual Symposium on Foundations of Computer Science*, pages 410–419. IEEE, 2012.

Avrim Blum, Cynthia Dwork, Frank McSherry, and Kobbi Nissim. Practical privacy: the sulq framework. In *Proceedings of the twenty-fourth ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems*, pages 128–138, 2005.

Avrim Blum, John Hopcroft, and Ravindran Kannan. *Foundations of data science*. Cambridge University Press, 2020.

Paul Bourgade and H-T Yau. The eigenvector moment flow and local quantum unique ergodicity. *Communications in Mathematical Physics*, 350(1):231–278, 2017.

Kamalika Chaudhuri, Anand Sarwate, and Kaushik Sinha. Near-optimal differentially private principal components. *Advances in neural information processing systems*, 25:989–997, 2012.

Chandler Davis and William Morton Kahan. The rotation of eigenvectors by a perturbation. III. *SIAM Journal on Numerical Analysis*, 7(1):1–46, 1970.

Cynthia Dwork and Aaron Roth. The algorithmic foundations of differential privacy. *Foundations and Trends in Theoretical Computer Science*, 9(3-4):211–407, 2014.

Cynthia Dwork, Krishnaram Kenthapadi, Frank McSherry, Ilya Mironov, and Moni Naor. Our data, ourselves: Privacy via distributed noise generation. In *Annual International Conference on the Theory and Applications of Cryptographic Techniques*, pages 486–503. Springer, 2006a.

Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. Calibrating noise to sensitivity in private data analysis. In *Theory of cryptography conference*, pages 265–284. Springer, 2006b.

Cynthia Dwork, Kunal Talwar, Abhradeep Thakurta, and Li Zhang. Analyze Gauss: Optimal bounds for privacy-preserving principal component analysis. In *Proceedings of the forty-sixth annual ACM symposium on Theory of computing*, pages 11–20, 2014.

FJ Dyson and M Lal Mehta. Random matrices and the statistical theory of energy levels IV. *J. math. Phys*, 4:701–12, 1963.

Freeman J Dyson. A Brownian-motion model for the eigenvalues of a random matrix. *Journal of Mathematical Physics*, 3(6):1191–1198, 1962.

László Erdős, Benjamin Schlein, and Horng-Tzer Yau. Universality of random matrices and local relaxation flow. *Inventiones mathematicae*, 185(1):75–119, 2011.

László Erdős, Horng-Tzer Yau, and Jun Yin. Rigidity of eigenvalues of generalized Wigner matrices. *Advances in Mathematics*, 229(3):1435–1515, 2012.

Jean Ginibre. Statistical ensembles of complex, quaternion, and real matrices. *Journal of Mathematical Physics*, 6(3):440–449, 1965.

Vyacheslav L Girko. Circular law. *Theory of Probability & Its Applications*, 29(4):694–706, 1985.

Kiyokazu Inukai. Collision or non-collision problem for interacting Brownian particles. *Proceedings of the Japan Academy, Series A, Mathematical Sciences*, 82(4):66–70, 2006.

Kurt Johansson. Random matrices and determinantal processes. *arXiv preprint math-ph/0510038*, 2005.

Michael Kapralov and Kunal Talwar. On differentially private low rank approximation. In *Proceedings of the twenty-fourth annual ACM-SIAM symposium on Discrete algorithms*, pages 1395–1414. SIAM, 2013.

Benjamin Landon and Horng-Tzer Yau. Convergence of local statistics of Dyson Brownian motion. *Communications in Mathematical Physics*, 355(3):949–1000, 2017.

Gregory F Lawler. Stochastic calculus: An introduction with applications. *American Mathematical Society*, 2010.

Jonathan Leake, Colin S McSwiggen, and Nisheeth K Vishnoi. A polynomial-time algorithm and applications for matrix sampling from Harish-Chandra–Itzykson-Zuber densities. In *ACM symposium on theory of computing STOC*, 2021.

Ji Oon Lee, Kevin Schnelli, Ben Stetler, and Horng-Tzer Yau. Bulk universality for deformed Wigner matrices. *The Annals of Probability*, 44(3):2349–2425, 2016.

Oren Mangoubi and Nisheeth K. Vishnoi. Re-analyze Gauss: Bounds for private matrix approximation via Dyson Brownian motion. *NeurIPS*, 2022.

Oren Mangoubi, Yikai Wu, Satyen Kale, Abhradeep Thakurta, and Nisheeth K Vishnoi. Private matrix approximation and geometry of unitary orbits. In *Conference on Learning Theory*, pages 3547–3588. PMLR, 2022.

Peter Mörters and Yuval Peres. *Brownian motion*, volume 30. Cambridge University Press, 2010.

Hoi Nguyen, Terence Tao, and Van Vu. Random matrices: tail bounds for gaps between eigenvalues. *Probability Theory and Related Fields*, 167(3):777–816, 2017.

Sean O'Rourke, Van Vu, and Ke Wang. Random perturbation of low rank matrices: Improving classical bounds. *Linear Algebra and its Applications*, 540:26–59, 2018.

Tharmalingam Ratnarajah, Rémi Vaillancourt, and M Alvo. Eigenvalues and condition numbers of complex random matrices. *SIAM Journal on Matrix Analysis and Applications*, 26(2):441–456, 2004.

Irene Rodriguez-Lujan, Jordi Fonollosa, Alexander Vergara, Margie Homer, and Ramon Huerta. On the calibration of sensor arrays for pattern recognition using the minimal number of experiments. *Chemometrics and Intelligent Laboratory Systems*, 130:123–134, 2014.

Leonard CG Rogers and Zhan Shi. Interacting Brownian particles and the Wigner law. *Probability theory and related fields*, 95(4):555–570, 1993.

Or Sheffet. Old techniques in differentially private linear regression. In *Algorithmic Learning Theory*, pages 789–827. PMLR, 2019.

Terence Tao. *Topics in random matrix theory*, volume 132. American Mathematical Soc., 2012.

Terence Tao. The asymptotic distribution of a single eigenvalue gap of a Wigner matrix. *Probability Theory and Related Fields*, 157(1):81–106, 2013.

Jalaj Upadhyay. The price of privacy for low-rank factorization. *Advances in Neural Information Processing Systems*, 31, 2018.

Roman Vershynin. *High-dimensional probability: An introduction with applications in data science*, volume 47. Cambridge university press, 2018.

Contents

Appendix A. Preliminaries

A.1. Brownian motion and Itô calculus

In this section, we give preliminaries on Brownian motion and Stochastic calculus (also referred to as Itô calculus). A Brownian motion $W(t)$ is a continuous process that has stationary independent increments (see e.g., (Mörters and Peres, 2010)). In a multi-dimensional Brownian motion, each coordinate is an independent and identical Brownian motion. The filtration $\mathcal{F}_t$ generated by $W(t)$ is defined as $\sigma\left(\cup_{s\le t}\sigma(W(s))\right)$, where $\sigma(\Omega)$ is the σ-algebra generated by Ω. $W(t)$ is a martingale with respect to $\mathcal{F}_t$.

Definition 5 (Itô Integral) *Let $W(t)$ be a Brownian motion for $t \geq 0$, let $\mathcal{F}_t$ be the filtration generated by $W(t)$, and let $z(t) : \mathcal{F}_t \to \mathbb{R}$ be a stochastic process adapted to $\mathcal{F}_t$. The Itô integral is defined as*

$$\int_0^T z(t)\mathrm{d}W(t) := \lim_{\omega\to 0}\sum_{i=1}^{\frac{T}{\omega}} z(i\omega) \times [W((i+1)\omega) - W(i\omega)].$$

The following lemma (Itô's Lemma) generalizes the chain rule of deterministic derivatives to stochastic derivatives and allows one to compute the derivative of a function $f(X(t))$ of a stochastic process $X(t)$. We state Itô's Lemma in its integral form:

Lemma 6 (Itô's Lemma, integral form with no drift; Theorem 3.7.1 of Lawler (2010)) *Let $f : \mathbb{R}^n \to \mathbb{R}$ be any twice-differentiable function. Let $W(t) \in \mathbb{R}^n$ be a Brownian motion, and let $X(t) \in \mathbb{R}^n$ be an Itô diffusion process with mean zero defined by the following stochastic differential equation:*

$$\mathrm{d}X_j(t) = \sum_{i=1}^{d} R_{ij}(t)\mathrm{d}W_i(t), \tag{23}$$

for some Itô diffusion $R(t) \in \mathbb{R}^{n\times n}$ adapted to the filtration generated by the Brownian motion $W(t)$. Then for any $T \geq 0$,

$$f(X(T)) - f(X(0)) = \int_0^T \sum_{i=1}^{n}\sum_{\ell=1}^{n}\left(\frac{\partial}{\partial X_\ell} f(X(t))\right) R_{i\ell}(t)\mathrm{d}W_i(t) + \frac{1}{2}\int_0^T \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{\ell=1}^{n}\left(\frac{\partial^2}{\partial X_j \partial X_\ell} f(X(t))\right) R_{ij}(t)R_{i\ell}(t)\mathrm{d}t.$$

We note that the above version of Itô's Lemma (Lemma 6) is given for real-valued variables. We apply Itô's Lemma to complex matrix-valued stochastic processes, we will separate the real and imaginary parts of the Itô integral and apply Itô's Lemma separately to each part.

A.2. Dyson Brownian motion

In this section, we give additional results about the existence, continuity, and related properties of Dyson Brownian motion. Let $\mathcal{O}(d)$ denote the space of $d \times d$ real orthogonal matrices, and $\mathcal{U}(d)$ the space of $d \times d$ complex unitary matrices. The following lemma, which guarantees the existence and uniqueness of solutions to the eigenvalue (4) and eigenvector SDE's (5), is known – see Theorem 2.3(a) in Bourgade and Yau (2017) and Lemma 4.3.3 in Anderson et al. (2010) for solutions of just the eigenvalue process for any $\beta \geq 1$. While the solutions are random processes, the outcome of these solutions can be shown to be unique when coupled to the underlying Brownian motion processes driving the SDE. Such a coupling is referred to as a "strong solution" to the SDE (see e.g. Lawler (2010)).

Lemma 7 (Existence and uniqueness of solutions to Dyson Brownian motion) *Consider any $T \geq T_0 \geq 0$ and $\beta \in \{1, 2\}$. Let $\{\gamma(t)\}_{t \in [0,T_0]} \subseteq \mathcal{W}_d$ be a continuous initial path for (4) and let $\{u(t)\}_{t\in[0,T_0]} \subseteq \mathcal{U}(d)$ if $\beta = 2$ (or $\{u(t)\}_{t\in[0,T_0]} \subseteq \mathcal{O}(d)$ if $\beta = 1$) be a continuous initial path for (5). Then there exists a unique strong solution for the system of SDEs (4) on all of $[0, T]$. Moreover, there exists a unique strong solution on all of $[0, T]$ for the system of SDEs comprising (4) and (5).*

In particular (by the definition of strong solution) the existence of strong solutions implies that the paths of Dyson Brownian motion are almost surely continuous on $[0, \infty)$. This fact will be useful in proving our gap comparison theorem for coupled solutions of Dyson Brownian motions (Lemma 3). The following result shows that the paths of Dyson Brownian motion are continuous with respect to their initial conditions:

Lemma 8 (Continuity w.r.t. initial condition; Proposition 4.3.5 in Anderson et al. (2010)) *Let γ be a strong solution to (4) for any initial condition $\gamma(0) \in \mathcal{W}_d$. Then, at any time $t \geq 0$, $\gamma(t)$ is a continuous function of the initial condition $\gamma(0)$.*

The following lemma is known; see Theorem 1.1 in Inukai (2006) and also Rogers and Shi (1993).

Lemma 9 (Non-collision of Dyson Brownian motion for $\beta \geq 1$) *Let γ be a solution to (4) with any initial condition $\gamma(0) \in \mathcal{W}_d$. Let $\tau := \inf\{t > 0 : \gamma_i(t) = \gamma_j(t)$ for some $i \neq j\}$ be the first positive time any of the particles in $\gamma(t)$ collide. Then if $\beta \geq 1$, $\mathbb{P}(\tau < \infty) = 0$.*

A.3. Matrix inequalities

The following lemmas will help us bound the gaps in the eigenvalues of the Dyson Brownian motion:

Lemma 10 (Theorem 4.4.5 of Vershynin (2018), special case [1]) *Let $W \in \mathbb{R}^{d\times d}$ with i.i.d. $N(0, 1)$ entries. Then*

$$\mathbb{P}(\|W\|_2 > 2\sqrt{d} + s) < 2e^{-s^2}$$

for any $s > 0$.

Note that Lemma 10 also applies to complex Gaussian matrices $W_1 + \mathfrak{i}W_2$ where W_1, W_2 have i.i.d. real $N(0, 1)$ entries, since $\|W_1 + \mathfrak{i}W_2\|_2 \geq \|W_1\|_2$.

Lemma 11 (Weyl's Inequality (Bhatia, 2013)) *If $A, B \in \mathbb{C}^{d\times d}$ are two Hermitian matrices, and denoting the i'th-largest eigenvalue of any Hermitian matrix M by $\sigma_i(M)$, we have*

$$\sigma_i(A) + \sigma_d(B) \leq \sigma_i(A + B) \leq \sigma_i(A) + \sigma_1(B).$$

Lemma 12 (Spectral norm bound (Theorem 4.3 of Mangoubi and Vishnoi (2022))) *For some universal constant C, and every $T > 0$, we have,*

$$\mathbb{P}\left(\sup_{t\in[0,T]} \|B(t)\|_2 > T\sqrt{d} + \alpha)\right) \leq e^{-C\frac{\alpha^2}{T^2}}.$$

A.4. Davis-Kahan Sin-Theta theorem

The following lemma gives a deterministic bound on the change to the subspace spanned by the top-k eigenvectors of a Hermitian matrix when it is perturbed by the addition of another Hermitian matrix. Let A and $\hat{A}$ be two Hermitian matrices with eigenvalue decompositions

$$A = U\Lambda U^* = (U_1, U_2) \begin{pmatrix} \Lambda_1 & \\ & \Lambda_2 \end{pmatrix} \begin{pmatrix} U_1^* \\ U_2^* \end{pmatrix} \tag{24}$$

$$\hat{A} = \hat{U}\hat{\Lambda}\hat{U}^* = (\hat{U}_1, \hat{U}_2) \begin{pmatrix} \hat{\Lambda}_1 & \\ & \hat{\Lambda}_2 \end{pmatrix} \begin{pmatrix} \hat{U}_1^* \\ \hat{U}_2^* \end{pmatrix}, \tag{25}$$

(although when we apply the Sin-Theta theorem we will only need the special case where $\hat{\Lambda} = \Lambda$).

Lemma 13 (sin-Θ Theorem (Davis and Kahan, 1970)) *Let $A, \hat{A}$ be two Hermitian matrices with eigenvalue decompositions given in (24) and (25). Suppose that there are $\alpha > \beta > 0$ and $\Delta > 0$ such that the spectrum of Λ_1 is contained in the interval $[\alpha, \beta]$ and the spectrum of $\hat{\Lambda}_2$ lies entirely outside of the interval $(\alpha - \Delta, \beta + \Delta)$. Then*

$$|||U_1U_1^* - \hat{U}_1\hat{U}_1^*||| \leq \frac{|||\hat{A} - A|||}{\Delta},$$

where $|||\cdot|||$ denotes the operator norm or Frobenius norm (or, more generally, any unitarily invariant norm).

Appendix B. Differentially private rank-k approximation: Proof of Theorem 1

Proof [Proof of Theorem 1]

Privacy. The real Gaussian mechanism, $M + \sqrt{T}(W_1 + W_1^\top)$, where W_1 is a matrix with i.i.d. $N(0, 1)$ entries, was studied in Dwork et al. (2014) and shown to be (ε, δ)-differentially private for $T = \frac{2\log\frac{1.25}{\delta}}{\varepsilon^2}$. Our Algorithm 1 is (ε, δ)-differentially private since it is a post-processing of the real Gaussian mechanism. This is because any post-processing of an (ε, δ)-differentially private mechanism (which does not have access to the original input matrix M) is guaranteed to be (ε, δ)-differentially private (see e.g. Dwork et al. (2006a), Dwork and Roth (2014)). To see why Algorithm 1 is a post-processing of the real Gaussian mechanism, observe that

$$\begin{aligned}\hat{M} &= M + G \\ &= M + W + W^* \\ &= M + (W_1 + W_2\mathfrak{i}) + (W_1 + W_2\mathfrak{i})^* \\ &= M + W_1 + W_1^\top + [W_2\mathfrak{i} + (W_2\mathfrak{i})^*].\end{aligned}$$

Utility of complex matrix $\hat{M}_k$ implies Utility of real matrix Y. Let $M = V\Sigma V^\top$ be a diagonalization of the real symmetric input matrix M with eigenvalues $\sigma_1 \geq \cdots \geq \sigma_d \geq 0$. Let $M_k = V\Sigma_k V^\top$ be a (non-private) rank-k approximation of M, where $\Sigma_k = \mathrm{diag}(\sigma_1, \ldots, \sigma_k, 0, \ldots, 0)$.

Suppose we can show an upper bound on $\|\hat{M}_k - M_k\|_F$, where $\hat{M}_k$ is the complex matrix in Algorithm 1.

Let $\mathfrak{R}_k := \{A \in \mathbb{R}^{d\times d} : \mathrm{rank}(A) \leq k\}$ denote the set of real $d \times d$ rank-k matrices. Since $Y = \mathrm{Real}(\hat{V}\hat{\Sigma}_k\hat{V}^*)$, we have that $Y \in \mathrm{argmin}_{Z\in\mathfrak{R}_k}\{\|\hat{M}_k - Z\|_F\}$. This is because $\mathrm{Real}(\hat{V}\hat{\Sigma}_k\hat{V}^*)$ is a matrix of rank at most k and the real and imaginary parts of $\hat{V}\hat{\Sigma}_k\hat{V}^*$ are orthogonal to each other in the Frobenius inner product. Thus, since $Y \in \mathrm{argmin}_{Z\in\mathfrak{R}_k}\{\|\hat{M}_k - Z\|_F\}$ and $M_k \in \mathfrak{R}_k$ is also in the set of real-valued rank-k matrices, we have that

$$\|\hat{M}_k - Y\|_F \leq \|\hat{M}_k - M_k\|_F.$$

Therefore, we have

$$\|Y - M_k\|_F \leq \|\hat{M}_k - Y\|_F + \|\hat{M}_k - M_k\|_F \leq 2\|\hat{M}_k - M_k\|_F. \tag{26}$$

Plugging in our bound for $\sqrt{\mathbb{E}[\|\hat{M}_k - M_k\|_F^2]}$ from Theorem 2 into (26), we get that

$$\sqrt{\mathbb{E}[\|M_k - Y\|_F^2]} \leq \tilde{O}\left(\sqrt{kd}\frac{\sigma_k}{\sigma_k - \sigma_{k+1}} \times \frac{\sqrt{\log(\frac{1}{\delta})}}{\varepsilon}\right).$$

■

Appendix C. Complex Gaussian perturbations: Proof of Theorem 2

C.1. Defining the stochastic process on the space of rank-k matrices

At every time t, let $\Phi(t) = U(t)\Gamma(t)U(t)^*$ be a spectral decomposition of the symmetric matrix $\Phi(t)$, where $\Gamma(t)$ is a diagonal matrix with diagonal entries $\gamma_1(t) \geq \cdots \geq \gamma_d(t)$ that are the eigenvalues of $\Phi(t)$, and $U(t) = [u_1(t), \ldots, u_d(t)]$ is a $d \times d$ unitary matrix whose columns $u_1(t), \ldots, u_d(t)$ are an orthonormal basis of eigenvectors of $\Phi(t)$. At every time t, define $\Psi(t)$ to be the symmetric matrix with any eigenvalues $\lambda_1(t) \geq \cdots \lambda_d(t)$, where $\Lambda(t) := \mathrm{diag}(\lambda_1(t), \ldots, \lambda_d(t))$, and with eigenvectors given by the columns of $U(t)$:

$$\Psi(t) := U(t)\Lambda(t)U(t)^* \qquad \forall t \in [0, T].$$

For all t, will fix $\lambda_i(t) = \gamma_i(t)$ for $i \leq k$ and $\lambda_i(t) = 0$ for all $i > k$.

C.2. Preliminary results

For every $\alpha > 0$, let $\hat{E}_\alpha$ be the "bad" event that either $\sup_{t\in[0,T]} \|B(t)\|_2 > 4\sqrt{T}(\sqrt{d} + \alpha)$ or $\sup_{t\in[0,t_0]} \|B(t)\|_2 > 4\sqrt{t_0}(\sqrt{d} + \alpha)$, or $\inf_{t_0\leq t\leq T, 1\leq i<d} \gamma_i(t) - \gamma_{i+1}(t) \leq \frac{1}{d^{10}}\frac{\sqrt{t}}{\mathfrak{b}\sqrt{d}}$. In the following, we set $\alpha = 20\log^{\frac{1}{2}}(\sigma_1 + T)$ and $t_0 = \frac{1}{(kd)^{10} + k\alpha^2 + \sigma_1^2}$. The following Lemma shows that $\hat{E}_\alpha$ occurs with very low probability:

Lemma 14 (Probability of "bad" event occurring) *For every $T > 0$ and every $\alpha > 0$, we have,* $\mathbb{P}\left(\hat{E}_\alpha\right) \leq 4\sqrt{\pi}e^{-\frac{1}{8}\alpha^2} + \frac{T}{d^{600}}$.

Proof

$$\begin{aligned}\mathbb{P}\left(\hat{E}_\alpha\right) &\leq \mathbb{P}\left(\sup_{t\in[0,T]}\|B(t)\|_2 > 4\sqrt{T}(\sqrt{d}+\alpha)\right) + \mathbb{P}\left(\sup_{t\in[0,t_0]}\|B(t)\|_2 > 4\sqrt{t_0}(\sqrt{d}+\alpha)\right)\\ &\quad + \mathbb{P}\left(\inf_{t_0\leq t\leq T, 1\leq i<d}\gamma_i(t)-\gamma_{i+1}(t) \leq \frac{1}{d^{10}}\frac{\sqrt{t}}{\mathfrak{b}\sqrt{d}}\right)\\ &\overset{\text{Lemmas 22, 12}}{\leq} 4\sqrt{\pi}e^{-\frac{1}{8}\alpha^2} + \frac{T}{d^{600}}.\end{aligned}$$

■

Lemma 15 *If $\alpha \geq 20\log^{\frac{1}{2}}(\sigma_1+T)$, then we have*

$$\mathbb{E}[\|\Psi(T)-\Psi(0)\|_F^2] \leq 4\mathbb{E}[\|\Psi(T)-\Psi(0)\|_F^2 \times \mathbb{1}_{\hat{E}_\alpha^c}] + d.$$

Proof

$$\mathbb{E}[\|\Psi(T)-\Psi(0)\|_F^2] \leq 4\mathbb{E}[\|\Psi(T)-\Psi(0)\|_F^2 \times \mathbb{1}_{\hat{E}_\alpha^c}] + 4\mathbb{E}[\|\Psi(T)-\Psi(0)\|_F^2 \times \mathbb{1}_{\hat{E}_\alpha}]. \tag{27}$$

$$\begin{aligned}\|\Psi(T)-\Psi(0)\|_F &= \|U(T)\Gamma_k(T)U(T)^* - U(0)\Gamma_k(0)U(0)^*\|_F\\ &\leq \|U(T)\Gamma_k(T)U(T)^*\|_F + \|U(0)\Gamma_k(0)U(0)^*\|_F\\ &\leq \|U(T)\Gamma(T)U(T)^*\|_F + \|U(0)\Gamma(0)U(0)^*\|_F\\ &= \|\Phi(T)\|_F + \|\Phi(0)\|_F\\ &= \|M+B(T)\|_F + \|M\|_F\\ &\leq 2\|M\|_F + \|B(T)\|_F.\end{aligned}$$

Therefore,

$$\begin{aligned}
\mathbb{E}[\|\Psi(T)-\Psi(0)\|_F^2 \times \mathbb{1}_{\hat{E}_\alpha}] &\leq \mathbb{E}[(2\|M\|_F + \|B(T)\|_F)^2 \times \mathbb{1}_{\hat{E}_\alpha}] \\
&\leq \mathbb{E}[(16\|M\|_F^2 + 4\|B(T)\|_F^2) \times \mathbb{1}_{\hat{E}_\alpha}] \\
&= 16\|M\|_F^2 \times \mathbb{P}(\hat{E}_\alpha) + 4\mathbb{E}[\|B(T)\|_F^2 \times \mathbb{1}_{\hat{E}_\alpha}] \\
&\leq 16\|M\|_F^2 \times \mathbb{P}(\hat{E}_\alpha) + 4\sqrt{d}\mathbb{E}[\|B(T)\|_2^2 \times \mathbb{1}_{\hat{E}_\alpha}] \\
&= 16\|M\|_F^2 \times \mathbb{P}(\hat{E}_\alpha) + 4\sqrt{d}\int_{16T(\sqrt{d}+\alpha)^2}^{\infty} \mathbb{P}(\|B(T)\|_2^2 > s)\mathrm{d}s \\
&= 16\|M\|_F^2 \times \mathbb{P}(\hat{E}_\alpha) + 4\sqrt{d}\int_{4T(\sqrt{d}+\alpha)^2}^{\infty} \mathbb{P}(T\|W\|_2^2 > s)\mathrm{d}s \\
&= 16\|M\|_F^2 \times \mathbb{P}(\hat{E}_\alpha) + 4\sqrt{d}\int_{16T(\sqrt{d}+\alpha)^2}^{\infty} \mathbb{P}\left(\|W\|_2 > \frac{\sqrt{s}}{\sqrt{T}}\right)\mathrm{d}s \\
&= 16\|M\|_F^2 \times \mathbb{P}(\hat{E}_\alpha) + 4\sqrt{d}\int_{16T(\sqrt{d}+\alpha)^2}^{\infty} 2e^{-\left(\frac{\sqrt{s}}{\sqrt{T}}-2\sqrt{d}\right)^2}\mathrm{d}s \\
&= 16\|M\|_F^2 \times \mathbb{P}(\hat{E}_\alpha) + 4\sqrt{d}\int_{16T(\sqrt{d}+\alpha)^2}^{\infty} 2e^{-\left(\frac{s}{T}-2\sqrt{d}\frac{\sqrt{s}}{\sqrt{T}}+4d\right)}\mathrm{d}s \\
&\leq 16\|M\|_F^2 \times \mathbb{P}(\hat{E}_\alpha) + 4\sqrt{d}\int_{16T(\sqrt{d}+\alpha)^2}^{\infty} 2e^{-\left(\frac{s}{2T}+4d\right)}\mathrm{d}s \\
&= 16\|M\|_F^2 \times \mathbb{P}(\hat{E}_\alpha) + 4\sqrt{d}e^{-4d}\int_{16T(\sqrt{d}+\alpha)^2}^{\infty} 2e^{-\frac{s}{2T}}\mathrm{d}s \\
&= 16\|M\|_F^2 \times \mathbb{P}(\hat{E}_\alpha) + 4\sqrt{d}e^{-4d}4Te^{-\frac{16T(\sqrt{d}+\alpha)^2}{2T}} \\
&= 16\|M\|_F^2 \times \mathbb{P}(\hat{E}_\alpha) + 4\sqrt{d}e^{-4d}4Te^{-8(\sqrt{d}+\alpha)^2} \\
&\leq 16\|M\|_F^2 \times \mathbb{P}(\hat{E}_\alpha) + 1 \\
&\overset{\text{Lemma 14}}{\leq} 16\|M\|_F^2 \times \left(4\sqrt{\pi}e^{-\frac{1}{8}\alpha^2} + \frac{T}{d^{600}}\right) + 1 \\
&\leq 16d\sigma_1^2 \times \left(4\sqrt{\pi}e^{-\frac{1}{8}\alpha^2} + \frac{T}{d^{600}}\right) + 1 \\
&\leq \frac{1}{4}d + \frac{T}{d^{200}}, \qquad (28)
\end{aligned}$$

where W is a matrix with i.i.d. $N(0,1)$ entries, and $Y \sim (0, \frac{1}{2})$. The last inequality is because $\alpha \geq 20\log^{\frac{1}{2}}(d\sigma_1 + T)$ and $\sigma_1^2 \leq d^{100}$. Plugging in (28) into (27) completes the proof. ■

The following lemma will be useful in bounding the gaps $\gamma_i(t) - \gamma_j(t)$ for $i \leq k < j$.

Lemma 16 ("Worst-case" eigenvalue gap bound) *Whenever $\gamma_i(0) - \gamma_{i+1}(0) \geq 8\sqrt{T}\sqrt{d}$ for every $i \in S$ and $T > 0$ and some subset $S \subset [d-1]$, we have that for any $\alpha > 0$,*

$$\bigcup_{i\in S}\left\{\inf_{t\in[0,T]} \gamma_i(t) - \gamma_{i+1}(t) < \frac{1}{2}(\gamma_i(0) - \gamma_{i+1}(0)) - \alpha\right\} \subseteq \hat{E}_\alpha.$$

Proof [Proof of Lemma 16] Since, at every time t, $\Phi(t) = M + B(t)$ and $\gamma_1(t) \geq \cdots \geq \gamma_d(t)$ are the eigenvalues of $\Phi(t)$, Weyl's Inequality implies that

$$\gamma_i(t) - \gamma_{i+1}(t) \geq \gamma_i(0) - \gamma_{i+1}(0) - \|B(t)\|_2, \qquad \forall t \in [0,T], i \in [d]. \tag{29}$$

Therefore, (29) implies that

$$\begin{aligned}
&\bigcup_{i\in S} \left\{ \inf_{t\in[0,T]} \gamma_i(t) - \gamma_{i+1}(t) < \frac{1}{2}(\gamma_i(0) - \gamma_{i+1}(0)) - \alpha) \right\} \\
&\overset{\text{Eq. (29)}}{\subseteq} \bigcup_{i\in S} \left\{ \gamma_i(0) - \gamma_{i+1}(0) - \sup_{t\in[0,T]} 2\|B(t)\|_2 < \frac{1}{2}(\gamma_i(0) - \gamma_{i+1}(0)) - \alpha) \right\} \\
&= \bigcup_{i\in S} \left\{ \sup_{t\in[0,T]} \|B(t)\|_2 > \frac{1}{4}(\gamma_i(0) - \gamma_{i+1}(0)) + \frac{1}{2}\alpha) \right\} \\
&\subseteq \bigcup_{i\in S} \left\{ \sup_{t\in[0,T]} \|B(t)\|_2 > 2\sqrt{T}\sqrt{d} + \frac{1}{2}\alpha) \right\} \\
&= \left\{ \sup_{t\in[0,T]} \|B(t)\|_2 > 2\sqrt{T}\sqrt{d} + \frac{1}{2}\alpha \right\} \\
&= \hat{E}_\alpha.
\end{aligned}$$

The first inequality holds by (29), and the second inequality holds since the statement of Lemma 16 assumes that $\gamma_i(0) - \gamma_{i+1}(0) \geq 8\sqrt{T}\sqrt{d}$. ■

The following proposition provides a crude bound on the Frobenius distance over the very short time interval $[0, t_0]$, which we will use to "jump-start" our more sophisticated bound on the much longer interval $[t_0, T]$:

Proposition 17 *Suppose that* $\sigma_k - \sigma_{k+1} \geq 4T\sqrt{d} + 2\alpha$. *Then for every* $0 \leq t_0 < 1$ *we have*

$$\|\Psi(t_0) - \Psi(0)\|_F \times \mathbb{1}_{\hat{E}^c_\alpha} \leq \sqrt{t_0}\left(2\sqrt{k}(\sqrt{d} + \alpha) + 8\sigma_1\right)$$

with probability 1.

Proof At every time $t \geq 0$, let $U_k(t)$ denote the $d \times k$ matrix consisting of the first k columns of $U(t)$. Further, let $\Gamma_k(t)$ denote the $k \times k$ matrix consisting of the first k rows and columns of $\Gamma(t)$.

$$\begin{aligned}
\|\Psi(t_0) - \Psi(0)\|_F &= \|U_k(t_0)\Gamma_k(t_0)U_k(t_0)^* - U_k(0)\Gamma_k(0)U_k(0)^*\|_F \\
&\leq \|U_k(t_0)\Gamma_k(t_0)U_k(t_0)^* - U_k(t_0)\Gamma_k(0)U_k(t_0)^*\|_F \\
&+ \|U_k(t_0)\Gamma_k(0)U_k(t_0)^* - U_k(t_0)\Gamma_k(0)U_k(0)^*\|_F \\
&+ \|U_k(t_0)\Gamma_k(0)U_k(0)^* - U_k(0)\Gamma_k(0)U_k(0)^*\|_F \\
&\leq \|U_k(t_0)\|_2^2 \times \|\Gamma_k(t_0) - \Gamma_k(0)\|_F \\
&+ \|U_k(t_0)\|_2 \times \|\Gamma_k(0)\|_2 \times \|U_k(t_0)^* - U_k(0)^*\|_F \\
&+ \|U_k(t_0) - U_k(0)\|_F \times \|\Gamma_k(0)\|_2 \times \|U_k(0)^*\|_2 \\
&= \|\Gamma_k(t_0) - \Gamma_k(0)\|_F + \sigma_1\|U_k(t_0)^* - U_k(0)^*\|_F + \sigma_1\|U_k(t_0) - U_k(0)\|_F \\
&= \|\Gamma_k(t_0) - \Gamma_k(0)\|_F + 2\sigma_1\|U_k(t_0) - U_k(0)\|_F,
\end{aligned} \tag{30}$$

where second equality holds since $\|U_k(t)\|_2 = 1$ for all $t \geq 0$, and since $\|\Gamma_k(0)\|_2 = \sigma_1$ since $\Gamma_k(0) = M$.

By Lemma 13, we have

$$\|U_k(t_0)U_k^*(t_0) - U_k(0)U_k^*(0)\|_F \overset{\text{Lemma 13}}{\leq} \frac{\|\Phi(t_0) - \Phi(0)\|_F}{\gamma_k(0) - \gamma_{k+1}(t_0)} = \frac{\|B(t_0)\|_F}{\gamma_k(0) - \gamma_{k+1}(t_0)}. \tag{31}$$

By Weyl's Inequality (Lemma 11), we have that, whenever the event $\hat{E}_\alpha^c$ occurs,

$$\begin{aligned} \gamma_{k+1}(t_0) &\overset{\text{Lemma 11}}{\leq} \gamma_{k+1}(0) + \|B(t_0)\|_2 \\ &\leq \gamma_{k+1}(0) + 2\sqrt{t_0}(\sqrt{d} + \alpha) \\ &= \sigma_{k+1} + 2\sqrt{t_0}(\sqrt{d} + \alpha) \end{aligned} \tag{32}$$

where the second inequality is by the definition of the event $\hat{E}_\alpha^c$. Thus, (32) implies that

$$\begin{aligned} \gamma_k(0) - \gamma_{k+1}(t_0) &\overset{\text{Eq. 32}}{\geq} \gamma_k(0) - \sigma_{k+1} - 2\sqrt{t_0}(\sqrt{d} + \alpha) \\ &= \sigma_k - \sigma_{k+1} - 2\sqrt{t_0}(\sqrt{d} + \alpha) \\ &\geq \frac{1}{2}(\sigma_k - \sigma_{k+1}). \end{aligned} \tag{33}$$

where the second inequality holds because $\sigma_k - \sigma_{k+1} \geq 4T\sqrt{d} + 2\alpha$ and $T \geq 1 > t_0$. Thus, plugging (33) into (31), we have that whenever the event $\hat{E}_\alpha^c$ occurs,

$$\|U_k(t_0)U_k^*(t_0) - U_k(0)U_k^*(0)\|_F \leq \frac{2\|B(t_0)\|_F}{\sigma_k - \sigma_{k+1}} \tag{34}$$

$$\leq \frac{4\sqrt{t_0}(\sqrt{d} + \alpha)}{\sigma_k - \sigma_{k+1}}, \tag{35}$$

where the second inequality is by the definition of the event $\hat{E}_\alpha^c$. We also have (by, e.g., Inequality (27) in Mangoubi et al. (2022)) that

$$\|U_k(t_0) - U_k(0)\|_F \leq \|U_k(t_0)U_k^*(t_0) - U_k(0)U_k^*(0)\|_F. \tag{36}$$

Therefore, plugging in (36) into (34), we get that, whenever the event $\hat{E}_\alpha^c$ occurs,

$$\|U_k(t_0) - U_k(0)\|_F \leq \frac{4\sqrt{t_0}(\sqrt{d} + \alpha)}{\sigma_k - \sigma_{k+1}}. \tag{37}$$

Plugging in (37) into (30) we get

$$\begin{aligned}
\|\Psi(t_0)-\Psi(0)\|_F \times \mathbb{1}_{\hat{E}^c_\alpha} &\overset{\text{Eq. (30)}}{\leq} \|\Gamma_k(t_0)-\Gamma_k(0)\|_F \times \mathbb{1}_{\hat{E}^c_\alpha} + 2\sigma_1\|U_k(t_0)-U_k(0)\|_F \times \mathbb{1}_{\hat{E}^c_\alpha} \\
&\overset{\text{Eq. (37)}}{\leq} \|\Gamma_k(t_0)-\Gamma_k(0)\|_F \times \mathbb{1}_{\hat{E}^c_\alpha} + 2\sigma_1\frac{4\sqrt{t_0}(\sqrt{d}+\alpha)}{\sigma_k-\sigma_{k+1}} \\
&\leq \sqrt{k}\times\|B(t_0)\|_2 \times \mathbb{1}_{\hat{E}^c_\alpha} + 2\sigma_1\frac{4\sqrt{t_0}(\sqrt{d}+\alpha)}{\sigma_k-\sigma_{k+1}} \\
&\leq \sqrt{k}\times 2\sqrt{t_0}(\sqrt{d}+\alpha) + 2\sigma_1\frac{4\sqrt{t_0}(\sqrt{d}+\alpha)}{\sigma_k-\sigma_{k+1}} \\
&\leq \sqrt{k}\times 2\sqrt{t_0}(\sqrt{d}+\alpha) + 2\sigma_1\frac{4\sqrt{t_0}(\sqrt{d}+\alpha)}{\sigma_k-\sigma_{k+1}} \\
&\leq \sqrt{k}\times 2\sqrt{t_0}(\sqrt{d}+\alpha) + 2\sigma_1\frac{4\sqrt{t_0}(\sqrt{d}+\alpha)}{\sqrt{d}+\alpha} \\
&\leq 2\sqrt{t_0}\sqrt{k}(\sqrt{d}+\alpha) + 8\sigma_1\sqrt{t_0} \\
&= \sqrt{t_0}(2\sqrt{k}(\sqrt{d}+\alpha)+8\sigma_1), \qquad (38)
\end{aligned}$$

where the fourth inequality is by the definition of the event $\hat{E}^c_\alpha$, and the fifth inequality holds by our assumption that $\sigma_k - \sigma_{k+1} \geq 4T\sqrt{d} + 2\alpha$ and since $T \geq 1$. ■

C.3. Proof of Theorem 2

Lemma 18 (Itô derivative $\mathrm{d}u_i(t)u_j^*(t)$) *For all $t \in [0, T]$,*

$$\begin{aligned}
&\mathrm{d}(u_i(t)u_i^*(t)) \\
&= \sum_{j\neq i}\frac{1}{\gamma_i(t)-\gamma_j(t)}(u_i(t)u_j^*(t)\mathrm{d}B_{ij}(t) + u_j(t)u_i^*(t)\mathrm{d}B_{ij}^*(t)) \\
&\quad - \sum_{j\neq i}\frac{\mathrm{d}t}{(\gamma_i(t)-\gamma_j(t))^2}(u_i(t)u_i^*(t) - u_j(t)u_j^*(t)).
\end{aligned}$$

The proof of Lemma 18 is given in Section F.2 and is an adaptation of the Proof of Lemma 4.1 of Mangoubi and Vishnoi (2022) to the setting of complex matrices.

Define $\Delta_{ij}(t) := \gamma_i(t) - \gamma_j(t)$ for all $i, j \in [d]$ and all $t \geq 0$. Fix any $\eta \in \mathbb{R}^{d\times d}$, define $\mu_{ij}(t) := \max(\Delta_{ij}(t), \eta_{ij})$ for $i \leq j$ and $\mu_{ij}(t) := -\mu_{ji}(t)$ for $i > j$. Further, define the following matrix-valued Itô diffusion $Z_\eta(t)$ via its Itô derivative $\mathrm{d}Z_\eta(t)$:

$$\begin{aligned}
\mathrm{d}Z_\eta(t) \quad := \quad & \frac{1}{2}\sum_{i=1}^d\sum_{j\neq i}(\lambda_i(t)-\lambda_j(t))\frac{1}{\mu_{ij}(t)}(u_i(t)u_j^*(t)\mathrm{d}B_{ij}(t) + u_j(t)u_i^*(t)\mathrm{d}B_{ij}^*(t)) \\
&+\sum_{i=1}^d\sum_{j\neq i}(\lambda_i(t)-\lambda_j(t))\frac{\mathrm{d}t}{\mu_{ij}^2(t)}u_i(t)u_i^*(t), \qquad (39)
\end{aligned}$$

with initial condition $Z_\eta(0) := \Psi(0)$. Thus, $Z_\eta(t) = \Psi(0) + \int_0^t \mathrm{d}Z_\eta(s)$ for all $t \geq 0$. We then integrate $\mathrm{d}Z_\eta(t)$ over the time interval $[0, T]$, and apply Itô's Lemma (Lemma 6) to obtain an expression for the Frobenius norm of this integral. In the following, we fix $\eta_{ij} = 0$ for all i, j.

Lemma 19 *For any $T \geq t_0 \geq 0$,*

$$\mathbb{E}\left[\|Z_\eta(T) - Z_\eta(t_0)\|_F^2\right] = 32\int_{t_0}^T \mathbb{E}\left[\sum_{i=1}^d \sum_{j\neq i} \frac{(\lambda_i(t) - \lambda_j(t))^2}{\mu_{ij}^2(t)} \mathrm{d}t\right] + T\int_{t_0}^T \mathbb{E}\left[\sum_{i=1}^d \left(\sum_{j\neq i} \frac{\lambda_i(t) - \lambda_j(t)}{\mu_{ij}^2(t)}\right)^2\right] \mathrm{d}t.$$

The proof of Lemma 19 is a slight modification of the proof of Lemma 4.5 in Mangoubi and Vishnoi (2022), to accommodate the fact that $\lambda_i(t)$ varies over time (which is in contrast to Mangoubi and Vishnoi (2022) where λ_i is constant), and the fact that our matrices have complex-valued entries. The proof of Lemma 19 is given in Section F.1. We now proceed to the proof of Theorem 2.

Proof [Proof of theorem 2] In the following, we set $t_0 = \frac{1}{(kd)^{10} + k\alpha^2 + \sigma_1^2}$. We first compute the Ito derivative of $\Psi(t) := \sum_{i=1}^d \lambda_i(t) u_i(t) u_i^*(t)$:

$$\begin{aligned}\mathrm{d}\Psi(t) &= \sum_{i=1}^d (\lambda_i(t) + \mathrm{d}\lambda_i(t))(u_i(t)u_i^*(t) + \mathrm{d}(u_i(t)u_i^*(t))) - \lambda_i(t)u_i(t)u_i^*(t) \\ &= \sum_{i=1}^d \lambda_i(t)\mathrm{d}(u_i(t)u_i^*(t))) + (\mathrm{d}\lambda_i(t))(u_i(t)u_i^*(t)) + \mathrm{d}\lambda_i(t)\mathrm{d}(u_i(t)u_i^*(t)). \end{aligned} \tag{40}$$

From Lemma 18, we have that, for all $t \in [0, T]$,

$$\mathrm{d}(u_i(t)u_i^*(t)) = \sum_{j\neq i} \frac{1}{\gamma_i(t) - \gamma_j(t)}(u_i(t)u_j^*(t)\mathrm{d}B_{ij}(t) + u_j(t)u_i^*(t)\mathrm{d}B_{ij}^*(t)) + \sum_{j\neq i} \frac{\mathrm{d}t}{(\gamma_i(t) - \gamma_j(t))^2}(u_i(t)u_i^*(t) - u_j(t)u_j^*(t)). \tag{41}$$

For all $i \leq k$, we have that $\lambda_i(t) = \gamma_i(t)$ for all $t \geq 0$. Thus, by (41), we have that

$$
\begin{aligned}
&\mathrm{d}\lambda_i(t)\mathrm{d}(u_i(t)u_i^*(t)) \\
&= \mathrm{d}\gamma_i(t)\mathrm{d}(u_i(t)u_i^*(t)) \\
&\overset{\text{Eq. (41)}}{=} \mathrm{d}\gamma_i(t)\Big[\sum_{j\neq i} \frac{1}{\gamma_i(t)-\gamma_j(t)}(u_i(t)u_j^*(t)\mathrm{d}B_{ij}(t) + u_j(t)u_i^*(t)\mathrm{d}B_{ij}^*(t)) \\
&\qquad\qquad + \sum_{j\neq i} \frac{\mathrm{d}t}{(\gamma_i(t)-\gamma_j(t))^2}(u_i(t)u_i^*(t) - u_j(t)u_j^*(t))\Big] \\
&= \sum_{j\neq i} \frac{\mathrm{d}\gamma_i(t)}{\gamma_i(t)-\gamma_j(t)}(u_i(t)u_j^*(t)\mathrm{d}B_{ij}(t) + u_j(t)u_i^*(t)\mathrm{d}B_{ij}^*(t)) \\
&\qquad\qquad + \sum_{j\neq i} \frac{\mathrm{d}\gamma_i(t)\mathrm{d}t}{(\gamma_i(t)-\gamma_j(t))^2}(u_i(t)u_i^*(t) - u_j(t)u_j^*(t)) \\
&= \sum_{j\neq i}\left(\mathrm{d}B_{ii}(t) + \sum_{j\neq i}\frac{1}{\gamma_i(t)-\gamma_j(t)}\mathrm{d}t\right)\frac{1}{\gamma_i(t)-\gamma_j(t)}(u_i(t)u_j^*(t)\mathrm{d}B_{ij}(t) + u_j(t)u_i^*(t)\mathrm{d}B_{ij}^*(t)) \\
&+ \sum_{j\neq i}\left(\mathrm{d}B_{ii}(t) + \sum_{j\neq i}\frac{1}{\gamma_i(t)-\gamma_j(t)}\mathrm{d}t\right)\frac{\mathrm{d}t}{(\gamma_i(t)-\gamma_j(t))^2}(u_i(t)u_i^*(t) - u_j(t)u_j^*(t)) \\
&= \sum_{j\neq i}\left(\mathrm{d}B_{ii}(t)\mathrm{d}B_{ij}(t) + 2\sum_{j\neq i}\frac{1}{\gamma_i(t)-\gamma_j(t)}\mathrm{d}t\,\mathrm{d}B_{ij}(t)\right)\frac{1}{\gamma_i(t)-\gamma_j(t)}u_i(t)u_j^*(t) \\
&+ \sum_{j\neq i}\left(\mathrm{d}B_{ii}(t)\mathrm{d}B_{ij}^*(t) + 2\sum_{j\neq i}\frac{1}{\gamma_i(t)-\gamma_j(t)}\mathrm{d}t\,\mathrm{d}B_{ij}^*(t)\right)\frac{1}{\gamma_i(t)-\gamma_j(t)}u_j(t)u_i^*(t) \\
&+ \sum_{j\neq i}\left(\mathrm{d}B_{ii}(t)\mathrm{d}t + 2\sum_{j\neq i}\frac{1}{\gamma_i(t)-\gamma_j(t)}(\mathrm{d}t)^2\right)\frac{1}{(\gamma_i(t)-\gamma_j(t))^2}(u_i(t)u_i^*(t) - u_j(t)u_j^*(t)) \\
&= 0, \qquad\qquad \forall i \leq k,
\end{aligned} \tag{42}
$$

where the last equality holds since, for all $i, j \in [d]$, the Ito differentials $\mathrm{d}B_{ii}(t)\mathrm{d}B_{ij}(t)$ and $\mathrm{d}B_{ii}(t)\mathrm{d}B_{ij}^*(t)$ vanish because $\mathrm{d}B_{ii}(t)$ and $\mathrm{d}B_{ij}(t)$ are uncorrelated with mean zero, and the Ito differentials $\mathrm{d}B_{ii}(t)\mathrm{d}t$ and $(\mathrm{d}t)^2$ vanish because they are higher-order terms. Therefore, plugging in (42) into (40), we have that

$$
\begin{aligned}
\mathrm{d}\Psi(t) &\overset{\text{Eq. (40)}}{=} \sum_{i=1}^{d} \lambda_i(t)\mathrm{d}(u_i(t)u_i^*(t))) + (\mathrm{d}\lambda_i(t))(u_i(t)u_i^*(t)) + \mathrm{d}\lambda_i(t)\mathrm{d}(u_i(t)u_i^*(t)) \\
&= \sum_{i=1}^{k} \lambda_i(t)\mathrm{d}(u_i(t)u_i^*(t))) + (\mathrm{d}\lambda_i(t))(u_i(t)u_i^*(t)) + \mathrm{d}\lambda_i(t)\mathrm{d}(u_i(t)u_i^*(t)) \\
&\overset{\text{Eq. (42)}}{=} \sum_{i=1}^{k} \lambda_i(t)\mathrm{d}(u_i(t)u_i^*(t))) + (\mathrm{d}\lambda_i(t))(u_i(t)u_i^*(t)),
\end{aligned} \tag{43}
$$

where the second equality holds since $\lambda_i(t) = 0$ for all $i > k$ and all $t \geq 0$. Therefore, we have

$$
\begin{aligned}
&\mathbb{E}\left[\|\Psi(T) - \Psi(t_0)\|_F^2 \times \mathbb{1}_{\hat{E}_\alpha^c}\right] \\
&= \mathbb{E}\left[\left\|\int_{t_0}^T \sum_{i=1}^d \lambda_i(t)\mathrm{d}(u_i(t)u_i^*(t)) + (\mathrm{d}\lambda_i(t))u_i(t)u_i^*(t)\right\|_F^2 \times \mathbb{1}_{\hat{E}_\alpha^c}\right] \\
&\leq \mathbb{E}\left[\left\|\int_{t_0}^T \sum_{i=1}^d \lambda_i(t)\mathrm{d}(u_i(t)u_i^*(t))\right\|_F^2 \times \mathbb{1}_{\hat{E}_\alpha^c}\right] \\
&\quad + \mathbb{E}\left[\left\|\int_{t_0}^T \sum_{i=1}^d (\mathrm{d}\lambda_i(t))u_i(t)u_i^*(t)\right\|_F^2 \times \mathbb{1}_{\hat{E}_\alpha^c}\right] \\
&\leq \mathbb{E}\left[\left\|\int_{t_0}^T \sum_{i=1}^d \lambda_i(t)\mathrm{d}(u_i(t)u_i^*(t))\right\|_F^2 \times \mathbb{1}_{\hat{E}_\alpha^c}\right] \\
&\quad + \mathbb{E}\left[\left\|\int_{t_0}^T \sum_{i=1}^d (\mathrm{d}\lambda_i(t))u_i(t)u_i^*(t)\right\|_F^2\right] \\
&\overset{\text{Lem. 16}}{\leq} \mathbb{E}\left[\|Z_\eta(T) - Z_\eta(t_0)\|_F^2 \times \mathbb{1}_{\hat{E}_\alpha^c}\right] \\
&\quad + \mathbb{E}\left[\left\|\int_{t_0}^T \sum_{i=1}^d (\mathrm{d}\lambda_i(t))u_i(t)u_i^*(t)\right\|_F^2 \times \mathbb{1}_{\hat{E}_\alpha^c}\right] \\
&\overset{\text{Lem. 19}}{\leq} 32\int_{t_0}^T \mathbb{E}\left[\sum_{i=1}^d \sum_{j\neq i} \frac{(\lambda_i(t) - \lambda_j(t))^2}{(\gamma_i(t) - \gamma_j(t))^2} \times \mathbb{1}_{\hat{E}_\alpha^c}\right]\mathrm{d}t \\
&\quad + T\int_{t_0}^T \mathbb{E}\left[\sum_{i=1}^d \left(\sum_{j\neq i} \frac{\lambda_i(t) - \lambda_j(t)}{(\gamma_i(t) - \gamma_j(t))^2}\right)^2 \times \mathbb{1}_{\hat{E}_\alpha^c}\right]\mathrm{d}t \\
&\quad + \mathbb{E}\left[\left\|\int_{t_0}^T \sum_{i=1}^d (\mathrm{d}\lambda_i(t))u_i(t)u_i^*(t)\right\|_F^2\right],
\end{aligned}
\tag{44}
$$

Where the last inequality holds by Lemma 19. Plugging in $\lambda_i(t) = \gamma_i(t)$ for $i \leq k$ and $\lambda_i(t) = 0$ for $i > k$ into (44), we have

$$
\begin{aligned}
&\mathbb{E}\left[\|\Psi(T)-\Psi(t_0)\|_F^2 \times \mathbb{1}_{\hat{E}_\alpha^c}\right]\\
&\leq 2\int_{t_0}^{T}\mathbb{E}\left[\sum_{i=1}^{d}\sum_{j\neq i}\frac{(\lambda_i(t)-\lambda_j(t))^2}{(\gamma_i(t)-\gamma_j(t))^2}\times \mathbb{1}_{\hat{E}_\alpha^c}\right]\mathrm{d}t\\
&+T\int_{t_0}^{T}\mathbb{E}\left[\sum_{i=1}^{d}\left(\sum_{j\neq i}\frac{\lambda_i(t)-\lambda_j(t)}{(\gamma_i(t)-\gamma_j(t))^2}\right)^2\times \mathbb{1}_{\hat{E}_\alpha^c}\right]\mathrm{d}t+\mathbb{E}\left[\left\|\int_{t_0}^{T}\sum_{i=1}^{d}(\mathrm{d}\lambda_i(t))u_i(t)u_i^*(t)\right\|_F^2\times \mathbb{1}_{\hat{E}_\alpha^c}\right]\\
&\leq 32\int_{t_0}^{T}\mathbb{E}\left[\sum_{i=1}^{k}\left(k+\sum_{j>k}\frac{(\gamma_i(t))^2}{(\gamma_i(t)-\gamma_j(t))^2}\right)\times \mathbb{1}_{\hat{E}_\alpha^c}\right]\mathrm{d}t\\
&+T\int_{t_0}^{T}\mathbb{E}\left[\sum_{i=1}^{k}\left(\sum_{j\neq i:j\leq k}\frac{1}{\gamma_i(t)-\gamma_j(t)}+\sum_{j>k}\frac{\gamma_i(t)}{(\gamma_i(t)-\gamma_j(t))^2}\right)^2\times \mathbb{1}_{\hat{E}_\alpha^c}\right]\mathrm{d}t\\
&+\mathbb{E}\left[\left\|\int_{t_0}^{T}\sum_{i=1}^{d}(\mathrm{d}\lambda_i(t))u_i(t)u_i^*(t)\right\|_F^2\times \mathbb{1}_{\hat{E}_\alpha^c}\right]\\
&\leq 32\int_{t_0}^{T}\mathbb{E}\left[\sum_{i=1}^{k}\left(k+\sum_{j>k}\frac{(\gamma_i(t))^2}{(\gamma_i(t)-\gamma_j(t))^2}\right)\times \mathbb{1}_{\hat{E}_\alpha^c}\right]\mathrm{d}t\\
&+4T\int_{t_0}^{T}\sum_{i=1}^{k}\mathbb{E}\left[\left(\sum_{j\neq i:j\leq k}\frac{1}{\gamma_i(t)-\gamma_j(t)}\right)^2\times \mathbb{1}_{\hat{E}_\alpha^c}\right]+\mathbb{E}\left[\left(\sum_{j>k}\frac{\gamma_i(t)}{(\gamma_i(t)-\gamma_j(t))^2}\right)^2\times \mathbb{1}_{\hat{E}_\alpha^c}\right]\mathrm{d}t\\
&+\mathbb{E}\left[\left\|\int_{t_0}^{T}\sum_{i=1}^{d}(\mathrm{d}\lambda_i(t))u_i(t)u_i^*(t)\right\|_F^2\times \mathbb{1}_{\hat{E}_\alpha^c}\right].
\end{aligned}
\tag{45}
$$

Bounding the second moment of the inverse gaps: By Corollary 24 we have that for every $1 \leq i < j \leq d$,

$$
\mathbb{P}\left(\left\{\gamma_i(t)-\gamma_j(t)\leq (j-i)\times s\frac{\sqrt{t}}{\mathfrak{b}\sqrt{d}}\right\}\cap \hat{E}_\alpha^c\right)\leq s^3 \qquad \forall s>0, t>0. \tag{46}
$$

Thus, for $t \le T$,

$$\begin{aligned}
&\mathbb{E}\left[\frac{1}{(\gamma_i(t)-\gamma_j(t))^2} \times \mathbb{1}_{\hat{E}_\alpha^c}\right] \\
&\le \mathbb{E}\left[\frac{1}{(\gamma_i(t)-\gamma_j(t))^2} \times \mathbb{1}\left\{\gamma_i(t)-\gamma_j(t) \le (j-i) \times \frac{\sqrt{t}}{\mathfrak{b}\sqrt{d}}\right\} \times \mathbb{1}_{\hat{E}_\alpha^c}\right] + \frac{\mathfrak{b}^2 d}{(j-i)^2 t} \\
&= \int_{\frac{\mathfrak{b}^2 d}{(j-i)^2 t}}^{\infty} \mathbb{P}\left(\left\{\frac{1}{(\gamma_i(t)-\gamma_{i+1}(t))^2} \ge s\right\} \cap \hat{E}_\alpha^c\right) \mathrm{d}s + \frac{\mathfrak{b}^2 d}{(j-i)^2 t} \\
&= \int_{\frac{\mathfrak{b}^2 d}{(j-i)^2 t}}^{\infty} \mathbb{P}\left(\left\{(\gamma_i(t)-\gamma_{i+1}(t))^2 \le s^{-1}\right\} \cap \hat{E}_\alpha^c\right) \mathrm{d}s + \frac{\mathfrak{b}^2 d}{(j-i)^2 t} \\
&= \int_{\frac{\mathfrak{b}^2 d}{(j-i)^2 t}}^{\infty} \mathbb{P}\left(\left\{\gamma_i(t)-\gamma_{i+1}(t) \le s^{-\frac{1}{2}}\right\} \cap \hat{E}_\alpha^c\right) \mathrm{d}s + \frac{\mathfrak{b}^2 d}{(j-i)^2 t} \\
&\overset{\text{Eq. (46)}}{\le} \int_{\frac{\mathfrak{b}^2 d}{(j-i)^2 t}}^{\infty} \left(\frac{\mathfrak{b}^2 d}{(j-i)^2 t}\right)^{\frac{3}{2}} s^{-\frac{3}{2}} \mathrm{d}s + \frac{\mathfrak{b}^2 d}{(j-i)^2 t} \\
&= -\frac{3}{2}\left(\frac{\mathfrak{b}^2 d}{(j-i)^2 t}\right)^{\frac{3}{2}} s^{-\frac{1}{2}}\Big|_{\frac{\mathfrak{b}^2 d}{(j-i)^2 t}}^{\infty} + \frac{\mathfrak{b}^2 d}{(j-i)^2 t} \\
&= \frac{3}{2}\frac{\mathfrak{b}^2 d}{(j-i)^2 t} + \frac{\mathfrak{b}^2 d}{(j-i)^2 t} \\
&\le 3\frac{\mathfrak{b}^2 d}{(j-i)^2 t}. \qquad (47)
\end{aligned}$$

Thus, for any $t_0 > 0$,

$$\begin{aligned}
\mathbb{E}\left[\int_{t_0}^{T} \frac{1}{(\gamma_i(t)-\gamma_j(t))^2}\mathrm{d}t \times \mathbb{1}_{\hat{E}_\alpha^c}\right] &= \int_{t_0}^{T} \mathbb{E}\left[\frac{1}{(\gamma_i(t)-\gamma_j(t))^2} \times \mathbb{1}_{\hat{E}_\alpha^c}\right]\mathrm{d}t \\
&\le 3\int_{t_0}^{T} \frac{\mathfrak{b}^2 d}{(j-i)^2 t}\mathrm{d}t \\
&= 3\frac{\mathfrak{b}^2 d}{(j-i)^2} \log(t)|_{t_0}^{T} \\
&= 3\frac{\mathfrak{b}^2 d}{(j-i)^2} \times (\log(T) - \log(t_0)).
\end{aligned}$$

Bounding the term $\mathbb{E}\left[\left\|\int_0^T \sum_{i=1}^d (\mathrm{d}\lambda_i(t)) u_i(t) u_i^*(t)\right\|_F^2 \times \mathbb{1}_{\hat{E}_\alpha^c}\right]$**:** For $i > k$, $\mathrm{d}\lambda_i(t) = 0$. For $i \le k$, we have $\lambda_i(t) = \gamma_i(t)$ and thus,

$$\begin{aligned}
(\mathrm{d}\lambda_i(t)) u_i(t) u_i^*(t) &= (\mathrm{d}\gamma_i(t)) u_i(t) u_i^*(t) \\
&= \left(\mathrm{d}B_{ii}(t) + 2\sum_{j \ne i} \frac{1}{\gamma_i(t)-\gamma_j(t)}\mathrm{d}t\right) u_i(t) u_i^*(t).
\end{aligned}$$

Thus,

$$\mathbb{E}\left[\left\|\int_{t_0}^{T}\sum_{i=1}^{d}(\mathrm{d}\lambda_i(t))u_i(t)u_i^*(t)\right\|_F^2 \times \mathbb{1}_{\hat{E}_\alpha^c}\right] = \mathbb{E}\left[\left\|\int_{t_0}^{T}\sum_{i=1}^{k}(\mathrm{d}\gamma_i(t))u_i(t)u_i^*(t)\right\|_F^2 \times \mathbb{1}_{\hat{E}_\alpha^c}\right]$$
$$= \mathbb{E}\left[\left\|\int_{t_0}^{T}\sum_{i=1}^{k}\left(\mathrm{d}B_{ii}(t) + 2\sum_{j\neq i}\frac{1}{\gamma_i(t)-\gamma_j(t)}\mathrm{d}t\right)u_i(t)u_i^*(t)\right\|_F^2 \times \mathbb{1}_{\hat{E}_\alpha^c}\right]$$
$$\leq 3\mathbb{E}\left[\left\|\int_{t_0}^{T}\sum_{i=1}^{k}u_i(t)u_i^*(t)\mathrm{d}B_{ii}(t)\right\|_F^2 \times \mathbb{1}_{\hat{E}_\alpha^c}\right]$$
$$+ 6\mathbb{E}\left[\left\|\int_{t_0}^{T}\sum_{i=1}^{k}\sum_{j\neq i}\frac{1}{\gamma_i(t)-\gamma_j(t)}u_i(t)u_i^*(t)\mathrm{d}t\right\|_F^2 \times \mathbb{1}_{\hat{E}_\alpha^c}\right]. \tag{48}$$

To bound the first term on the r.h.s. of (48), we note that since $\mathrm{d}B_{ii}(t)$ are i.i.d. for all $t > 0$, we have if we plug in $f(X) := \|X\|_F^2 = \sum_{i=1}^{d}\sum_{j=1}^{d} X_{ij}^2$ into Ito's Lemma (Lemma 6) that

$$\mathbb{E}\left[\left\|\int_{t_0}^{T}\sum_{i=1}^{d}u_i(t)u_i^*(t)\mathrm{d}B_{ii}(t)\right\|_F^2 \times \mathbb{1}_{\hat{E}_\alpha^c}\right] = \mathbb{E}\left[\left\|\int_{t_0}^{T}\sum_{i=1}^{d}u_i(t)u_i^*(t)\mathrm{d}B_{ii}(t)\right\|_F^2\right]$$
$$\overset{\text{Lemma 6}}{=} \sum_{i=1}^{d}\int_{t_0}^{T}\|u_i(t)u_i^*(t)\|_F^2\mathrm{d}t$$
$$= (T-t_0)d. \tag{49}$$

To bound the second term on the R.H.S. of (48), we have

$$\mathbb{E}\left[\left\|\int_{t_0}^{T}\sum_{i=1}^{k}\sum_{j\neq i}\frac{1}{\gamma_i(t)-\gamma_j(t)}u_i(t)u_i^*(t)\mathrm{d}t\right\|_F^2 \times \mathbb{1}_{\hat{E}_\alpha^c}\right]$$
$$\leq \mathbb{E}\left[\int_{t_0}^{T}\left\|\sum_{i=1}^{k}\sum_{j\neq i}\frac{1}{\gamma_i(t)-\gamma_j(t)}u_i(t)u_i^*(t)\right\|_F^2\mathrm{d}t \times \int_{t_0}^{T}1^2\mathrm{d}t \times \mathbb{1}_{\hat{E}_\alpha^c}\right]$$
$$= (T-t_0)\mathbb{E}\left[\int_{t_0}^{T}\left\|\sum_{i=1}^{k}\sum_{j\neq i}\frac{1}{\gamma_i(t)-\gamma_j(t)}u_i(t)u_i^*(t)\right\|_F^2\mathrm{d}t \times \mathbb{1}_{\hat{E}_\alpha^c}\right]$$
$$= (T-t_0)\mathbb{E}\left[\int_{t_0}^{T}\sum_{i=1}^{k}\left\|\sum_{j\neq i}\frac{1}{\gamma_i(t)-\gamma_j(t)}u_i(t)u_i^*(t)\right\|_F^2\mathrm{d}t \times \mathbb{1}_{\hat{E}_\alpha^c}\right]$$
$$= (T-t_0)\mathbb{E}\left[\int_{t_0}^{T}\sum_{i=1}^{k}\left(\sum_{j\neq i}\frac{1}{\gamma_i(t)-\gamma_j(t)}\right)^2\|u_i(t)u_i^*(t)\|_F^2\,\mathrm{d}t \times \mathbb{1}_{\hat{E}_\alpha^c}\right],$$
$$= (T-t_0)\int_{t_0}^{T}\sum_{i=1}^{k}\mathbb{E}\left[\left(\sum_{j\neq i}\frac{1}{\gamma_i(t)-\gamma_j(t)}\right)^2 \times \mathbb{1}_{\hat{E}_\alpha^c}\right]\mathrm{d}t, \tag{50}$$

where the first inequality is by the Cauchy-Schwartz inequality, and the second equality holds since $\langle u_i(t)u_i^*(t),\ u_\ell(t)u_\ell^*(t)\rangle = 0$ for all $i \neq \ell$. Therefore, plugging in (49) and (50) into (48), we have

$$\mathbb{E}\left[\left\|\int_{t_0}^{T}\sum_{i=1}^{d}(\mathrm{d}\lambda_i(t))u_i(t)u_i^*(t)\right\|_F^2 \times \mathbb{1}_{\hat{E}_\alpha^c}\right]$$
$$\leq 3(T-t_0)d + 6(T-t_0)\int_{t_0}^{T}\sum_{i=1}^{k}\mathbb{E}\left[\left(\sum_{j\neq i}\frac{1}{\gamma_i(t)-\gamma_j(t)}\right)^2 \times \mathbb{1}_{\hat{E}_\alpha^c}\right]\mathrm{d}t. \tag{51}$$

Bounding the term $\mathbb{E}\left[\left(\sum_{j\neq i}\frac{1}{\gamma_i(t)-\gamma_j(t)}\right)^2 \times \mathbb{1}_{\hat{E}_\alpha^c}\right]$: Consider any subset $S \subseteq \{1,\ldots,d\}$. Then

$$\mathbb{E}\left[\left(\sum_{j\in S, j\neq i}\frac{1}{\gamma_i(t)-\gamma_j(t)}\right)^2 \times \mathbb{1}_{\hat{E}_\alpha^c}\right]$$
$$= \mathbb{E}\left[\sum_{j\in S, j\neq i}\sum_{\ell\in S,\ell\neq i}\frac{1}{(\gamma_i(t)-\gamma_j(t))(\gamma_i(t)-\gamma_\ell(t))}\times \mathbb{1}_{\hat{E}_\alpha^c}\right]$$
$$= \sum_{j\in S, j\neq i}\sum_{\ell\in S,\ell\neq i}\mathbb{E}\left[\frac{1}{(j-i)(\ell-i)\frac{\gamma_i(t)-\gamma_j(t)}{j-i}\times\frac{\gamma_i(t)-\gamma_\ell(t)}{\ell-i}}\times \mathbb{1}_{\hat{E}_\alpha^c}\right]$$
$$= \sum_{j\in S, j\neq i}\sum_{\ell\in S,\ell\neq i}\frac{1}{(j-i)(\ell-i)}\mathbb{E}\left[\frac{1}{\frac{\gamma_i(t)-\gamma_j(t)}{j-i}\times\frac{\gamma_i(t)-\gamma_\ell(t)}{\ell-i}}\times \mathbb{1}_{\hat{E}_\alpha^c}\right]$$
$$\leq 2\sum_{j\in S, j\neq i}\sum_{\ell\in S,\ell\neq i}\frac{1}{|(j-i)(\ell-i)|}\left(\mathbb{E}\left[\frac{1}{\left(\frac{\gamma_i(t)-\gamma_j(t)}{j-i}\right)^2}\times \mathbb{1}_{\hat{E}_\alpha^c}\right]+\mathbb{E}\left[\frac{1}{\left(\frac{\gamma_i(t)-\gamma_\ell(t)}{\ell-i}\right)^2}\times \mathbb{1}_{\hat{E}_\alpha^c}\right]\right)$$
$$\overset{\text{Eq. (47)}}{\leq} 2\sum_{j\in S, j\neq i}\sum_{\ell\in S,\ell\neq i}\frac{1}{|(j-i)(\ell-i)|}\left(3\mathfrak{b}^2\frac{d}{t}+3\mathfrak{b}^2\frac{d}{t}\right)$$
$$= 12\mathfrak{b}^2\frac{d}{t}\sum_{j\in S, j\neq i}\sum_{\ell\in S,\ell\neq i}\frac{1}{|(j-i)(\ell-i)|}$$
$$= 12\mathfrak{b}^2\frac{d}{t}\sum_{j\in S, j\neq i}\frac{1}{|j-i|}\sum_{\ell\in S,\ell\neq i}\frac{1}{|\ell-i|}$$
$$\leq 12\mathfrak{b}^2\frac{d}{t}\sum_{\ell\in S,\ell\neq i}\frac{1}{|j-i|}\log d$$
$$\leq 12\mathfrak{b}^2\frac{d}{t}\log^2 d. \tag{52}$$

Completing the proof:

$$
\begin{aligned}
&\mathbb{E}\left[\|\Psi(T)-\Psi(t_0)\|_F^2 \times \mathbb{1}_{\hat{E}_\alpha^c}\right] \\
&\overset{\text{Eq. (45), (51)}}{\leq} 32\int_{t_0}^{T} \mathbb{E}\left[\sum_{i=1}^{k}\left(k+\sum_{j>k}\frac{(\gamma_i(t))^2}{(\gamma_i(t)-\gamma_j(t))^2}\right)\times \mathbb{1}_{\hat{E}_\alpha^c}\right]\mathrm{d}t \\
&+4T\int_{t_0}^{T}\sum_{i=1}^{k}\mathbb{E}\left[\left(\sum_{j\neq i: j\leq k}\frac{1}{\gamma_i(t)-\gamma_j(t)}\right)^2\times \mathbb{1}_{\hat{E}_\alpha^c}\right]+\mathbb{E}\left[\left(\sum_{j>k}\frac{\gamma_i(t)}{(\gamma_i(t)-\gamma_j(t))^2}\right)^2\times \mathbb{1}_{\hat{E}_\alpha^c}\right]\mathrm{d}t \\
&+3(T-t_0)d+6(T-t_0)\int_{t_0}^{T}\sum_{i=1}^{k}\mathbb{E}\left[\left(\sum_{j\neq i}\frac{1}{\gamma_i(t)-\gamma_j(t)}\right)^2\times \mathbb{1}_{\hat{E}_\alpha^c}\right]\mathrm{d}t. \\
&\overset{\text{Eq. (52)}}{\leq} 32\int_{t_0}^{T} \mathbb{E}\left[\sum_{i=1}^{k}\left(k+\sum_{j>k}\frac{(\gamma_i(t))^2}{(\gamma_i(t)-\gamma_j(t))^2}\right)\times \mathbb{1}_{\hat{E}_\alpha^c}\right]\mathrm{d}t \\
&+4T\int_{t_0}^{T}\sum_{i=1}^{k}12\mathfrak{b}^2\frac{d}{t}\log^2 d+\mathbb{E}\left[\left(\sum_{j>k}\frac{\gamma_i(t)}{(\gamma_i(t)-\gamma_j(t))^2}\right)^2\times \mathbb{1}_{\hat{E}_\alpha^c}\right]\mathrm{d}t \\
&+3(T-t_0)d+6(T-t_0)\int_{t_0}^{T}\sum_{i=1}^{k}12\mathfrak{b}^2\frac{d}{t}\log^2 d \\
&\leq 32\int_{t_0}^{T}\mathbb{E}\left[\sum_{i=1}^{k}\left(k+\sum_{j>k}16\frac{\sigma_k^2}{(\sigma_k-\sigma_{k+1})^2}\right)\right]\mathrm{d}t \\
&+48\mathfrak{b}^2Tkd(\log T-\log t_0)\log^2 d+\int_{t_0}^{T}\mathbb{E}\left[\left(\sum_{j>k}16\frac{\sigma_k}{(\sigma_k-\sigma_{k+1})\sqrt{T}\sqrt{d}}\right)^2\right]\mathrm{d}t \\
&+3(T-t_0)d+6(T-t_0)12\mathfrak{b}^2kd(\log^2 d)(\log(T)-\log(t_0)) \\
&\leq 32T\left(k^2+16kd\frac{\sigma_k^2}{(\sigma_k-\sigma_{k+1})^2}\right) \\
&+48\mathfrak{b}^2Tkd(\log T-\log t_0)\log^2 d+\frac{16^2}{T}d\frac{\sigma_k^2}{(\sigma_k-\sigma_{k+1})^2}(T-t_0) \\
&+3(T-t_0)d+6(T-t_0)12\mathfrak{b}^2kd(\log^2 d)(\log(T)-\log(t_0)) \\
&\leq \frac{1}{2}10^4\mathfrak{b}^2kdT\frac{\sigma_k^2}{(\sigma_k-\sigma_{k+1})^2}(\log^3 d)\log(\sigma_1+T),
\end{aligned} \tag{53}
$$

where the last inequality of (53) holds because $-\log t_0 \leq 20\log(d)$ since $t_0=\frac{1}{(kd)^{10}+k\alpha^2+\sigma_1^2}\geq\frac{1}{(kd)^{10}+400k\log(\sigma_1+T)+\sigma_1^2}$. Moreover, the third inequality of (53) holds because Lemma 16 implies that, since $\sigma_k-\sigma_{k+1}\geq\sqrt{T}\sqrt{d}+40\log^{\frac{1}{2}}(\sigma_1+T)$, whenever $\hat{E}_\alpha^c$ occurs, we have

$$
\gamma_k(t)-\gamma_{k+1}(t)\geq\frac{1}{2}((\sigma_k-\sigma_{k+1})-\alpha)\geq\frac{1}{4}((\sigma_k-\sigma_{k+1})-\alpha)\geq\frac{1}{4}\sqrt{T}\sqrt{d}\qquad\forall t\geq 0,
$$

because $\alpha = 20\log^{\frac{1}{2}}(\sigma_1 + T)$. Therefore, plugging in (53) into Lemma 15, we have that

$$\begin{aligned}
&\mathbb{E}[\|\Psi(T) - \Psi(0)\|_F^2] \\
&\overset{\text{Lemma 15}}{\leq} 4\mathbb{E}[\|\Psi(T) - \Psi(0)\|_F^2 \times \mathbb{1}_{\hat{E}_\alpha^c}] + d + \frac{T}{d^{200}} \\
&\leq 16\mathbb{E}[\|\Psi(T) - \Psi(t_0)\|_F^2 \times \mathbb{1}_{\hat{E}_\alpha^c}] + 16\mathbb{E}[\|\Psi(t_0) - \Psi(0)\|_F^2 \times \mathbb{1}_{\hat{E}_\alpha^c}] + d + \frac{T}{d^{200}} \\
&\overset{\text{Eq. (53)}}{\leq} \frac{1}{4}10^6\mathfrak{b}^2 kdT\frac{\sigma_k^2}{(\sigma_k - \sigma_{k+1})^2}(\log^3 d) + 16\mathbb{E}[\|\Psi(t_0) - \Psi(0)\|_F^2 \times \mathbb{1}_{\hat{E}_\alpha^c}] + d + \frac{T}{d^{200}} \\
&\leq \frac{1}{2}10^6\mathfrak{b}^2 kdT\frac{\sigma_k^2}{(\sigma_k - \sigma_{k+1})^2}(\log^3 d) + \mathbb{E}[\|\Psi(t_0) - \Psi(0)\|_F^2 \times \mathbb{1}_{\hat{E}_\alpha^c}] \\
&\overset{\text{Lemma 17}}{\leq} \frac{1}{2}10^6\mathfrak{b}^2 kdT\frac{\sigma_k^2}{(\sigma_k - \sigma_{k+1})^2}(\log^3 d) + 40^2 \\
&\leq 10^6\mathfrak{b}^2 kdT\frac{\sigma_k^2}{(\sigma_k - \sigma_{k+1})^2}(\log^3 d)\log(\sigma_1 + T),
\end{aligned} \tag{54}$$

where the fifth inequality holds by Lemma 17 since $t_0 = \frac{1}{(kd)^{10} + k\alpha^2 + \sigma_1^2}$. ■

Appendix D. Eigenvalue gap comparison result: Proof of Lemma 3

Proposition 20 *Consider any strong solutions γ, ξ to (4), for any $\beta \geq 1$. Suppose that for some $i \in [d]$ and at some time $t \geq 0$,*

$$\gamma_i(t) - \gamma_{i+1}(t) = \xi_i(t) - \xi_{i+1}(t) > 0 \tag{55}$$

and

$$\gamma_j(t) - \gamma_{j+1}(t) \geq \xi_j(t) - \xi_{j+1}(t) > 0 \qquad \forall j \in [d-1]. \tag{56}$$

Then

$$\mathrm{d}\gamma_i(t) - \mathrm{d}\gamma_{i+1}(t) \geq \mathrm{d}\xi_i(t) - \mathrm{d}\xi_{i+1}(t). \tag{57}$$

Proof First, note that for any numbers $b \geq c > 0$ and all $a > 0$ we have that

$$\frac{1}{a+b} - \frac{1}{b} \geq \frac{1}{a+c} - \frac{1}{c}. \tag{58}$$

Bounding the repulsion forces when $j > i+1$: For any $j > i+1$ we have that by (58) (setting $a = \gamma_i(t) - \gamma_{i+1}(t)$, $b = \gamma_{i+1}(t) - \gamma_j(t)$, and $c = \xi_{i+1}(t) - \xi_j(t)$, and noting that (56) implies that $b \geq c > 0$ since $j > i+1$),

$$\begin{aligned}
&\frac{1}{\gamma_i(t) - \gamma_{i+1}(t) + (\gamma_{i+1}(t) - \gamma_j(t))} - \frac{1}{\gamma_{i+1}(t) - \gamma_j(t)} \\
&\geq \frac{1}{\gamma_i(t) - \gamma_{i+1}(t) + (\xi_{i+1}(t) - \xi_j(t))} - \frac{1}{\xi_{i+1}(t) - \xi_j(t)}.
\end{aligned} \tag{59}$$

Plugging in (55) into (59), we get that

$$\frac{1}{\gamma_i(t)-\gamma_{i+1}(t)+(\gamma_{i+1}(t)-\gamma_j(t))}-\frac{1}{\gamma_{i+1}(t)-\gamma_j(t)} \geq \frac{1}{\xi_i(t)-\xi_{i+1}(t)+(\xi_{i+1}(t)-\xi_j(t))}-\frac{1}{\xi_{i+1}(t)-\xi_j(t)}. \tag{60}$$

Simplifying (60), we get

$$\frac{1}{\gamma_i(t)-\gamma_j(t)}-\frac{1}{\gamma_{i+1}(t)-\gamma_j(t)} \geq \frac{1}{\xi_i(t)-\xi_j(t)}-\frac{1}{\xi_{i+1}(t)-\xi_j(t)} \qquad \forall j > i+1. \tag{61}$$

Bounding the repulsion forces when $j < i$: Next, consider any $j < i$. Then by (58) (setting $a = \gamma_i(t) - \gamma_{i+1}(t)$, $b = \gamma_j(t) - \gamma_i(t)$, and $c = \xi_j(t) - \xi_i(t)$ into (58), and noting that (56) implies that $b \geq c > 0$ since $j < i$), we get

$$\frac{1}{\gamma_j(t)-\gamma_i(t)+(\gamma_i(t)-\gamma_{i+1}(t))}-\frac{1}{\gamma_j(t)-\gamma_i(t)} \geq \frac{1}{\xi_j(t)-\xi_i(t)+(\gamma_i(t)-\gamma_{i+1}(t))}-\frac{1}{\xi_j(t)-\xi_i(t)}. \tag{62}$$

Then plugging in (55) into (62), we have

$$\frac{1}{\gamma_j(t)-\gamma_i(t)+(\gamma_i(t)-\gamma_{i+1}(t))}-\frac{1}{\gamma_j(t)-\gamma_i(t)} \geq \frac{1}{\xi_j(t)-\xi_i(t)+(\xi_i(t)-\xi_{i+1}(t))}-\frac{1}{\xi_j(t)-\xi_i(t)}. \tag{63}$$

Simplifying (63), we get

$$\frac{1}{\gamma_i(t)-\gamma_j(t)}-\frac{1}{\gamma_{i+1}(t)-\gamma_j(t)} \geq \frac{1}{\xi_i(t)-\xi_j(t)}-\frac{1}{\xi_{i+1}(t)-\xi_j(t)} \qquad \forall j < i. \tag{64}$$

Therefore, (61) and (64) together imply that

$$\frac{1}{\gamma_i(t)-\gamma_j(t)}-\frac{1}{\gamma_{i+1}(t)-\gamma_j(t)} \geq \frac{1}{\xi_i(t)-\xi_j(t)}-\frac{1}{\xi_{i+1}(t)-\xi_j(t)} \qquad \forall j \in [d]\backslash\{i, i+1\}. \tag{65}$$

Bounding the gap derivative: By (4) and (65) we have that

$$\begin{aligned}
&\mathrm{d}\gamma_i(t) - \mathrm{d}\gamma_{i+1}(t) \\
&\overset{\text{Eq. (4)}}{=} \left(\mathrm{d}B_{i,i}(t) + \beta\sum_{j\neq i}\frac{1}{\gamma_i(t)-\gamma_j(t)}\mathrm{d}t\right) - \left(\mathrm{d}B_{i+1,i+1}(t) + \beta\sum_{j\neq i+1}\frac{1}{\gamma_{i+1}(t)-\gamma_j(t)}\mathrm{d}t\right) \\
&= \mathrm{d}B_{i,i}(t) - \mathrm{d}B_{i+1,i+1}(t) + \beta\mathrm{d}t\sum_{j\in[d]\backslash\{i,i+1\}}\frac{1}{\gamma_i(t)-\gamma_j(t)} - \frac{1}{\gamma_{i+1}(t)-\gamma_j(t)} \\
&\overset{\text{Eq. (65)}}{\geq} \mathrm{d}B_{i,i}(t) - \mathrm{d}B_{i+1,i+1}(t) + \beta\mathrm{d}t\sum_{j\in[d]\backslash\{i,i+1\}}\frac{1}{\xi_i(t)-\xi_j(t)} - \frac{1}{\xi_{i+1}(t)-\xi_j(t)} \\
&= \left(\mathrm{d}B_{i,i}(t) + \beta\sum_{j\neq i}\frac{1}{\xi_i(t)-\xi_j(t)}\mathrm{d}t\right) - \left(\mathrm{d}B_{i+1,i+1}(t) + \beta\sum_{j\neq i+1}\frac{1}{\xi_{i+1}(t)-\xi_j(t)}\mathrm{d}t\right) \\
&= \mathrm{d}\xi_i(t) - \mathrm{d}\xi_{i+1}(t).
\end{aligned}$$

This proves (57) and completes the proof of the proposition. ■

Proposition 21 *Consider any strong solutions γ, ξ to (4), for any $\beta \geq 1$. Suppose that for some $i \in [d]$ and at some time $t \geq 0$,*

$$\gamma_i(t) - \gamma_{i+1}(t) = \xi_i(t) - \xi_{i+1}(t) > 0 \qquad \forall i \in [d-1]. \tag{66}$$

Then

$$\mathrm{d}\gamma_i(t) - \mathrm{d}\gamma_{i+1}(t) = \mathrm{d}\xi_i(t) - \mathrm{d}\xi_{i+1}(t) \qquad \forall i \in [d-1]. \tag{67}$$

Proof

$$\begin{aligned}
&\mathrm{d}\gamma_i(t) - \mathrm{d}\gamma_{i+1}(t) \\
&\overset{Eq.~(4)}{=} \left(\mathrm{d}B_{i,i}(t) + \beta \sum_{j \neq i} \frac{1}{\gamma_i(t) - \gamma_j(t)} \mathrm{d}t\right) - \left(\mathrm{d}B_{i+1,i+1}(t) + \beta \sum_{j \neq i+1} \frac{1}{\gamma_{i+1}(t) - \gamma_j(t)} \mathrm{d}t\right) \\
&\overset{Eq.~(65)}{=} \mathrm{d}B_{i,i}(t) - \mathrm{d}B_{i+1,i+1}(t) + \beta \mathrm{d}t \sum_{j \in [d] \setminus \{i, i+1\}} \frac{1}{\xi_i(t) - \xi_j(t)} - \frac{1}{\xi_{i+1}(t) - \xi_j(t)} \\
&= \mathrm{d}\xi_i(t) - \mathrm{d}\xi_{i+1}(t).
\end{aligned}$$

■

Proof [Proof of Lemma 3] First, we note that since by Lemma 8 at every time $t \geq 0$ the strong solution $\gamma(t)$ is a continuous function of the initial conditions $\gamma(0)$, without loss of generality we may assume that that the initial eigenvalue gaps of γ are strictly greater than the corresponding eigenvalue gaps of ξ:

$$\xi_i(0) - \xi_{i+1}(0) < \gamma_i(0) - \gamma_{i+1}(0) \qquad 1 \leq i < d. \tag{68}$$

We prove Lemma 3 by contradiction. Let $\tau := \inf\{t \geq 0 : \xi_i(t) - \xi_{i+1}(t) > \gamma_i(t) - \gamma_{i+1}(t)$ for some $i \in [d]\}$ to be the first time where the size of the i'th gaps "cross" for some $i \in [d]$ (in other words τ be the first time when the conclusion of Lemma 3 fails to hold).

Assumption towards a contradiction: Suppose (towards a contradiction) that $\tau < \infty$. By definition of strong solutions, strong solutions to stochastic differential are almost surely continuous on $[0, \infty)$, we have that both γ and ξ are almost surely continuous on $[0, \infty)$. Therefore, since $\tau < \infty$, by the intermediate value theorem, we must have that, for some $i \in [d]$ i'th gap of ξ and the i'th gap of γ are equal at the time τ, and that at this time τ all the other gaps of γ are at least as large as the corresponding gaps of ξ:

$$\gamma_i(\tau) - \gamma_{i+1}(\tau) = \xi_i(\tau) - \xi_{i+1}(\tau), \tag{69}$$

$$\gamma_j(\tau) - \gamma_{j+1}(\tau) \geq \xi_j(\tau) - \xi_{j+1}(\tau) \qquad \forall j \in [d-1]. \tag{70}$$

Moreover, by Lemma 9 we have that, almost surely, the particles of Dyson Brownian motion do not collide with each other on all of $(0,\infty)$. Therefore, we have that, almost surely,

$$\gamma_i(t) - \gamma_{i+1}(t) > 0 \qquad \forall t \in (0,\infty),\ i \in [d-1] \tag{71}$$

$$\xi_i(t) - \xi_{i+1}(t) > 0 \qquad \forall t \in (0,\infty),\ i \in [d-1]. \tag{72}$$

Therefore, plugging in (69), (70) and (71) into Proposition 20, we have that

$$\mathrm{d}\gamma_i(\tau) - \mathrm{d}\gamma_{i+1}(\tau) \geq \mathrm{d}\xi_i(\tau) - \mathrm{d}\xi_{i+1}(\tau). \tag{73}$$

Next, we consider two cases: when $\mathrm{d}\gamma_i(\tau) - \mathrm{d}\gamma_{i+1}(\tau) > \mathrm{d}\xi_i(\tau) - \mathrm{d}\xi_{i+1}(\tau)$, and when $\mathrm{d}\gamma_i(\tau) - \mathrm{d}\gamma_{i+1}(\tau) = \mathrm{d}\xi_i(\tau) - \mathrm{d}\xi_{i+1}(\tau)$.

Case 1, $\mathrm{d}\gamma_i(\tau) - \mathrm{d}\gamma_{i+1}(\tau) > \mathrm{d}\xi_i(\tau) - \mathrm{d}\xi_{i+1}(\tau)$**:** For any $w \in \mathcal{W}_d$, define the "drift" function

$$\mu_i(w) := \beta \sum_{j\neq i} \frac{1}{w_i - w_j}. \tag{74}$$

Then we have that

$$\begin{aligned}
&(\mathrm{d}\gamma_i(t) - \mathrm{d}\gamma_{i+1}(t)) - (\mathrm{d}\xi_i(t) - \mathrm{d}\xi_{i+1}(t)) \qquad (75)\\
&\overset{\text{Eq. (4)}}{=} \left[\left(\mathrm{d}B_{i,i}(t) + \beta\sum_{j\neq i}\frac{1}{\gamma_i(t)-\gamma_j(t)}\mathrm{d}t\right) - \left(\mathrm{d}B_{i+1,i+1}(t) + \beta\sum_{j\neq i+1}\frac{1}{\gamma_{i+1}(t)-\gamma_j(t)}\mathrm{d}t\right)\right]\\
&\quad - \left[\left(\mathrm{d}B_{i,i}(t) + \beta\sum_{j\neq i}\frac{1}{\xi_i(t)-\xi_j(t)}\mathrm{d}t\right) - \left(\mathrm{d}B_{i+1,i+1}(t) + \beta\sum_{j\neq i+1}\frac{1}{\xi_{i+1}(t)-\xi_j(t)}\mathrm{d}t\right)\right]\\
&= \mu_i(\gamma(t)) - \mu_{i+1}(\gamma(t)) - (\mu_i(\xi(t)) - \mu_{i+1}(\xi(t))).
\end{aligned}$$

From (71), all the gaps of γ and ξ are strictly greater than zero at time τ. Therefore, since γ and ξ are almost surely continuous on $[0,\infty)$, we must have that, $\mu(\gamma(t))$ and $\mu(\xi(t))$ are also continuous on all of $t \in (0,\infty)$.

Since $\mathrm{d}\gamma_i(\tau) - \mathrm{d}\gamma_{i+1}(\tau) > \mathrm{d}\xi_i(\tau) - \mathrm{d}\xi_{i+1}(\tau)$, we have by (75) that

$$\begin{aligned}
\mu_i(\gamma(\tau)) - \mu_{i+1}(\gamma(\tau)) - (\mu_i(\xi(\tau)) - \mu_{i+1}(\xi(\tau))) &= (\mathrm{d}\gamma_i(\tau) - \mathrm{d}\gamma_{i+1}(\tau)) - (\mathrm{d}\xi_i(\tau) - \mathrm{d}\xi_{i+1}(\tau))\\
&> 0. \qquad (76)
\end{aligned}$$

Therefore, since $\mu(\gamma(t))$ and $\mu(\xi(t))$ is almost surely continuous on $(0,\infty)$, by (76) we must have that there exits some open interval $\mathcal{I} \subset (0,\infty)$ containing τ such that

$$(\mathrm{d}\gamma_i(t) - \mathrm{d}\gamma_{i+1}(t) - (\mathrm{d}\xi_i(t) - \mathrm{d}\xi_{i+1}(t)) > 0. \qquad \forall t \in \mathcal{I}. \tag{77}$$

Consider any $t \in \mathcal{I}$ such that $t > \tau$. Then

$$\begin{aligned}
&(\gamma_i(t) - \gamma_{i+1}(t)) - (\xi_i(t) - \xi_{i+1}(t))\\
&\overset{\text{Eq. (70)}}{=} [(\gamma_i(t) - \gamma_{i+1}(t)) - (\xi_i(t) - \xi_{i+1}(t))] - [(\gamma_i(\tau) - \gamma_{i+1}(\tau)) - (\xi_i(\tau) - \xi_{i+1}(\tau))]\\
&= \int_\tau^t (\mathrm{d}\gamma_i(s) - \mathrm{d}\gamma_{i+1}(s) - (\mathrm{d}\xi_i(s) - \mathrm{d}\xi_{i+1}(s))\mathrm{d}s\\
&\overset{\text{Eq. (77)}}{>} 0. \qquad (78)
\end{aligned}$$

Therefore (78) implies that there exits some $\tau' \in \mathcal{I}$ where $\tau' > \tau$ such that

$$\gamma_i(t) - \gamma_{i+1}(t) > \xi_i(t) - \xi_{i+1}(t) \qquad \forall \tau < t < \tau'.$$

Therefore $\tau \neq \inf\{t \geq 0 : \xi_i(t) - \xi_{i+1}(t) > \gamma_i(t) - \gamma_{i+1}(t)\}$ for any $i \in [d]$. This contradicts the definition of τ. Therefore, by contradiction our assumption that $\tau < \infty$ is false.

Case 2, $\mathrm{d}\gamma_i(\tau) - \mathrm{d}\gamma_{i+1}(\tau) = \mathrm{d}\xi_i(\tau) - \mathrm{d}\xi_{i+1}(\tau)$**:** Consider the system of stochastic differential equation for the process $\gamma_i(t) - \gamma_{i+1}(t)$:

$$\mathrm{d}\gamma_i(t) - \mathrm{d}\gamma_{i+1}(t) \overset{\text{Eq. (4)}}{=} \left(\mathrm{d}B_{i,i}(t) + \beta \sum_{j \neq i} \frac{1}{\gamma_i(t) - \gamma_j(t)} \mathrm{d}t\right) - \left(\mathrm{d}B_{i+1,i+1}(t) + \beta \sum_{j \neq i+1} \frac{1}{\gamma_{i+1}(t) - \gamma_j(t)} \mathrm{d}t\right) \qquad \forall i \in [d] \quad (79)$$

and the system of stochastic differential equation for the process $\xi_i(t) - \xi_{i+1}(t)$:

$$\mathrm{d}\xi_i(t) - \mathrm{d}\xi_{i+1}(t) \overset{\text{Eq. (4)}}{=} \left(\mathrm{d}B_{i,i}(t) + \beta \sum_{j \neq i} \frac{1}{\xi_i(t) - \xi_j(t)} \mathrm{d}t\right) - \left(\mathrm{d}B_{i+1,i+1}(t) + \beta \sum_{j \neq i+1} \frac{1}{\xi_{i+1}(t) - \xi_j(t)} \mathrm{d}t\right) \qquad \forall i \in [d]. \quad (80)$$

Then we have that

$$\begin{aligned} 0 &= (\mathrm{d}\gamma_i(\tau) - \mathrm{d}\gamma_{i+1}(\tau)) - (\mathrm{d}\xi_i(\tau) - \mathrm{d}\xi_{i+1}(\tau)) \\ &\overset{\text{Eq. (4)}}{=} \left[\left(\beta \sum_{j \neq i} \frac{1}{\gamma_i(\tau) - \gamma_j(\tau)} \mathrm{d}t\right) - \left(\beta \sum_{j \neq i+1} \frac{1}{\gamma_{i+1}(\tau) - \gamma_j(\tau)} \mathrm{d}t\right)\right] \\ &\quad - \left[\left(\beta \sum_{j \neq i} \frac{1}{\xi_i(\tau) - \xi_j(\tau)} \mathrm{d}t\right) - \left(\beta \sum_{j \neq i+1} \frac{1}{\xi_{i+1}(\tau) - \xi_j(\tau)} \mathrm{d}t\right)\right]. \end{aligned} \quad (81)$$

But we also have from (69) that $\gamma_i(\tau) - \gamma_{i+1}(\tau) = \xi_i(\tau) - \xi_{i+1}(\tau)$ and from (70) that $\gamma_j(\tau) - \gamma_{j+1}(\tau) \geq \xi_j(\tau) - \xi_{j+1}(\tau)$ for all $j \in [d-1]$. Thus, the only way for the r.h.s. of (81) to be equal to zero is if we have $\xi_i(\tau) - \xi_{i+1}(\tau) = \gamma_i(\tau) - \gamma_{i+1}(\tau)$ for all $i \in [d-1]$. Therefore, since the $\xi_i(\tau) - \xi_{i+1}(\tau) = \gamma_i(\tau) - \gamma_{i+1}(\tau)$ for all $i \in [d-1]$.

Moreover, by Lemma 7, for any initial conditions $\gamma(\tau)$ and $\xi(\tau)$, the processes γ and ξ have unique strong solutions on $(0, \infty)$. Therefore, since the stochastic differential equation (4) for γ and ξ are invariant to spatial translation, we must have that

$$\xi_i(t) - \xi_{i+1}(t) = \gamma_i(t) - \gamma_{i+1}(t) \qquad \forall t \geq \tau, i \in [d]. \quad (82)$$

By (82), we have that $\tau = \inf\{t \geq 0 : \xi_i(t) - \xi_{i+1}(t) > \gamma_i(t) - \gamma_{i+1}(t) \text{ for some } i \in [d]\} = \infty$. This contradicts our assumption that $\tau < \infty$. Therefore, by contradiction our assumption that $\tau < \infty$ is false.

■

D.1. Showing gaps are uniformly bounded below over time with high probability

Lemma 22 *Let $\gamma(t) = (\gamma_1(t), \ldots, \gamma_d(t))$ be a strong solution to (4), with initial condition $\gamma(0) = (0, \cdots, 0)$. Then for any $t_0 \geq \frac{1}{d^{40}}$ and any $T > 0$ we have*

$$\mathbb{P}\left(\inf_{t_0 \leq t \leq T, 1 \leq i < d} \gamma_i(t) - \gamma_{i+1}(t) \leq \frac{1}{d^{10}} \frac{\sqrt{t}}{\mathfrak{b}\sqrt{d}}\right) \leq \frac{T}{d^{600}}, \tag{83}$$

for any $d \geq N_0$ where N_0 is a universal constant.

Proof By Weyl's inequality (Lemma 11), we have that for any $z \geq t_0$,

$$\begin{aligned}
&\mathbb{P}\left(\gamma_i(t) - \gamma_{i+1}(t) \leq \frac{1}{d^{10}} \frac{\sqrt{t}}{\mathfrak{b}\sqrt{d}} \qquad \text{for some } t \in \left[z, z + \frac{1}{d^{200}}\right], i \in [d-1]\right) \\
&\overset{\text{Lemma 11}}{\leq} \mathbb{P}\left(\gamma_i(z) - \gamma_{i+1}(z) \leq \frac{1}{d^{10}} \frac{\sqrt{t}}{\mathfrak{b}\sqrt{d}} + 2\|B(t)\|_2 \text{ for some } t \in \left[z, z + \frac{1}{d^{200}}\right], i \in [d-1]\right) \\
&\leq \mathbb{P}\left(\gamma_i(z) - \gamma_{i+1}(z) \leq \frac{1}{d^{10}} \frac{\sqrt{t}}{\mathfrak{b}\sqrt{d}} + 4\frac{1}{d^{200}}\sqrt{d} \qquad \text{for some } t \in \left[z, z + \frac{1}{d^{200}}\right], i \in [d-1]\right) \\
&\qquad + \mathbb{P}\left(\sup_{t \in [0, \frac{1}{d^{200}}]} \|B(t)\|_2 > 2\frac{1}{d^{200}}\sqrt{d}\right) \\
&\overset{\text{Lemma 12}}{\leq} \mathbb{P}\left(\gamma_i(z) - \gamma_{i+1}(z) \leq \frac{1}{d^{10}} \frac{\sqrt{t}}{\mathfrak{b}\sqrt{d}} + 4\frac{1}{d^{200}}\sqrt{d} \quad \text{for some } t \in \left[z, z + \frac{1}{d^{200}}\right], i \in [d-1]\right) \\
&\qquad + \frac{1}{d^{1000}} \\
&\leq \mathbb{P}\left(\gamma_i(z) - \gamma_{i+1}(z) \leq \frac{2}{d^{10}} \frac{\sqrt{z}}{\mathfrak{b}\sqrt{d}} \qquad \text{for some } i \in [d-1]\right) + \frac{1}{d^{1000}} \\
&\leq \sum_{i=1}^{d-1} \mathbb{P}\left(\gamma_i(z) - \gamma_{i+1}(z) \leq \frac{2}{d^{10}} \frac{\sqrt{z}}{\mathfrak{b}\sqrt{d}}\right) + \frac{1}{d^{1000}} \\
&\overset{\text{Lemma 4}}{\leq} \sum_{i=1}^{d-1} \left(\frac{2}{d^{10}}\right)^3 + \frac{1}{d^{1000}} \\
&\leq \frac{1}{d^{997}},
\end{aligned} \tag{84}$$

where the third Inequality holds by Lemma 12 whenever $d \geq N_0$ for some sufficiently large universal constant N_0, and the sixth inequality holds by Lemma 4. Thus, we have,

$$
\begin{aligned}
&\mathbb{P}\left(\inf_{t_0 \le t \le T, 1 \le i < d} \gamma_i(t) - \gamma_{i+1}(t) \le \frac{1}{d^{10}} \frac{\sqrt{t}}{\mathfrak{b}\sqrt{d}}\right) \\
&= \mathbb{P}\left(\gamma_i(t) - \gamma_{i+1}(t) \le \frac{1}{d^{10}} \frac{\sqrt{t}}{\mathfrak{b}\sqrt{d}} \qquad \text{for some } t \in [t_0, T], i \in [d-1]\right) \\
&\le \mathbb{P}\left(\bigcup_{z \in [t_0, T] \cap \frac{1}{d^{200}}\mathbb{Z}} \inf_{1 \le i < d} \gamma_i(t) - \gamma_{i+1}(t) \le \frac{1}{d^{10}} \frac{\sqrt{t}}{\mathfrak{b}\sqrt{d}} \qquad \forall t \in [z, z + \frac{1}{d^{200}}]\right) \\
&\le \sum_{z \in [t_0, T] \cap \frac{1}{d^{200}}\mathbb{Z}} \mathbb{P}\left(\inf_{1 \le i < d} \gamma_i(t) - \gamma_{i+1}(t) \le \frac{1}{d^{10}} \frac{\sqrt{t}}{\mathfrak{b}\sqrt{d}} \qquad \forall t \in [z, z + \frac{1}{d^{200}}]\right) \\
&\le \sum_{z \in [t_0, T] \cap \frac{1}{d^{200}}\mathbb{Z}} \frac{1}{d^{998}} \\
&\le d^{200} T \times \frac{1}{d^{998}} \\
&\le \frac{T}{d^{600}}.
\end{aligned}
$$

■

D.2. Gaps between not necessarily neighboring eigenvalues

Proposition 23 *Suppose that $X_1, \ldots, X_r$ are (not necessarily independent) random variables satisfying $\mathbb{P}(X_i \le s) \le F(s)$ for all $i \in [r]$, where $F : \mathbb{R} \to \mathbb{R}$ is some nondecreasing function. Then*

$$\mathbb{P}\left(\sum_{i=1}^{r} X_i \le \frac{1}{2} rs\right) \le 2F(s). \tag{85}$$

Proof Let E be the “bad” event that $|\{i : X_i \le s\}| \ge \frac{r}{2}$. Choose J uniformly at random from $\{1, \ldots, r\}$. Then $\mathbb{P}(J \in \{i : X_i \le s\} | E) \ge \frac{1}{2}$. Therefore,

$$
\begin{aligned}
\mathbb{P}(X_j \le s) &= \mathbb{P}(X_j \le s | E) \times \mathbb{P}(E) \\
&= \mathbb{P}(J \in \{i : X_i \le s\} | E) \times \mathbb{P}(E) \\
&\ge \frac{1}{2} \mathbb{P}(E).
\end{aligned}
$$

Thus, $\mathbb{P}(E) \le 2\mathbb{P}(X_j \le s)$ But $\{\sum_{i=1}^r X_i \le \frac{1}{2} rs\} \subseteq E$. Therefore,

$$\mathbb{P}\left(\sum_{i=1}^{r} X_i \le \frac{1}{2} rs\right) \le \mathbb{P}(E) \le 2\mathbb{P}(X_J \le s) \le 2F(s),$$

where the last inequality holds since $\mathbb{P}(X_i \le s) \le F(s)$ for all $i \in [r]$. ■

Corollary 24 (Gaps between not-necessarily neighboring eigenvalues)
Let $\gamma(t) = (\gamma_1(d), \ldots, \gamma_d(t))$ be a strong solution of (4) starting from any initial $\gamma(0) \in \mathcal{W}_d$. Then for every $i, j \in [d]$ and every $\alpha > 0$,

$$\mathbb{P}\left(\left\{\gamma_i(t) - \gamma_j(t) \leq (j-i) \times s\frac{\sqrt{t}}{\mathfrak{b}\sqrt{d}}\right\} \cap \hat{E}_\alpha^c\right) \leq s^3 \qquad \forall s > 0, t > 0.$$

Proof Let $\xi(t) = (\xi_1(d), \ldots, \xi_d(t))$ be a strong solution of (4) starting from $\xi(0) = (0, \cdots, 0)$. Further, let $\tilde{E}$ be the event that $\inf_{t_0 \leq t \leq T, 1 \leq i < d} \xi_i(t) - \xi_{i+1}(t) \leq \frac{1}{d^{10}} \frac{\sqrt{t}}{\mathfrak{b}\sqrt{d}}$.

Since $\xi(t) = (\xi_1(t), \ldots, \xi_d(t))$ have the same joint distribution as the eigenvalues of $\sqrt{t}(G + G^*)$ where G is a $d \times d$ matrix with i.i.d. complex standard Gaussian entries, by Lemma 4 we have that,

$$\mathbb{P}\left(\left\{\xi_i(t) - \xi_{i+1}(t) \leq s\frac{\sqrt{t}}{\mathfrak{b}\sqrt{d}}\right\} \cap \tilde{E}^c\right) \leq 2s^3 \qquad \forall s > 0, \forall 1 \leq i < d. \tag{86}$$

Define $X_\ell := \xi_{i+\ell}(t) - \xi_{i+\ell+1}(t)$ for all $\ell \in \{1, \ldots, j-i\}$. Then using the fact that by Lemma 3 all the eigenvalue gaps of $\gamma(t)$ are at least as large as the corresponding gaps of $\xi(t)$, and plugging (86) into Proposition 23, we have that

$$\begin{aligned}
&\mathbb{P}\left(\left\{\gamma_i(t) - \gamma_j(t) \leq (j-i) \times s\frac{\sqrt{t}}{2\mathfrak{b}\sqrt{d}}\right\} \cap \hat{E}_\alpha^c\right) \\
&\overset{\text{Lemmas 3}}{\leq} \mathbb{P}\left(\left\{\xi_i(t) - \xi_j(t) \leq (j-i) \times s\frac{\sqrt{t}}{2\mathfrak{b}\sqrt{d}}\right\} \cap \tilde{E}^c\right) \\
&= \mathbb{P}\left(\left\{\sum_{\ell=1}^{j-i} X_j \leq \frac{1}{2}(j-i) \times s\frac{\sqrt{t}}{\mathfrak{b}\sqrt{d}}\right\} \cap \tilde{E}^c\right) \\
&\overset{\text{Lemma 23, Eq. (86)}}{\leq} 4s^3.
\end{aligned}$$

Redefining $\mathfrak{b}$ to be 4 times the original value of $\mathfrak{b}$ completes the proof. ■

Appendix E. Eigenvalue gaps of Gaussian Unitary Ensemble: Proof of Lemma 4

E.1. Eigenvalue ridgidity

Denote by $\eta_1, \ldots, \eta_d$ the eigenvalues of the GUE random matrix– that is the matrix $G + G^*$ where each entry of G is an independent standard complex Gaussian. The eigenvalue gaps of the GUE satisfy a rigidity property ((Erdős et al., 2012); restated here as Lemma 26). Roughly, for every $i \in [d]$ the i'th eigenvalue η_i does not deviate by more than $\mathrm{polylog}(d)$ times the average gap size $\eta_i - \eta_{i+1}$. More formally, for every $i \in [d]$ we define the "classical" eigenvalue location ω_i to be the number such that

$$d\int_{\frac{\omega_i}{\sqrt{d}}}^{\infty} \rho(x)\mathrm{d}x = i - 1, \tag{87}$$

where $\rho(x) := \frac{1}{2\pi}\sqrt{\max(4 - x^2,\ 0)}$ is the semi-circle law. For convenience, we also define $\omega_{d+1} := -2\sqrt{d}$ (that way, the locations of the $\omega_{d+1} \leq \omega_d \leq \cdots \leq \omega_1$ are symmetric about 0).

Proposition 25 *The classical eigenvalues ω_i satisfy*

$$2\sqrt{d} - 3d^{-\frac{1}{6}}(i-1)^{\frac{2}{3}} \le \omega_i \le 2\sqrt{d} - d^{-\frac{1}{6}}(i-1)^{\frac{2}{3}} \qquad \forall\, 1 \le i \le \frac{d}{2}, \tag{88}$$

$$d^{-\frac{1}{6}}(d-i+1)^{\frac{2}{3}} - 2\sqrt{d} \le \omega_i \le 3d^{-\frac{1}{6}}(d-i+1)^{\frac{2}{3}} - 2\sqrt{d} \qquad \forall\, \frac{d}{2} \le i \le d. \tag{89}$$

Moreover, their gaps satisfy

$$d^{-\frac{1}{6}}\min(i, d-i+1)^{-\frac{1}{3}} \le \omega_i - \omega_{i+1} \le 2\pi d^{-\frac{1}{6}}\min(i, d-i+1)^{-\frac{1}{3}} \qquad \forall\, 1 \le i \le d, \tag{90}$$

Proof For every $x \in [-2, 0]$, we have

$$\frac{1}{2\pi}\sqrt{x+2} \le \rho(x) \le \frac{1}{2\pi}2\sqrt{x+2} \tag{91}$$

Furthermore, since $\rho(x)$ is symmetric about 0, for every $x \in [0, 2]$, we have

$$\frac{1}{2\pi}\sqrt{2-x} \le \rho(x) \le \frac{1}{2\pi}2\sqrt{2-x}. \tag{92}$$

Thus, (91) implies that,

$$\int_{-2}^{x} \rho(s)\mathrm{d}s \ge \int_{-2}^{x} \frac{1}{2\pi}\sqrt{x+2}\mathrm{d}s = \frac{3}{2}(x+2)^{\frac{3}{2}}$$

and that

$$\int_{-2}^{x} \rho(s)\mathrm{d}s \le 2\int_{-2}^{x} \frac{1}{2\pi}\sqrt{x+2}\mathrm{d}s = 2(x+2)^{\frac{3}{2}}.$$

Then for every $i \in [d]$ we have

$$\int_{-2}^{d^{-\frac{2}{3}}i^{\frac{2}{3}}-2} \rho(x)\mathrm{d}x \le \frac{1}{2\pi}2(d^{-\frac{2}{3}}i^{\frac{2}{3}} - 2)^{1.5} \le \frac{i}{d}, \tag{93}$$

and

$$\int_{-2}^{3d^{-\frac{2}{3}}i^{\frac{2}{3}}-2} \rho(x)\mathrm{d}x \ge \frac{1}{2\pi}\frac{3}{2}(3d^{-\frac{2}{3}}i^{\frac{2}{3}})^{1.5} \ge \frac{i}{d}. \tag{94}$$

Since $\rho(x)$ is nonnegative, $\int_{-2}^{x}\rho(s)\mathrm{d}s$ is nondecreasing in x. Therefore, from (93) and (94), we have by the definition of ω_i (Equation (87)) that

$$d^{-\frac{2}{3}}(d-i+1)^{\frac{2}{3}} - 2 \le \frac{\omega_i}{\sqrt{d}} \le 3d^{-\frac{2}{3}}(d-i+1)^{\frac{2}{3}} - 2 \qquad \forall 1 \le i \le d+1, \tag{95}$$

which proves (90). Moreover, since the density $\rho(x)$ is symmetric about 0, (95) implies that

$$2 - 3d^{-\frac{2}{3}}(i-1)^{\frac{2}{3}} \le \frac{\omega_i}{\sqrt{d}} \le 2 - d^{-\frac{2}{3}}(i-1)^{\frac{2}{3}} \qquad \forall 1 \le i \le d+1, \tag{96}$$

which proves (89). Since $\rho(x)$ is nonincreasing on [0,2] we also have that for all $2 \le i \le \frac{d}{2}+1$,

$$\frac{\omega_i}{\sqrt{d}} - \frac{\omega_{i+1}}{\sqrt{d}} \le \frac{1}{d \times \rho(\omega_i)} \overset{\text{Eq. (92)}}{\le} \frac{1}{d\frac{1}{2\pi}\sqrt{2-\omega_i}} \overset{\text{Eq. (96)}}{\le} \frac{2\pi}{d\sqrt{d^{-\frac{2}{3}}(i-1)^{\frac{2}{3}}}} \le 4\pi d^{-\frac{2}{3}}i^{-\frac{1}{3}}. \tag{97}$$

and that, for all $1 \leq i \leq \frac{d}{2}+1$,

$$\frac{\omega_i}{\sqrt{d}} - \frac{\omega_{i+1}}{\sqrt{d}} \geq \frac{1}{d \times \rho(\omega_{i+1})} \overset{\text{Eq. (92)}}{\geq} \frac{1}{2d\frac{1}{2\pi}\sqrt{2-\omega_{i+1}}} \overset{\text{Eq. (96)}}{\geq} \frac{1}{2d\frac{1}{2\pi}\sqrt{3d^{-\frac{2}{3}}i^{\frac{2}{3}}}} \geq \frac{\pi}{\sqrt{3}}d^{-\frac{2}{3}}i^{-\frac{1}{3}}. \quad (98)$$

Therefore,

$$\frac{\pi}{\sqrt{3}}d^{-\frac{1}{6}}i^{-\frac{1}{3}} \overset{\text{Eq. (98)}}{\leq} \omega_i - \omega_{i+1} \overset{\text{Eq. (97)}}{\leq} 4\pi d^{-\frac{1}{6}}i^{-\frac{1}{3}} \qquad \forall 2 \leq i \leq \frac{d}{2}+1. \quad (99)$$

Moreover, plugging in $i=2$ to (88) plugging in the fact that $\omega_1 = 2\sqrt{d}$, we have that

$$d^{-\frac{1}{6}} \leq \omega_1 - \omega_2 \leq 3d^{-\frac{1}{6}}. \quad (100)$$

Therefore (99) and (100) together imply that,

$$\frac{\pi}{\sqrt{3}}d^{-\frac{1}{6}}i^{-\frac{1}{3}} \overset{\text{Eq. (98)}}{\leq} \omega_i - \omega_{i+1} \overset{\text{Eq. (97)}}{\leq} 4\pi d^{-\frac{1}{6}}i^{-\frac{1}{3}} \qquad \forall 1 \leq i \leq \frac{d}{2}+1. \quad (101)$$

Finally, since the density $\rho(x)$ is symmetric about 0, (101) implies that

$$d^{-\frac{1}{6}}\min(i, d-i+1)^{-\frac{1}{3}} \leq \omega_i - \omega_{i+1} \leq 2\pi d^{-\frac{1}{6}}\min(i, d-i+1)^{-\frac{1}{3}} \qquad \forall 1 \leq i \leq d,$$

which proves (90). ■

Lemma 26 (Eigenvalue rigidity of GUE (Theorem 2.2 of Erdős et al. (2012))) *There exist universal constants $C \geq 1$ and $c_1, c_2, N_0 > 0$ such that for every $L \in \left[c_1, \frac{\log(10d)}{10(\log\log d)^2}\right]$ and every $d \geq N_0$,*

$$\mathbb{P}\left(\exists j \in [d] : |\eta_j - \omega_j| \geq (\log d)^{L\log\log d}\min(j, d-j+1)^{-\frac{1}{3}}d^{-\frac{1}{6}}\right) \leq C\exp[-(\log d)^{c_2 L\log\log d}].$$

E.2. Bounding the eigenvalue gaps of the GUE matrix

In this section, we prove high-probability bounds for the eigenvalue gaps of the GUE random matrix (Lemma 4).

Step 1. Define the rigidity event E and prove that it holds with high probability (Use Lemma 26). Set $L := \max(\frac{2}{c_2}\log\log(C), c_1, 1)$; thus, L is a universal constant. Define the event E as follows:

$$E := \left\{\exists j \in [d] : |\eta_j - \omega_j| \geq (\log d)^{L\log\log d}\min(j, d-j+1)^{-\frac{1}{3}}d^{-\frac{1}{6}})\right\}^c.$$

Then (replacing the universal constant N_0 with a universal constant such that $\max(\frac{2}{c_2}\log\log(C), c_1) \leq \frac{\log(10d)}{10(\log\log d)^2}$ and $N_0 \geq e^4$), we have by Lemma 26 that

$$\mathbb{P}(E^c) \leq C\exp[-(\log d)^{c_2 L\log\log d}] \leq \exp[-(\log d)^{2\log\log d}] \leq \exp[-(\log d)^2] \leq d^{-\log d} \leq \frac{1}{d^{1000}}, \quad (102)$$

for all $d \geq N_0$, where N_0 is a universal constant. Define $\mathfrak{b} := 10^6(\log d)^{L\log\log d}$. Further, define $\omega_j = \eta_j = +\infty$ for all $j < d$ and $\omega_j = \eta_j = -\infty$ for all $j > d$.

Step 2. Consider any $\eta \in \mathcal{W}_d$ such that the event E holds. Define $j_{\min} := \max(i - \mathfrak{b}^2, 1)$ and $j_{\max} := \min(i + \mathfrak{b}^2, d)$. Define the following quantities:

- $a_{\min} := \omega_{j_{\min}} - \frac{1}{30}\mathfrak{b}^2 d^{-\frac{1}{6}} \min(i, d-i)^{-\frac{1}{3}}$
- $a_{\max} := \omega_{j_{\min}} + \frac{1}{30}\mathfrak{b}^2 d^{-\frac{1}{6}} \min(i, d-i)^{-\frac{1}{3}}$
- $b_{\min} := \omega_{j_{\max}} - \frac{1}{30}\mathfrak{b}^2 d^{-\frac{1}{6}} \min(i, d-i)^{-\frac{1}{3}}$
- $b_{\max} := \omega_{j_{\max}} + \frac{1}{30}\mathfrak{b}^2 d^{-\frac{1}{6}} \min(i, d-i)^{-\frac{1}{3}}$.

Proposition 27 *Suppose that the event E occurs. Then for all $\mathfrak{b}^2 \le i \le d - \mathfrak{b}^2$ we have*

$$\eta_{i-\mathfrak{b}^2} - \eta_{i+\mathfrak{b}^2} \ge \frac{29}{30}\mathfrak{b}^2 d^{-\frac{1}{6}} \min(i, d-i)^{-\frac{1}{3}} \ge \frac{29}{30}\mathfrak{b}^2 \frac{1}{\sqrt{d}}. \tag{103}$$

and $\eta_{j_{\max}} \in [a_{\min}, a_{\max}]$ and $\eta_{j_{\min}} \in [b_{\min}, b_{\max}]$.

Proof Without loss of generality, we may assume that $i > \mathfrak{b}^2$ (since otherwise we have $\eta_{i-\mathfrak{b}^2} - \eta_{i+\mathfrak{b}^2} \ge \infty$), and that $i \le \frac{1}{2}$ (since the GUE matrix G and $-G$ have the same distribution and hence the eigenvalue distribution of the GUE is symmetric about 0).

If E occurs, then by the definition of the event E we have

$$\begin{aligned}
\eta_{i-\mathfrak{b}^2} - \eta_{i+\mathfrak{b}^2} &\ge \omega_{i-\mathfrak{b}^2} - \omega_{i+\mathfrak{b}^2} - 2\mathfrak{b}(i - \mathfrak{b}^2)^{-\frac{1}{3}} d^{-\frac{1}{6}} \\
&\overset{\text{Prop. 25}}{\ge} \mathfrak{b}^2 \times d^{-\frac{1}{6}} i^{-\frac{1}{3}} - 2\mathfrak{b}(i - \mathfrak{b}^2)^{-\frac{1}{3}} d^{-\frac{1}{6}} \\
&\ge \mathfrak{b}^2 \times d^{-\frac{1}{6}} i^{-\frac{1}{3}} - 2\mathfrak{b}\left(\frac{i}{2\mathfrak{b}^2}\right)^{-\frac{1}{3}} d^{-\frac{1}{6}} \\
&\ge \mathfrak{b}^2 \times d^{-\frac{1}{6}} i^{-\frac{1}{3}} - 2^{\frac{4}{3}}\mathfrak{b}^{\frac{5}{3}} i^{-\frac{1}{3}} d^{-\frac{1}{6}} \\
&\ge \frac{29}{30}\mathfrak{b}^2 d^{-\frac{1}{6}} i^{-\frac{1}{3}},
\end{aligned}$$

where the third inequality holds since $\frac{i}{2\mathfrak{b}^2} \le i - \mathfrak{b}^2$ because $i \ge \mathfrak{b}^2 + 1 > 4$, and the last inequality holds since $\mathfrak{b} \ge 10^6$. This proves (103).

Moreover, by the definition of the event E, we also have that

$$\begin{aligned}
|\eta_{i-\mathfrak{b}^2} - \omega_{i-\mathfrak{b}^2}| &\le \mathfrak{b}(i - \mathfrak{b}^2)^{-\frac{1}{3}} d^{-\frac{1}{6}} \\
&\le \mathfrak{b}\left(\frac{i}{2\mathfrak{b}^2}\right)^{-\frac{1}{3}} d^{-\frac{1}{6}} \\
&\le 2^{\frac{1}{3}}\mathfrak{b}^{\frac{5}{3}} i^{-\frac{1}{3}} d^{-\frac{1}{6}} \\
&\le \frac{1}{30}\mathfrak{b}^2 d^{-\frac{1}{6}} i^{-\frac{1}{3}}
\end{aligned} \tag{104}$$

where the second inequality holds since $\frac{i}{2\mathfrak{b}^2} \le i - \mathfrak{b}^2$ because $i \ge \mathfrak{b}^2 + 1 > 4$, and the last inequality holds since $\mathfrak{b} \ge 10^6$. Thus, (104) implies that $\eta_{j_{\max}} \in [a_{\min}, a_{\max}]$.

Again, by the definition of the event E, we also have that

$$
\begin{aligned}
|\eta_{i+\mathfrak{b}^2} - \omega_{i+\mathfrak{b}^2}| &\leq \mathfrak{b}(i+\mathfrak{b}^2)^{-\frac{1}{3}} d^{-\frac{1}{6}} \\
&\leq \mathfrak{b} i^{-\frac{1}{3}} d^{-\frac{1}{6}} \\
&\leq \frac{1}{30}\mathfrak{b}^2 d^{-\frac{1}{6}} i^{-\frac{1}{3}} \qquad (105)
\end{aligned}
$$

where the last inequality holds since $\mathfrak{b} \geq 10^6$. Thus, (105) implies that $\eta_{j_{\min}} \in [b_{\min}, b_{\max}]$. ■

Then, whenever the event E occurs, we have that $\eta_{j_{\max}} \in [a_{\min}, a_{\max}]$ and $\eta_{j_{\min}} \in [b_{\min}, b_{\max}]$. Consider any a, b such that $a_{\min} \leq a \leq a_{\max}$ and $b_{\min} \leq b \leq b_{\max}$. Define the sets

- $S_0(a,b) := \cap\{\eta \in \mathbb{R}^d : \eta_{j_{\max}} = a, \eta_{j_{\min}} = b\}$,
- $S_3(a,b;y) := \{\eta \in \mathbb{R}^d : \eta_i - \eta_{i+1} = y\} \cap S_0(a,b)$ for any $y \leq s\frac{1}{8\mathfrak{b}^4\sqrt{d}}$, and
- $S_4(a,b) := \{\eta \in \mathbb{R}^d : \eta_i - \eta_{i+1} \geq s\} \cap S_0(a,b)$.

Step 3. Define an injective map from the "bad" set S_3 to the good set S_4, which, roughly speaking, shows that the good set has a much bigger volume and a much larger probability density than the bad set. For any $y \leq s\frac{1}{\mathfrak{b}\sqrt{d}}$, we want to define an injective map $g : S_2(a,b;y) \to S_4(a,b)$, such that its Jacobian $J_g(\eta)$ satisfies $\det(J_g(\eta)) \geq \frac{1}{2}$ for all $\eta \in \mathbb{R}^d$, and

$$\frac{f(\eta)}{f(g(\eta))} \leq (\mathfrak{b}\sqrt{d})^2 \times y^2, \qquad (106)$$

for any $\eta \in S_3(a,b;y)$. Towards this end, we consider the map $g : \mathcal{W}_d \to \mathcal{W}_d$ such that

- $$g(\eta)[j] = \eta_j \qquad \forall j \notin [j_{\max}, j_{\min}] \qquad (107)$$
- $$g(\eta)[j_{\max}] = \eta_{j_{\max}} = a, \qquad (108)$$
- $$g(\eta)[j] = g(\eta)[j+1] + (1-\alpha) \times (\eta_j - \eta_{j+1}) \qquad \forall j \in [j_{\max}, j_{\min}]\backslash\{i\} \qquad (109)$$
- $$g(\eta)[i] = g(\eta)[i+1] + \left(\frac{2}{s}(\eta_i - \eta_{i+1}) + 2\frac{1}{8\mathfrak{b}^4\sqrt{d}}\right) \times \frac{b-a-(\eta_i-\eta_{i+1})}{b-a}, \qquad (110)$$

where $\alpha := \frac{\frac{2}{s}(\eta_i - \eta_{i+1}) + 2\frac{1}{8\mathfrak{b}^4\sqrt{d}}}{b-a}$.

Proposition 28 *Suppose that $\mathfrak{b}^2 \leq i \leq d - \mathfrak{b}^2$. Then the following properties hold for g:*

- *g is injective.*
- $g(\eta)[j_{\min}] = \eta_{j_{\min}} = b$,

- $g(\eta)[i] - g(\eta)[i+1] \geq \frac{1}{8\mathfrak{b}^4\sqrt{d}}$, *and hence*

$$\frac{\eta_i - \eta_{i+1}}{g(\eta)[i] - g(\eta)[i+1]} \leq 8\mathfrak{b}^4\sqrt{d} \times (\eta_i - \eta_{i+1}) = 8\mathfrak{b}^4\sqrt{d} \times y \tag{111}$$

for any $\eta \in S_3(a, b; y)$ *and any* $y \leq s\frac{1}{8\mathfrak{b}^4\sqrt{d}}$.

- $$g(\eta)[j] - g(\eta)[j+1] \geq (1-\alpha)(\eta_j - \eta_{j+1}) \qquad \forall j \in [d]. \tag{112}$$

Proof
Injectivity: To prove that g is injective, we note that, given any vector $z \in \mathbb{R}^d$ we can solve for the *unique* $\eta \in \mathcal{W}_d$ such that $g(\eta) = z$ whenever such a value of η exits.

First, we solve for $\eta_i - \eta_{i+1}$ by solving the quadratic equation (110) for $\eta_i - \eta_{i+1}$, and noting that $\eta_i - \eta_{i+1} \geq 0$ (and hence we have a unique solution for $\eta_i - \eta_{i+1}$, whenever this solution exists). This gives us the value of $\eta_i - \eta_{i+1}$ in terms of $g(\eta)[i] - g(\eta)[i+1]$.

Next, we use plug in the value of $\eta_i - \eta_{i+1}$ to compute $\alpha = \frac{\frac{2}{s}(\eta_i - \eta_{i+1}) + 2\frac{1}{\mathfrak{b}\sqrt{d}}}{b-a}$, and for every $j \in [j_{\max}, j_{\min}]$, plug in this value of α to (109), to solve for $\eta_j - \eta_{j+1}$ in terms of $g(\eta)[j] - g(\eta)[j+1]$.

Finally, $\eta_{j_{\max}} = a$ we can compute each $\eta_j = a + \sum_{\ell=j}^{j_{\max}-1} \eta_\ell - \eta_{\ell+1}$ for each $j \in [j_{\max} + 1, j_{\min}]$. Thus, given any vector $z \in \mathbb{R}^d$ we can solve for the unique $\eta \in \mathbb{R}^d$ such that $g(\eta) = z$ whenever such a value of η exits. Therefore, g is injective.

Showing that $g(\eta)[j_{\min}] = b$*:*

$$\begin{aligned}
g(\eta)[j_{\min}] &= a + \sum_{\ell=j_{\min}}^{j_{\max-1}} g(\eta)[\ell] - g(\eta)[\ell+1] \\
&= a + \left(\frac{2}{s}(\eta_i - \eta_{i+1}) + 2\frac{1}{8\mathfrak{b}^4\sqrt{d}}\right) \times \frac{b - a - (\eta_i - \eta_{i+1})}{b-a} \\
&\quad + \sum_{\ell \in [j_{\min}, j_{\max}] \setminus \{i\}} (1-\alpha) \times (\eta_j - \eta_{j+1}) \\
&= b.
\end{aligned}$$

Showing (111)*:* Since $y \leq s\frac{1}{8\mathfrak{b}^4\sqrt{d}}$ and $\eta \in S_3(a, b; y)$, we have that $\eta_i - \eta_{i+1} = y \leq s\frac{1}{8\mathfrak{b}^4\sqrt{d}}$. Thus by (110),

$$\begin{aligned}
g(\eta)[i] - g(\eta)[i+1] &= \left(\frac{2}{s}(\eta_i - \eta_{i+1}) + 2\frac{1}{8\mathfrak{b}^4\sqrt{d}}\right) \times \frac{b - a - (\eta_i - \eta_{i+1})}{b-a} \\
&\geq 2\frac{1}{8\mathfrak{b}^4\sqrt{d}} \times \frac{1}{2} \\
&= \frac{1}{8\mathfrak{b}^4\sqrt{d}}.
\end{aligned}$$

Hence,

$$\frac{\eta_i - \eta_{i+1}}{g(\eta)[i] - g(\eta)[i+1]} \leq 8\mathfrak{b}^4\sqrt{d} \times (\eta_i - \eta_{i+1}) = 8\mathfrak{b}^4\sqrt{d} \times y, \tag{113}$$

which proves (111).

Showing (112)*:* By (110), we have

$$\begin{aligned} g(\eta)[i] - g(\eta)[i+1] &= \left(\frac{2}{s}(\eta_i - \eta_{i+1}) + 2\frac{1}{\mathfrak{b}\sqrt{d}}\right) \times \frac{b - a - (\eta_i - \eta_{i+1})}{b-a} \\ &\geq \left(\frac{2}{s}(\eta_i - \eta_{i+1}) + 2\frac{1}{\mathfrak{b}\sqrt{d}}\right) \times \frac{1}{2} \\ &\geq \frac{1}{s}(\eta_i - \eta_{i+1}) \\ &\geq (\eta_i - \eta_{i+1}) \\ &\geq (1-\alpha) \times (\eta_i - \eta_{i+1}), \end{aligned}$$

where the second-to-last inequality holds since $s \leq 1$. Thus, (112) holds for $j = i$. Moreover, (112) holds for all $j \in [j_{\max}, j_{\min}] \backslash \{i\}$ by (109) and (112) holds for all $j \notin [j_{\max}, j_{\min}]$ by (107). Therefore (112) holds for all $j \in [d]$. ■

Step 4. Bounding the Jacobian determinant of the map g.

Proposition 29 (Jacobian determinant of g) *If $y \leq s\frac{1}{8\mathfrak{b}^4\sqrt{d}}$ and $\eta \in S_3(a, b; y)$, we have that*

$$\det(J_g(\eta)) \geq \frac{1}{2s}.$$

Proof Consider the map $h : \mathcal{W}_d \to \mathbb{R}^d$, where $h(\eta)[j] = \eta_j - \eta_{j+1}$ for $j \in [j_{\min}, j_{\max} - 1]$ and $h(\eta)[j] = \eta_j$ for $j \in [1, d] \backslash [j_{\min}, j_{\max} - 1]$. Then for every $\Delta \in \mathcal{W}_{j_{\max} - j_{\min}}$, by (107)-(110), we have that

$$\frac{\partial g \circ h^{-1}(\Delta)[\ell]}{\partial \Delta_j} = 0 \qquad \forall \ell \neq j, j \neq i \tag{114}$$

$$\frac{\partial g \circ h^{-1}(\Delta)[\ell]}{\partial \Delta_j} = (1-\alpha) \qquad \forall \ell = j \neq i, \quad j \in [j_{\min}, j_{\max}]$$

$$\begin{aligned} \frac{\partial g \circ h^{-1}(\Delta)[i]}{\partial \Delta_i} &= \frac{2}{s} - 2\frac{1}{\mathfrak{b}\sqrt{d}(b-a)} - \frac{1}{s(b-a)}(\eta_i - \eta_{i+1}) \\ &\geq \frac{1}{2s}, \end{aligned}$$

$$\frac{\partial g \circ h^{-1}(\Delta)[j]}{\partial \Delta_j} = 1 \qquad \forall j \notin [j_{\min}, j_{\max}].$$

where the inequality holds since $s \leq \frac{1}{\mathfrak{b}}$, $\mathfrak{b} < 1$, and $(\eta_i - \eta_{i+1}) \leq b - a$. Moreover, since $\eta_i - \eta_{i+1} \leq s\frac{1}{8\mathfrak{b}^4\sqrt{d}}$ because $y \leq s\frac{1}{8\mathfrak{b}^4\sqrt{d}}$ and $\eta \in S_3(a, b; y)$, we also have that $\alpha = \frac{\frac{2}{s}(\eta_i - \eta_{i+1}) + 2\frac{1}{8\mathfrak{b}^4\sqrt{d}}}{b-a} \leq \frac{1}{(b-a)\mathfrak{b}^4\sqrt{d}}$. Thus, $J_{g \circ h^{-1}}(\Delta)$ has diagonal entries $1 - \alpha \geq 1 - \frac{1}{(b-a)\mathfrak{b}^2\sqrt{d}}$ for $j \in [j_{\min}, j_{\max} - 1] \backslash \{i\}$, and i'th entry $\geq \frac{1}{2s}$, and all other diagonal entries equal to $= 1$. Moreover, if one exchanges the i'th row and column of $J_{g \circ h^{-1}}(\Delta)$ with its first row and column, by (114) the resulting matrix is a $d \times d$

upper triangular matrix with the same determinant as $J_{g\circ h^{-1}}(\Delta)$. Thus, by Sylvester's formula, the determinant of $J_{g\circ h^{-1}}(\Delta)$ is equal to the product of its diagonal entries, and hence

$$\begin{aligned}\det(J_{g\circ h^{-1}}(\Delta)) &\geq \frac{1}{2s}\left(1-\frac{1}{(b-a)\mathfrak{b}^4\sqrt{d}}\right)^{j_{\max}-j_{\min}-2} \times 1\\ &\geq \left(1-\frac{1}{(b-a)\mathfrak{b}^4\sqrt{d}}\right)^{2\mathfrak{b}^2}\\ &\geq \left(1-\frac{1}{(b-a)\mathfrak{b}^4\sqrt{d}}\right)^{2\mathfrak{b}^2}\\ &\geq \left(1-\frac{1}{\mathfrak{b}^2}\right)^{2\mathfrak{b}^2}\\ &\geq \frac{1}{16s},\end{aligned} \tag{115}$$

where the fourth inequality holds by Proposition (25) we have that $b-a \geq \frac{\mathfrak{b}^2}{\sqrt{d}}$.

Hence,

$$\begin{aligned}\det(J_g(\eta)) = \det(J_{g\circ h^{-1}\circ h}(\eta)) &= \det(J_{g\circ h^{-1}}(h(\eta)) \times J_h(\eta))\\ &= \det(J_{g\circ h^{-1}}(h(\eta)) \times \det(J_h(\eta))\\ &= \det(J_{g\circ h^{-1}}(h(\eta)) \times 1\\ &\overset{\text{Eq. (115)}}{\geq} \frac{1}{16s},\end{aligned}$$

where the third equality holds since $J_h(\eta)$ is the bidiagonal matrix with diagonal entries 1 and entries -1 above the diagonal, and this matrix has determinant 1. ■

Step 5. This step is a mean-field approximation for far-away eigenvalues.

Lemma 30 (Mean field approximation for far-away eigenvalues)

$$\prod_{j\in[j_{\min},j_{\max}],\ \ell\notin[j_{\min}-2\mathfrak{b},j_{\max}+2\mathfrak{b}]} \frac{|\eta_j-\eta_\ell|^2}{|g(\eta)[j]-g(\eta)[\ell]|^2} \leq 2, \tag{116}$$

Proof Consider any $j \in [j_{\min}, j_{\max}]$ and any $r \geq 2\mathfrak{b}$. Then, if $\eta \in E$, by the definition of the event E we have that

$$|\eta_j - \eta_{j+r}| = r\frac{1}{2\sqrt{d}} + \rho \tag{117}$$

for some $\rho \geq 0$. Moreover, from (109) and (110) that

$$\begin{aligned}|(g(\eta)[j]-g(\eta)[j+r]) - (\eta_j-\eta_{j+r})| &\leq \alpha(b-a)\\ &= \frac{2}{s}(\eta_i-\eta_{i+1}) + 2\frac{1}{8\mathfrak{b}^4\sqrt{d}}\\ &\leq \frac{1}{2\mathfrak{b}^4\sqrt{d}}\end{aligned} \tag{118}$$

where the last inequality holds since $y \leq s\frac{1}{8\mathfrak{b}^4\sqrt{d}}$. Thus, by (117) and (118) we have that for some $\zeta \in \mathbb{R}$ where $|\zeta| \leq \frac{1}{2\mathfrak{b}^4\sqrt{d}}$, we have

$$\begin{aligned}\frac{|\eta_j - \eta_{j+r}|^2}{|g(\eta)[j] - g(\eta)[j+r]|^2} &\leq \frac{|r\frac{1}{2\sqrt{d}} + \rho|^2}{|r\frac{1}{2\sqrt{d}} + \rho + \zeta|^2} \\ &\leq \left(1 + \frac{1}{r} \times \frac{1}{\mathfrak{b}^4}\right)^2 \\ &= \left(1 + \frac{1}{r} \times \frac{1}{\mathfrak{b}^4}\right)^2, \end{aligned} \tag{119}$$

where the second inequality holds since $|\zeta| \leq \frac{1}{4\mathfrak{b}^4\sqrt{d}}$ and $\rho \geq 0$. Therefore, we have

$$\begin{aligned}&\prod_{j\in[j_{\min},j_{\max}],\ \ell\notin[j_{\min}-2\mathfrak{b},j_{\max}+2\mathfrak{b}]} \frac{|\eta_j - \eta_\ell|^2}{|g(\eta)[j] - g(\eta)[\ell]|^2} \\ &\overset{\text{Eq. (119)}}{\leq} \prod_{\ell\in[j_{\min},j_{\max}]} \prod_{r=2\mathfrak{b}}^{d} \left(1 + \frac{1}{r} \times \frac{1}{\mathfrak{b}^4}\right)^2 \\ &\leq (e^{\frac{1}{\mathfrak{b}^3}})^{j_{\max}-j_{\min}} \\ &= (e^{\frac{1}{\mathfrak{b}^3}})^{2\mathfrak{b}^2} \\ &= e^{\frac{1}{\mathfrak{b}}} \\ &\geq 2,\end{aligned}$$

where the second inequality holds since $\prod_{r=1}^{d}(1+\frac{1}{r}) = d+1$ and hence that $\prod_{r=2}^{d}(1+\frac{1}{\kappa r}) \geq (d+1)^{\frac{1}{\kappa}}$ for every $\kappa \geq 1$. Plugging in $\kappa = \mathfrak{b}^4$, we have $\prod_{r=2}^{n}(1+\frac{1}{\kappa r}) \geq (d+1)^{-\frac{1}{\mathfrak{b}^4}} \geq \left(d^{-\frac{1}{\log(d)^{\log\log d}}}\right)^{\frac{1}{\mathfrak{b}^3}} = e^{-\frac{1}{\mathfrak{b}^3}}$. ■

Step 6. Bounding the density ratio to show that $\frac{f(\eta)}{f(g(\eta))} \leq \tilde{O}((\sqrt{d}\log d)^2 \times y^2)$.

Lemma 31 *For any $y \leq s\frac{1}{8\mathfrak{b}^4\sqrt{d}}$ and any $\eta \in S_3(a,b;y)$, we have that*

$$\frac{f(\eta)}{f(g(\eta))} \leq 50(8\mathfrak{b}^4\sqrt{d})^2 \times y^2,$$

Proof Since $\eta_i - \eta_{i+1} \leq s\frac{1}{8\mathfrak{b}^4\sqrt{d}}$ because $y \leq s\frac{1}{8\mathfrak{b}^4\sqrt{d}}$ and $\eta \in S_3(a,b;y)$, we also have that $\alpha = \frac{\frac{2}{s}(\eta_i - \eta_{i+1}) + 2\frac{1}{8\mathfrak{b}^4\sqrt{d}}}{b-a} \leq \frac{1}{(b-a)\mathfrak{b}^4\sqrt{d}}$. Therefore,

$$1 - \alpha \geq 1 - \frac{1}{(b-a)\mathfrak{b}^4\sqrt{d}} \geq 1 - \frac{1}{\mathfrak{b}^2}, \tag{120}$$

where the last inequality holds by Proposition (25) we have that $b - a \geq \frac{\mathfrak{b}^2}{\sqrt{d}}$.

$$\frac{f(\eta)}{f(g(\eta))} = \prod_{\ell<j,: \ \ell,j\in[j_{\min}-2\mathfrak{b},j_{\max}+2\mathfrak{b}]} \frac{|\eta_\ell-\eta_j|^2}{|g(\eta)[\ell]-g(\eta)[j]|^2} e^{-\frac{1}{2}\sum_{\ell=j_{\min}}^{j_{\max}}(\eta_\ell^2-g(\eta)[\ell]^2)}$$
$$\times \prod_{j\in[j_{\min},j_{\max}], \ \ell\notin[j_{\min}-2\mathfrak{b},j_{\max}+2\mathfrak{b}]} \frac{|\eta_j-\eta_\ell|^2}{|g(\eta)[j]-g(\eta)[\ell]|^2}$$
$$\overset{\text{Eq. (112), (111), Lemma 30}}{\leq} (8\mathfrak{b}^4\sqrt{d}\times y)^2\times(1-\alpha)^{j_{\max}-j_{\min}+2\mathfrak{b}}\times e^{\frac{1}{2}(j_{\max}-j_{\min}+2\mathfrak{b})\alpha(b-a)\times 2\sqrt{d}}\times 2$$
$$\leq (8\mathfrak{b}^4\sqrt{d}\times y)^2\times(1-\alpha)^{3\mathfrak{b}^2}\times e^{\frac{1}{2}(j_{\max}-j_{\min}+2\mathfrak{b})\alpha(b-a)\times 2\sqrt{d}}\times 2$$
$$\leq 2e^3(8\mathfrak{b}^4\sqrt{d})^2\times y^2$$
$$\leq 50(8\mathfrak{b}^4\sqrt{d})^2\times y^2,$$

where the first inequality holds since $|\eta_\ell|\leq 2\sqrt{d}$ whenever $\eta\in E$, and the third inequality holds since $1-\alpha\geq 1-\frac{1}{\mathfrak{b}^2}$ by (120) and since $\alpha(b-a)\leq\frac{1}{2\mathfrak{b}^4\sqrt{d}}$ by (118). ■

Step 7. Dealing with the eigenvalues near the edge of the spectrum.

In this step, we extend the results of the previous steps to the eigenvalues which are near the edge of the spectrum.

Suppose that $i\leq 2\mathfrak{b}^2$. In place of the map g, we instead consider the map $\phi:\mathcal{W}_d\to\mathcal{W}_d$ such that

- $$\phi(\eta)[j]=\eta_j \qquad \forall j>i \tag{121}$$
- $$\phi(\eta)[i]=\eta_{i+1}+\frac{2}{s}(\eta_i-\eta_{i+1})+2\frac{1}{8\mathfrak{b}^4\sqrt{d}}. \tag{122}$$
- $$\phi(\eta)[j]=\phi(\eta)[j+1]+(\eta_j-\eta_{j+1}) \qquad \forall j\leq i. \tag{123}$$

Proposition 32 *Suppose that $i\leq 2\mathfrak{b}^2$. Then the following properties hold for ϕ:*

- *ϕ is injective.*
- *$\phi(\eta)[i]-\phi(\eta)[i+1]\geq\frac{1}{8\mathfrak{b}^4\sqrt{d}}$, and hence*
$$\frac{\eta_i-\eta_{i+1}}{\phi(\eta)[i]-\phi(\eta)[i+1]}\leq 8\mathfrak{b}^4\sqrt{d}\times(\eta_i-\eta_{i+1})=8\mathfrak{b}^4\sqrt{d}\times y \tag{124}$$
for any $\eta\in S_3(a,b;y)$ and any $y\leq s\frac{1}{8\mathfrak{b}^4\sqrt{d}}$.
- $$\phi(\eta)[j]-\phi(\eta)[j+1]\geq\eta_j-\eta_{j+1} \qquad \forall j\in[d]. \tag{125}$$

Proof

Injectivity: To prove that g is injective, we note that, given any vector $z \in \mathbb{R}^d$ we can find the *unique* $\eta \in \mathcal{W}_d$ such that $g(\eta) = z$ whenever such a value of η exits. We can do this by solving the system of linear equations given by (121)-(123): First, we note that by (121), we can solve for η_j for all $j > i$. Then we can plug in the value we found for η_{i+1} into (122) to solve for η_i. Finally, we can use (123) to solve for η_j for all $j < i$ recursively, starting with η_{i-1}.

Showing (124)*:* Since $y \leq s\frac{1}{8\mathfrak{b}^4\sqrt{d}}$ and $\eta \in S_3(a, b; y)$, we have that $\eta_i - \eta_{i+1} = y \leq s\frac{1}{8\mathfrak{b}^4\sqrt{d}}$. Thus, by (122),

$$\begin{aligned} g(\eta)[i] - g(\eta)[i+1] &= \frac{2}{s}(\eta_i - \eta_{i+1}) + 2\frac{1}{8\mathfrak{b}^4\sqrt{d}} \\ &\geq 2\frac{1}{8\mathfrak{b}^4\sqrt{d}} \\ &\geq \frac{1}{8\mathfrak{b}^4\sqrt{d}}. \end{aligned}$$

and, hence,

$$\frac{\eta_i - \eta_{i+1}}{g(\eta)[i] - g(\eta)[i+1]} \leq 8\mathfrak{b}^4\sqrt{d} \times (\eta_i - \eta_{i+1}) = 8\mathfrak{b}^4\sqrt{d} \times y,$$

which proves (124).

Showing (125)*:* By (122), we have

$$\begin{aligned} g(\eta)[i] - g(\eta)[i+1] &= \frac{2}{s}(\eta_i - \eta_{i+1}) + 2\frac{1}{8\mathfrak{b}^4\sqrt{d}} \\ &\geq \frac{1}{s}(\eta_i - \eta_{i+1}) \\ &\geq \eta_i - \eta_{i+1}. \end{aligned}$$

where the last inequality holds since $s \leq 1$. Thus, (125) holds for $j = i$. Moreover, (125) holds for all $j \neq i$ by (122). Therefore (125) holds for all $j \in [d]$. ■

Proposition 33 (Jacobian determinant of ϕ) *If $y \leq s\frac{1}{8\mathfrak{b}^4\sqrt{d}}$ and $\eta \in S_3(a, b; y)$, we have that*

$$\det(J_\phi(\eta)) = \frac{2}{s}.$$

Proof Consider the map $h : \mathcal{W}_d \to \mathbb{R}^d$, where $h(\eta)[j] = \eta_j - \eta_{j+1}$ for $j \in [j_{\min}, j_{\max} - 1]$ and $h(\eta)[j] = \eta_j$ for $j \in [1, d] \backslash [j_{\min}, j_{\max} - 1]$. Then, for every $\Delta \in \mathcal{W}_{j_{\max} - j_{\min}}$, by (107)-(110), we have that

$$\frac{\partial \phi \circ h^{-1}(\Delta)[\ell]}{\partial \Delta_j} = 0 \qquad \forall \ell \neq j \tag{126}$$

$$\frac{\partial \phi \circ h^{-1}(\Delta)[\ell]}{\partial \Delta_j} = 1 \qquad \forall \ell = j \neq i$$

$$\frac{\partial \phi \circ h^{-1}(\Delta)[i]}{\partial \Delta_i} = \frac{2}{s},$$

Thus, $J_{\phi \circ h^{-1}}(\Delta)$ has diagonal entries 1 for $j \neq i$ and i'th entry $= \frac{2}{s}$. Moreover, if one exchanges the i'th row and column of $J_{\phi \circ h^{-1}}(\Delta)$ with its first row and column, by (126) the resulting matrix is a $d \times d$ upper triangular matrix with the same determinant as $J_{\phi \circ h^{-1}}(\Delta)$. Thus, by Sylvester's formula, the determinant of $J_{g \circ h^{-1}}(\Delta)$ is equal to the product of its diagonal entries, and hence

$$\det(J_{\phi \circ h^{-1}}(\Delta)) = \frac{2}{s} \times 1 = \frac{2}{s}. \tag{127}$$

Hence,

$$\begin{aligned}
\det(J_\phi(\eta)) &= \det(J_{\phi \circ h^{-1} \circ h}(\eta)) = \det(J_{\phi \circ h^{-1}}(h(\eta)) \times J_h(\eta)) \\
&= \det(J_{\phi \circ h^{-1}}(h(\eta)) \times \det(J_h(\eta)) \\
&= \det(J_{\phi \circ h^{-1}}(h(\eta)) \times 1 \\
&\overset{\text{Eq. (127)}}{=} \frac{2}{s},
\end{aligned}$$

where the third equality holds since $J_h(\eta)$ is the bidiagonal matrix with diagonal entries 1 and entries -1 above the diagonal, and this matrix has determinant 1. ■

Lemma 34 (Mean field approximation, edge case)

$$\prod_{j \in [j_{\min}, j_{\max}],\ \ell \notin [j_{\min} - 2\mathfrak{b}, j_{\max} + 2\mathfrak{b}]} \frac{|\eta_j - \eta_\ell|^2}{|\phi(\eta)[j] - \phi(\eta)[\ell]|^2} \leq 2, \tag{128}$$

Proof Consider any $j \in [j_{\min}, j_{\max}]$ and any $r \geq 2\mathfrak{b}$. Then, if $\eta \in E$, by the definition of the event E we have that

$$|\eta_j - \eta_{j+r}| = r\frac{1}{2\sqrt{d}} + \rho \tag{129}$$

for some $\rho \geq 0$. Further, from (121)-(123) we have that

$$\begin{aligned}
|(\phi(\eta)[j] - \phi(\eta)[j+r]) - (\eta_j - \eta_{j+r})| &\leq \frac{2}{s}(\eta_i - \eta_{i+1}) + 2\frac{1}{8\mathfrak{b}^4\sqrt{d}} \\
&\leq \frac{1}{2\mathfrak{b}^4\sqrt{d}}
\end{aligned} \tag{130}$$

where the last inequality holds since $y \leq s\frac{1}{8\mathfrak{b}^4\sqrt{d}}$. Thus, by (129) and (130) we have that for some $\zeta \in \mathbb{R}$ where $|\zeta| \leq \frac{1}{2\mathfrak{b}^4\sqrt{d}}$, we have

$$\begin{aligned}
\frac{|\eta_j - \eta_{j+r}|^2}{|\phi(\eta)[j] - \phi(\eta)[j+r]|^2} &\leq \frac{|r\frac{1}{2\sqrt{d}} + \rho|^2}{|r\frac{1}{2\sqrt{d}} + \rho + \zeta|^2} \\
&\leq \left(1 + \frac{1}{r} \times \frac{1}{\mathfrak{b}^4}\right)^2 \\
&= \left(1 + \frac{1}{r} \times \frac{1}{\mathfrak{b}^4}\right)^2,
\end{aligned} \tag{131}$$

where the second inequality holds since $|\zeta| \leq \frac{1}{4\mathfrak{b}^4\sqrt{d}}$ and $\rho \geq 0$. Therefore, we have

$$\begin{aligned}
&\prod_{j\in[j_{\min},j_{\max}],\;\; \ell\notin[j_{\min}-2\mathfrak{b},j_{\max}+2\mathfrak{b}]} \frac{|\eta_j-\eta_\ell|^2}{|\phi(\eta)[j]-\phi(\eta)[\ell]|^2}\\
&\overset{\text{Eq. (131)}}{\leq} \prod_{\ell\in[j_{\min},j_{\max}]} \prod_{r=2\mathfrak{b}}^{d} \left(1+\frac{1}{r}\times\frac{1}{\mathfrak{b}^4}\right)^2\\
&\leq (e^{\frac{1}{\mathfrak{b}^3}})^{j_{\max}-j_{\min}}\\
&= (e^{\frac{1}{\mathfrak{b}^3}})^{2\mathfrak{b}^2}\\
&= e^{\frac{1}{\mathfrak{b}}}\\
&\geq 2,
\end{aligned}$$

where the second inequality holds since $\prod_{r=1}^{d}(1+\frac{1}{r}) = d+1$ and hence that $\prod_{r=2}^{d}(1+\frac{1}{\kappa r}) \geq (d+1)^{\frac{1}{\kappa}}$ for every $\kappa \geq 1$. Plugging in $\kappa = \mathfrak{b}^4$, we have $\prod_{r=2}^{n}(1+\frac{1}{\kappa r}) \geq (d+1)^{-\frac{1}{\mathfrak{b}^4}} \geq \left(d^{-\frac{1}{\log(d)^{\log\log d}}}\right)^{\frac{1}{\mathfrak{b}^3}} = e^{-\frac{1}{\mathfrak{b}^3}}$. ■

Lemma 35 *For any $y \leq s\frac{1}{8\mathfrak{b}^4\sqrt{d}}$ and any $\eta \in S_3(a,b;y)$, we have that*

$$\frac{f(\eta)}{f(\phi(\eta))} \leq 4(8\mathfrak{b}^4\sqrt{d})^2 \times y^2,$$

Proof

$$\begin{aligned}
\frac{f(\eta)}{f(\phi(\eta))} &= \prod_{\ell<j,:\;\; \ell,j\in[j_{\min}-2\mathfrak{b},j_{\max}+2\mathfrak{b}]} \frac{|\eta_\ell-\eta_j|^2}{|\phi(\eta)[\ell]-\phi(\eta)[j]|^2} e^{-\frac{1}{2}\sum_{\ell=i}^{1}(\eta_\ell^2-\phi(\eta)[\ell]^2)}\\
&\times \prod_{j\in[j_{\min},j_{\max}],\;\; \ell\notin[j_{\min}-2\mathfrak{b},j_{\max}+2\mathfrak{b}]} \frac{|\eta_j-\eta_\ell|^2}{|\phi(\eta)[j]-\phi(\eta)[\ell]|^2}\\
&\overset{\text{Eq. (125), (124), Lemma 34}}{\leq} (8\mathfrak{b}^4\sqrt{d}\times y)^2 \times 1 \times e^{\mathfrak{b}^2\times\frac{1}{8\mathfrak{b}^4\sqrt{d}}\times 2\sqrt{d}} \times 2\\
&\leq (8\mathfrak{b}^4\sqrt{d}\times y)^2 \times e^{\frac{1}{4\mathfrak{b}^2}} \times 2\\
&\leq 4(8\mathfrak{b}^4\sqrt{d})^2 \times y^2,
\end{aligned}$$

where the first inequality holds since $|\eta_\ell| \leq 2\sqrt{d}$ whenever $\eta \in E$. ■

Step 8. Completing the proof.
Proof [Proof of Lemma 4]

Bulk case ($\mathfrak{b}^2 \leq i \leq d-\mathfrak{b}^2$): By Proposition 28 we have that g is invertible. Therefore, since f is a probability density,

$$\int_{a_{\min}}^{a_{\max}}\int_{b_{\min}}^{b_{\max}}\int_0^{s\frac{1}{8\mathfrak{b}^4\sqrt{d}}}\int_{S_3(a,b;y)\cap E} f(g(\eta))\det(J_g(\eta))\mathrm{d}\eta\mathrm{d}y\mathrm{d}a\mathrm{d}b \leq \int_{\mathcal{W}_d} f(\eta)\mathrm{d}\eta = 1.$$

Therefore,

$$\begin{aligned}
1 &\geq \int_{a_{\min}}^{a_{\max}} \int_{b_{\min}}^{b_{\max}} \int_{0}^{s\frac{1}{8\mathfrak{b}^4\sqrt{d}}} \int_{S_3(a,b;y)\cap E} \frac{f(g(\eta))}{f(\eta)} \det(J_g(\eta)) \times f(\eta)\mathrm{d}\eta\mathrm{d}y\mathrm{d}a\mathrm{d}b \\
&\geq \int_{a_{\min}}^{a_{\max}} \int_{b_{\min}}^{b_{\max}} \int_{0}^{s\frac{1}{8\mathfrak{b}^4\sqrt{d}}} \int_{S_3(a,b;y)\cap E} \frac{1}{50(8\mathfrak{b}^4\sqrt{d})^2 \times y^2} \times \frac{1}{2s} \times f(\eta)\mathrm{d}\eta\mathrm{d}y\mathrm{d}a\mathrm{d}b \\
&\geq \int_{a_{\min}}^{a_{\max}} \int_{b_{\min}}^{b_{\max}} \int_{0}^{s\frac{1}{8\mathfrak{b}^4\sqrt{d}}} \int_{S_3(a,b;y)\cap E} \frac{1}{50s^2} \times \frac{1}{2s} \times f(\eta)\mathrm{d}\eta\mathrm{d}y\mathrm{d}a\mathrm{d}b \\
&= \frac{1}{100s^3} \int_{a_{\min}}^{a_{\max}} \int_{b_{\min}}^{b_{\max}} \int_{0}^{s\frac{1}{8\mathfrak{b}^4\sqrt{d}}} \int_{S_3(a,b;y)\cap E} f(\eta)\mathrm{d}\eta\mathrm{d}y\mathrm{d}a\mathrm{d}b,
\end{aligned}$$

where the second equality holds by Lemma 31 and Proposition 29. Therefore,

$$\int_{a_{\min}}^{a_{\max}} \int_{b_{\min}}^{b_{\max}} \int_{0}^{s\frac{1}{8\mathfrak{b}^4\sqrt{d}}} \int_{S_3(a,b;y)\cap E} f(\eta)\mathrm{d}\eta\mathrm{d}y\mathrm{d}a\mathrm{d}b \leq 100s^3. \tag{132}$$

Hence,

$$\begin{aligned}
&\mathbb{P}\left(\eta_i - \eta_{i+1} \leq s\frac{1}{8\mathfrak{b}^4\sqrt{d}}\right) \leq \mathbb{P}\left(\left\{\eta_i - \eta_{i+1} \leq s\frac{1}{8\mathfrak{b}^4\sqrt{d}}\right\} \cap E\right) + \mathbb{P}(E^c) \\
&= \int_{\left\{\eta\in\mathcal{W}_d \,:\, \eta_i-\eta_{i+1}\leq s\frac{1}{8\mathfrak{b}^4\sqrt{d}}\right\}\cap E} f(\eta)\mathrm{d}\eta + \mathbb{P}(E^c) \\
&= \int_{a_{\min}}^{a_{\max}} \int_{b_{\min}}^{b_{\max}} \int_{\left\{\eta\in\mathcal{W}_d \,:\, \eta_i-\eta_{i+1}\leq s\frac{1}{8\mathfrak{b}^4\sqrt{d}}\right\}\cap E\cap\{\eta\in\mathcal{W}_d \,:\, \eta_{j_{\min}}=a,\, \eta_{j_{\max}}=b\}} f(\eta)\mathrm{d}\eta\mathrm{d}a\mathrm{d}b + \mathbb{P}(E^c) \\
&= \int_{a_{\min}}^{a_{\max}} \int_{b_{\min}}^{b_{\max}} \int_{0}^{s\frac{1}{8\mathfrak{b}^4\sqrt{d}}} \int_{S_3(a,b;y)\cap E} f(\eta)\mathrm{d}\eta\mathrm{d}y\mathrm{d}a\mathrm{d}b + \mathbb{P}(E^c) \\
&\overset{\text{Eq. (132)}}{\leq} 100s^3 + \mathbb{P}(E^c) \\
&\overset{\text{Eq. (102)}}{\leq} 100s^3 + \frac{1}{d^{1000}}.
\end{aligned}$$

Redefining the universal constant L (and hence redefining $\mathfrak{b}$), we get that

$$\mathbb{P}\left(\eta_i - \eta_{i+1} \leq s\frac{1}{\mathfrak{b}\sqrt{d}}\right) \leq s^3 + \frac{1}{d^{1000}} \qquad \forall s > 0$$

which proves Lemma 4 for any $\mathfrak{b}^2 \leq i \leq d - \mathfrak{b}^2$.

Edge case ($\min(i, d-i) \leq \mathfrak{b}^2$): Since the joint density of the eigenvalues (9) is symmetric about 0, without loss of generality we may assume that $i \leq \mathfrak{b}^2$.

The proof of Lemma 4 for the edge case $i \leq \mathfrak{b}^2$ follows exactly the same steps as for the bulk case ($\mathfrak{b}^2 \leq i \leq d - \mathfrak{b}^2$), if we replace the map g with the map ϕ, Proposition 28 with Proposition 32, Lemma 31 with 35, and Proposition 29 with Proposition 33. ■

Appendix F. Extensions of prior results to complex matrices

F.1. Proof of Lemma 19

Proof [Proof of Lemma 19; modification of the proof of Lemma 4.5 in Mangoubi and Vishnoi (2022), for complex matrices] By the definition of $Z_\eta(t)$ we have that

$$\begin{aligned} Z_\eta(T) - Z_\eta(t_0) &= \int_{t_0}^T \mathrm{d}Z_\eta(t) \\ &= \frac{1}{2}\int_{t_0}^T \sum_{i=1}^d \sum_{j\neq i} |\lambda_i(t) - \lambda_j(t)| \frac{1}{\max(|\Delta_{ij}(t)|, \eta_{ij})} (u_i(t)u_j^*(t)\mathrm{d}B_{ij}(t) + u_j(t)u_i^*(t)\mathrm{d}B_{ij}^*(t)) \\ &\quad - \int_{t_0}^T \sum_{i=1}^d \sum_{j\neq i} (\lambda_i(t) - \lambda_j(t)) \frac{\mathrm{d}t}{\max(\Delta_{ij}^2(t), \eta_{ij}^2)} u_i(t)u_i^*(t). \end{aligned}$$

Therefore, we have that

$$\begin{aligned} &\|Z_\eta(T) - Z_\eta(t_0)\|_F^2 \\ &\leq \frac{1}{2}\left\| \int_{t_0}^T \sum_{i=1}^d \sum_{j\neq i} |\lambda_i(t) - \lambda_j(t)| \frac{1}{\max(|\Delta_{ij}(t)|, \eta_{ij})} (u_i(t)u_j^*(t)\mathrm{d}B_{ij}(t) + u_j(t)u_i^*(t)\mathrm{d}B_{ij}^*(t)) \right\|_F^2 \\ &+ \left\| \int_{t_0}^T \sum_{i=1}^d \sum_{j\neq i} (\lambda_i(t) - \lambda_j(t)) \frac{\mathrm{d}t}{\max(\Delta_{ij}^2(t), \eta_{ij}^2)} u_i(t)u_i^*(t) \right\|_F^2. \end{aligned} \tag{133}$$

The first term on the r.h.s. of (133) (inside its Frobenius norm) is a "diffusion" term–that is, the integral has mean 0 and Brownian motion differentials $\mathrm{d}B_{ij}(t)$ inside the integral. The second term on the r.h.s. (inside its Frobenius norm) is a "drift" term– that is, the integral has non-zero mean and deterministic differentials $\mathrm{d}t$ inside the integral. We bound the diffusion and drift terms separately.

Bounding the diffusion term: We first use Itô's Lemma (Lemma 6) to bound the diffusion term in (133). The idea is to apply Ito's Lemma separately to the real and complex parts of the integrand. Define

$$X(t) := \int_{t_0}^t \sum_{i=1}^d \sum_{j\neq i} |\lambda_i(t) - \lambda_j(t)| \frac{1}{\max(|\Delta_{ij}(s)|, \eta_{ij})} (u_i(s)u_j^*(s)\mathrm{d}B_{ij}(s) + u_j(s)u_i^*(s)\mathrm{d}B_{ij}^*(s))$$

for all $t \geq 0$.

Then

$$\mathrm{d}X_{\ell r}(t) = \sum_{j=1}^d R_{(\ell r)(ij)}(t)\mathrm{d}B_{(ij)}(t) + Q_{(\ell r)(ij)}(t)\mathrm{d}B_{(ij)}^*(t) \qquad \forall t \geq 0,$$

where

$$R_{(\ell r)(ij)}(t) := \left(\frac{|\lambda_i(t) - \lambda_j(t)|}{\max(|\Delta_{ij}(t)|, \eta_{ij})} u_i(t)u_j^*(t)\right)[\ell, r]$$

and

$$Q_{(\ell r)(ij)}(t) := \left(\frac{|\lambda_i(t) - \lambda_j(t)|}{\max(|\Delta_{ij}(t)|, \eta_{ij})} u_j(t)u_i^*(t)\right)[\ell, r],$$

and where we denote by either $H_{\ell r}$ or $H[\ell, r]$ the (ℓ, r)'th entry of any matrix H. Thus,

$$\begin{aligned}
\mathrm{d}X_{\ell r}(t) &= \sum_{j=1}^{d} R_{(\ell r)(ij)}(t)\mathrm{d}B_{(ij)}(t) + Q_{(\ell r)(ij)}(t)\mathrm{d}B^*_{(ij)}(t) \qquad \forall t \geq 0, \\
&= \sum_{j=1}^{d} [\mathcal{R}(R_{(\ell r)(ij)}(t)) + \mathrm{i}\mathcal{I}(R_{(\ell r)(ij)}(t))] \times [\mathcal{R}(\mathrm{d}B_{(ij)}(t)) + \mathrm{i}\mathcal{I}(\mathrm{d}B_{(ij)}(t))] \\
&+ \sum_{j=1}^{d} [\mathcal{R}(Q_{(\ell r)(ij)}(t)) + \mathrm{i}\mathcal{I}(Q_{(\ell r)(ij)}(t))] \times [\mathcal{R}(\mathrm{d}B^*_{(ij)}(t)) + \mathrm{i}\mathcal{I}(\mathrm{d}B^*_{(ij)}(t))] \\
&= \sum_{j=1}^{d} \mathcal{R}(R_{(\ell r)(ij)}(t))\mathcal{R}(\mathrm{d}B_{(ij)}(t)) + \mathrm{i}\mathcal{I}(R_{(\ell r)(ij)}(t))\mathcal{R}(\mathrm{d}B_{(ij)}(t)) \\
&\qquad + \mathrm{i}\mathcal{R}(R_{(\ell r)(ij)}(t))\mathcal{I}(\mathrm{d}B_{(ij)}(t)) - \mathcal{I}(R_{(\ell r)(ij)}(t))\mathcal{I}(\mathrm{d}B_{(ij)}(t)) \\
&+ \sum_{j=1}^{d} \mathcal{R}(Q_{(\ell r)(ij)}(t))\mathcal{R}(\mathrm{d}B^*_{(ij)}(t)) + \mathrm{i}\mathcal{I}(Q_{(\ell r)(ij)}(t))\mathcal{R}(\mathrm{d}B^*_{(ij)}(t)) \\
&\qquad + \mathrm{i}W\mathcal{R}(Q_{(\ell r)(ij)}(t))\mathcal{I}(\mathrm{d}B^*_{(ij)}(t)) - \mathcal{I}(Q_{(\ell r)(ij)}(t))\mathcal{I}(\mathrm{d}B^*_{(ij)}(t)). \qquad (134)
\end{aligned}$$

Our goal is to bound $\mathbb{E}[\|X(T) - X(t_0)\|_F^2]$. Towards this end, let $f : \mathbb{R}^{d\times d} :\to \mathbb{R}$ be the function which takes as input a $d \times d$ matrix and outputs the square of its Frobenius norm: $f(Y) := \|Y\|_F^2 = \sum_{i=1}^d \sum_{j=1}^d Y_{ij}^2$ for every $Y \in \mathbb{R}^{d\times d}$. Then

$$\frac{\partial^2}{\partial Y_{ij}\partial Y_{\alpha\beta}} f(Y) = \begin{cases} 2 & \text{if} \quad (i,j) = (\alpha, \beta) \\ 0 & \text{otherwise.} \end{cases} \qquad (135)$$

Then by (134) we have

$$\|X(T) - X(t_0)\|_F^2 = \left\| \sum_{\ell, r} \int_{t_0}^{T} \mathrm{d}X_{\ell r}(t) \right\|_F^2 \qquad (136)$$

$$\begin{aligned}
&\overset{\text{Eq.(134)}}{\leq} \left\| \int_{t_0}^{T} \sum_{\ell,r} \sum_{j=1}^{d} \mathcal{R}(R_{(\ell r)(ij)}(t))\mathcal{R}(\mathrm{d}B_{(ij)}(t)) \right\|_F^2 + \left\| \int_{t_0}^{T} \sum_{\ell,r} \sum_{j=1}^{d} \mathcal{I}(R_{(\ell r)(ij)}(t))\mathcal{R}(\mathrm{d}B_{(ij)}(t)) \right\|_F^2 \\
&+ \left\| \int_{t_0}^{T} \sum_{\ell,r} \sum_{j=1}^{d} \mathcal{R}(R_{(\ell r)(ij)}(t))\mathcal{I}(\mathrm{d}B_{(ij)}(t)) \right\|_F^2 + \left\| \int_{t_0}^{T} \sum_{\ell,r} \sum_{j=1}^{d} \mathcal{I}(R_{(\ell r)(ij)}(t))\mathcal{I}(\mathrm{d}B_{(ij)}(t)) \right\|_F^2 \\
&+ \left\| \int_{t_0}^{T} \sum_{\ell,r} \sum_{j=1}^{d} \mathcal{R}(Q_{(\ell r)(ij)}(t))\mathcal{R}(\mathrm{d}B^*_{(ij)}(t)) \right\|_F^2 + \left\| \int_{t_0}^{T} \sum_{\ell,r} \sum_{j=1}^{d} \mathcal{I}(Q_{(\ell r)(ij)}(t))\mathcal{R}(\mathrm{d}B^*_{(ij)}(t)) \right\|_F^2 \\
&+ \left\| \int_{t_0}^{T} \sum_{\ell,r} \sum_{j=1}^{d} \mathcal{R}(Q_{(\ell r)(ij)}(t))\mathcal{I}(\mathrm{d}B^*_{(ij)}(t)) \right\|_F^2 + \left\| \int_{t_0}^{T} \sum_{\ell,r} \sum_{j=1}^{d} \mathcal{I}(Q_{(\ell r)(ij)}(t))\mathcal{I}(\mathrm{d}B^*_{(ij)}(t)) \right\|_F^2.
\end{aligned} \qquad (137)$$

Since all of the terms on the r.h.s. of (136) are entirely real or imaginary for all $t \geq 0$, we can apply Itô's Lemma (Lemma 6) individually to each of these terms. The proof to bound each of these eight terms is identical (if we replace $\mathcal{R}$ with $\mathcal{I}$, R with Q, and/or $\mathrm{d}B_{(ij)}(t)$ with $\mathrm{d}B^*_{(ij)}(t)$)), since $\mathcal{R}(\mathrm{d}B_{(ij)}(t))$, $\mathcal{R}(\mathrm{d}B^*_{(ij)}(t))$, $\mathcal{I}(\mathrm{d}B_{(ij)}(t))$, $\mathcal{I}(\mathrm{d}B^*_{(ij)}(t))$ are equal in distribution. Thus, without loss of generality, we only present the proof of how to bound the term $\left\|\int_{t_0}^T \sum_{j=1}^d \mathcal{R}(R_{(\ell r)(ij)}(t))\mathcal{R}(\mathrm{d}B_{(ij)}(t))\right\|_F^2$.

Towards this end, define

$$Y(t) := \int_{t_0}^t \sum_{\ell,r}\sum_{j=1}^d \mathcal{R}(R_{(\ell r)(ij)}(t))\mathcal{R}(\mathrm{d}B_{(ij)}(t)) \qquad \forall t \geq 0$$

Then we have,

$$\begin{aligned}
&\left\|\int_{t_0}^T \sum_{\ell,r}\sum_{j=1}^d \mathcal{R}(R_{(\ell r)(ij)}(t))\mathcal{R}(\mathrm{d}B_{(ij)}(t))\right\|_F^2 \\
&= \mathbb{E}[f(Y(T)) - f(Y(t_0))] \\
&\overset{\text{Itô's Lemma (Lemma 6)}}{=} \mathbb{E}\left[\frac{1}{2}\int_{t_0}^t \sum_{\ell,r}\sum_{\alpha,\beta}\left(\frac{\partial}{\partial Y_{\alpha\beta}} f(Y(t))\right)\mathcal{R}(R_{(\ell r)(\alpha\beta)}(t))\mathcal{R}(\mathrm{d}B_{\ell r}(t))\right] \\
&\qquad + \mathbb{E}\left[\frac{1}{2}\int_{t_0}^t \sum_{\ell,r}\sum_{i,j}\sum_{\alpha,\beta}\left(\frac{\partial^2}{\partial Y_{ij}\partial Y_{\alpha\beta}} f(Y(t))\right)\mathcal{R}(R_{(\ell r)(ij)}(t))\mathcal{R}(R_{(\ell r)(\alpha\beta)}(t))\mathrm{d}t\right] \\
&= 0 + \mathbb{E}\left[\frac{1}{2}\int_{t_0}^t \sum_{\ell,r}\sum_{i,j}\sum_{\alpha,\beta}\left(\frac{\partial^2}{\partial Y_{ij}\partial Y_{\alpha\beta}} f(Y(t))\right)\mathcal{R}(R_{(\ell r)(ij)}(t))\mathcal{R}(R_{(\ell r)(\alpha\beta)}(t))\mathrm{d}t\right], \qquad (138)
\end{aligned}$$

where the third equality is Itô's Lemma (Lemma 6), and the last equality holds since

$$\mathbb{E}\left[\int_{t_0}^T \left(\frac{\partial}{\partial Y_{\alpha\beta}} f(Y(t))\right)\mathcal{R}(R_{(\ell r)(\alpha\beta)}(t))\mathcal{R}(\mathrm{d}B_{\ell r}(t))\right] = 0,$$

for each $\ell, r, \alpha, \beta \in [d]$ because $\mathrm{d}B_{\ell r}(s)$ is independent of both $Y(t)$ and $R(t)$ for all $s \geq t$ and the Brownian motion increments $\mathrm{d}B_{\alpha\beta}(s)$ satisfy $\mathbb{E}[\int_t^\tau \mathrm{d}B_{\alpha\beta}(s)] = \mathbb{E}[B_{\alpha\beta}(\tau) - B_{\alpha\beta}(t)] = 0$ for any $\tau \geq t$.

Thus, plugging (135) into (138), we have

$$\begin{aligned}
&\left\| \int_{t_0}^{T} \sum_{\ell,r} \sum_{j=1}^{d} \mathcal{R}(R_{(\ell r)(ij)}(t)) \mathcal{R}(\mathrm{d}B_{(ij)}(t)) \right\|_F^2 \\
&\overset{\text{Eq.(135),(139)}}{=} \mathbb{E}\left[\frac{1}{2} \int_{t_0}^{t} \sum_{\ell,r} \sum_{i,j} 2[\mathcal{R}(R_{(\ell r)(ij)}(t))]^2 \mathrm{d}t \right] \\
&= \mathbb{E}\left[\int_{t_0}^{t} \sum_{\ell,r} \sum_{i,j} \left(\left(\frac{|\lambda_i(t) - \lambda_j(t)|}{\max(|\Delta_{ij}(t)|, \eta_{ij})^2} \mathcal{R}\left(u_i(t) u_j^*(t) \right) \right) [\ell, r] \right)^2 \mathrm{d}t \right] \\
&= \mathbb{E}\left[\int_{t_0}^{t} \sum_{i,j} \sum_{\ell,r} \left(\left(\frac{|\lambda_i(t) - \lambda_j(t)|}{\max(|\Delta_{ij}(t)|, \eta_{ij})^2} \mathcal{R}\left(u_i(t) u_j^*(t) \right) \right) [\ell, r] \right)^2 \mathrm{d}t \right] \\
&= \mathbb{E}\left[\int_{t_0}^{t} \sum_{i,j} \left\| \frac{|\lambda_i(t) - \lambda_j(t)|}{\max(|\Delta_{ij}(t)|, \eta_{ij})} \mathcal{R}\left(u_i(t) u_j^*(t) \right) \right\|_F^2 \mathrm{d}t \right] \\
&= \mathbb{E}\left[\int_{t_0}^{t} \sum_{i,j} \left\| \mathcal{R}\left(\frac{|\lambda_i(t) - \lambda_j(t)|}{\max(|\Delta_{ij}(t)|, \eta_{ij})} u_i(t) u_j^*(t) \right) \right\|_F^2 \mathrm{d}t \right] \\
&\leq \mathbb{E}\left[\int_{t_0}^{t} \sum_{i,j} \left\| \frac{|\lambda_i(t) - \lambda_j(t)|}{\max(|\Delta_{ij}(t)|, \eta_{ij})} u_i(t) u_j^*(t) \right\|_F^2 \mathrm{d}t \right] \\
&= 2 \int_{t_0}^{T} \mathbb{E}\left[\sum_{i=1}^{d} \sum_{j \neq i} \frac{(\lambda_i(t) - \lambda_j(t))^2}{\max(\Delta_{ij}^2(t), \eta_{ij}^2)} \left\| u_i(t) u_j^*(t) \right\|_F^2 \mathrm{d}t \right] \\
&\leq 4 \int_{t_0}^{T} \mathbb{E}\left[\sum_{i=1}^{d} \sum_{j \neq i} \frac{(\lambda_i(t) - \lambda_j(t))^2}{\max(\Delta_{ij}^2(t), \eta_{ij}^2)} \mathrm{d}t \right],
\end{aligned} \tag{139}$$

where the sixth equality holds because $\langle u_i(t)u_j^*(t), u_\ell(t)u_h^*(t)\rangle = 0$ for all $(i,j) \neq (\ell, h)$, and the last equality holds because $\|u_i(t)u_j^*(t) + u_j(t)u_i^*(t)\|_F^2 = 2$ for all t with probability 1.

Thus, plugging (138) into (136) (and recalling that, from the discussion after (136), the bound we derive in (138) holds without loss of generality for all eight terms in (136)), we have that

$$\begin{aligned}
&\left\| \int_{t_0}^{t} \sum_{i=1}^{d} \sum_{j \neq i} |\lambda_i(t) - \lambda_j(t)| \frac{1}{\max(|\Delta_{ij}(s)|, \eta_{ij})} (u_i(s)u_j^*(s)\mathrm{d}B_{ij}(s) + u_j(s)u_i^*(s)\mathrm{d}B_{ij}^*(s)) \right\|_F^2 \\
&= \|X(T) - X(t_0)\|_F^2 \\
&\leq 32 \int_{t_0}^{T} \mathbb{E}\left[\sum_{i=1}^{d} \sum_{j \neq i} \frac{(\lambda_i(t) - \lambda_j(t))^2}{\max(\Delta_{ij}^2(t), \eta_{ij}^2)} \mathrm{d}t \right]
\end{aligned} \tag{140}$$

Bounding the drift term: To bound the drift term in (133), we use the Cauchy-Shwarz inequality:

$$\begin{aligned}
&\left\| \int_{t_0}^{T} \sum_{i=1}^{d} \sum_{j \neq i} (\lambda_i(t) - \lambda_j(t)) \frac{\mathrm{d}t}{\max(\Delta_{ij}^2(t), \eta_{ij}^2)} u_i(t) u_i^*(t) \right\|_F^2 \\
&= \left\| \int_{t_0}^{T} \sum_{i=1}^{d} \sum_{j \neq i} \frac{\lambda_i(t) - \lambda_j(t)}{\max(\Delta_{ij}^2(t), \eta_{ij}^2)} u_i(t) u_i^*(t) \times 1\mathrm{d}t \right\|_F^2 \\
&\overset{\text{Cauchy-Schwarz Inequality}}{\leq} \int_{t_0}^{T} \left\| \sum_{i=1}^{d} \sum_{j \neq i} \frac{\lambda_i(t) - \lambda_j(t)}{\max(\Delta_{ij}^2(t), \eta_{ij}^2)} u_i(t) u_i^*(t) \right\|_F^2 \mathrm{d}t \times \int_{t_0}^{T} 1^2 \mathrm{d}t \\
&= T \int_{t_0}^{T} \left\| \sum_{i=1}^{d} \sum_{j \neq i} \frac{\lambda_i(t) - \lambda_j(t)}{\max(\Delta_{ij}^2(t), \eta_{ij}^2)} u_i(t) u_i^*(t) \right\|_F^2 \mathrm{d}t \\
&= T \int_{t_0}^{T} \sum_{i=1}^{d} \left\| \sum_{j \neq i} \frac{\lambda_i(t) - \lambda_j(t)}{\max(\Delta_{ij}^2(t), \eta_{ij}^2)} u_i(t) u_i^*(t) \right\|_F^2 \mathrm{d}t \\
&= T \int_{t_0}^{T} \sum_{i=1}^{d} \left\| \left(\sum_{j \neq i} \frac{\lambda_i(t) - \lambda_j(t)}{\max(\Delta_{ij}^2(t), \eta_{ij}^2)} \right) u_i(t) u_i^*(t) \right\|_F^2 \mathrm{d}t \\
&= T \int_{t_0}^{T} \sum_{i=1}^{d} \left(\sum_{j \neq i} \frac{\lambda_i(t) - \lambda_j(t)}{\max(\Delta_{ij}^2(t), \eta_{ij}^2)} \right)^2 \|u_i(t) u_i^*(t)\|_F^2 \, \mathrm{d}t \\
&= T \int_{t_0}^{T} \sum_{i=1}^{d} \left(\sum_{j \neq i} \frac{\lambda_i(t) - \lambda_j(t)}{\max(\Delta_{ij}^2(t), \eta_{ij}^2)} \right)^2 \times 1\mathrm{d}t,
\end{aligned} \tag{141}$$

where the first inequality is by the Cauchy-Schwarz inequality for integrals (applied to each entry of the matrix-valued integral). The third equality holds since $\langle u_i(t)u_i^*(t), u_j(t)u_j^*(t) \rangle = 0$ for all $i \neq j$. The last equality holds since $\|u_i(t)u_i^*(t)\|_F^2 = 1$ with probability 1. Therefore, taking the expectation on both sides of (133), and plugging (140) and (141) into (133), we have

$$\begin{aligned}
\mathbb{E}\left[\|Z_\eta(T) - Z_\eta(t_0)\|_F^2\right] \leq 32 \int_{t_0}^{T} \mathbb{E}\left[\sum_{i=1}^{d} \sum_{j \neq i} \frac{(\lambda_i(t) - \lambda_j(t))^2}{\max(\Delta_{ij}^2(t), \eta_{ij}^2)} \right] \mathrm{d}t \\
+ T \int_{t_0}^{T} \mathbb{E}\left[\sum_{i=1}^{d} \left(\sum_{j \neq i} \frac{\lambda_i(t) - \lambda_j(t)}{\max(\Delta_{ij}^2(t), \eta_{ij}^2)} \right)^2 \right] \mathrm{d}t.
\end{aligned} \tag{142}$$

■

F.2. Proof of Lemma 18

Proof [Proof of Lemma 18] To compute the stochastic Ito derivative $\mathrm{d}(u_i(t)u_i^*(t))$ we apply the Dyson Brownian motion equations (5). For any $t \in [0, T]$, we have

$$\begin{aligned}
\mathrm{d}(u_i(t)u_i^*(t)) &= (u_i(t) + \mathrm{d}u_i(t))(u_i(t) + \mathrm{d}u_i(t))^* - u_i(t)u_i^*(t) \\
&= \left(u_i(t) + \sum_{j \neq i} \frac{\mathrm{d}B_{ij}(t)}{\gamma_i(t) - \gamma_j(t)} u_j(t) - \sum_{j \neq i} \frac{\mathrm{d}t}{(\gamma_i(t) - \gamma_j(t))^2} u_i(t)\right) \\
&\quad \times \left(u_i(t) + \sum_{j \neq i} \frac{\mathrm{d}B_{ij}(t)}{\gamma_i(t) - \gamma_j(t)} u_j(t) - \sum_{j \neq i} \frac{\mathrm{d}t}{(\gamma_i(t) - \gamma_j(t))^2} u_i(t)\right)^* - u_i(t)u_i^*(t) \\
&= u_i(t)u_i^*(t) + \sum_{j \neq i} \frac{1}{\gamma_i(t) - \gamma_j(t)} (u_i(t)u_j^*(t)\mathrm{d}B_{ij}(t) + u_j(t)u_i^*(t)\mathrm{d}B_{ij}^*(t)) \\
&\quad - \sum_{j \neq i} \frac{\mathrm{d}t}{(\gamma_i(t) - \gamma_j(t))^2} u_i(t)u_i^*(t) \\
&\quad + \sum_{j \neq i} \sum_{\ell \neq i} \frac{\mathrm{d}B_{ij}(t)\mathrm{d}B_{i\ell}^*(t)}{(\gamma_i(t) - \gamma_j(t))(\gamma_i(t) - \gamma_\ell(t))} u_j(t)u_\ell^*(t) \\
&\qquad - \varphi_1(t)\varphi_2^*(t) - \varphi_2(t)\varphi_1^*(t) + -\varphi_2(t)\varphi_2^*(t) - u_i(t)u_i^*(t),
\end{aligned} \tag{143}$$

where we define $\varphi_1(t) := \sum_{j \neq i} \frac{\mathrm{d}B_{ij}(t)}{\gamma_i(t) - \gamma_j(t)} u_j(t)$ and $\varphi_2(t) := \sum_{j \neq i} \frac{\mathrm{d}t}{(\gamma_i(t) - \gamma_j(t))^2} u_i(t)$. The terms $\varphi_1(t)\varphi_2^*(t)$ and $\varphi_2(t)\varphi_1^*(t)$ have differentials $O(\mathrm{d}B_{ij}\mathrm{d}t)$, and $\varphi_2(t)\varphi_2^*(t)$ has differentials $O(\mathrm{d}t^2)$; thus, all three terms vanish in the stochastic derivative by Ito's Lemma 6 (applied separately to the real and imaginary parts of these terms). Therefore, (143) implies that the stochastic derivative $\mathrm{d}(u_i(t)u_i^*(t))$ satisfies

$$\begin{aligned}
\mathrm{d}(u_i(t)u_i^*(t)) &= \sum_{j \neq i} \frac{1}{\gamma_i(t) - \gamma_j(t)} (u_i(t)u_j^*(t)\mathrm{d}B_{ij}(t) + u_j(t)u_i^*(t)\mathrm{d}B_{ij}^*(t)) \\
&\quad - \sum_{j \neq i} \frac{\mathrm{d}t}{(\gamma_i(t) - \gamma_j(t))^2} u_i(t)u_i^*(t) \\
&\quad + \sum_{j \neq i} \sum_{\ell \neq i} \frac{\mathrm{d}B_{ij}(t)\mathrm{d}B_{i\ell}^*(t)}{(\gamma_i(t) - \gamma_j(t))(\gamma_i(t) - \gamma_\ell(t))} u_j(t)u_\ell^*(t) \\
&= \sum_{j \neq i} \frac{1}{\gamma_i(t) - \gamma_j(t)} (u_i(t)u_j^*(t)\mathrm{d}B_{ij}(t) + u_j(t)u_i^*(t)\mathrm{d}B_{ij}^*(t)) - \sum_{j \neq i} \frac{\mathrm{d}t}{(\gamma_i(t) - \gamma_j(t))^2} u_i(t)u_i^*(t) \\
&\quad + \sum_{j \neq i} \frac{\mathrm{d}B_{ij}(t)\mathrm{d}B_{ij}^*(t)}{(\gamma_i(t) - \gamma_j(t))^2} u_j(t)u_j^*(t) \\
&= \sum_{j \neq i} \frac{1}{\gamma_i(t) - \gamma_j(t)} (u_i(t)u_j^*(t)\mathrm{d}B_{ij}(t) + u_j(t)u_i^*(t)\mathrm{d}B_{ij}^*(t)) - \sum_{j \neq i} \frac{\mathrm{d}t}{(\gamma_i(t) - \gamma_j(t))^2} u_i(t)u_i^*(t) \\
&\quad + \sum_{j \neq i} \frac{\mathrm{d}t}{(\gamma_i(t) - \gamma_j(t))^2} u_j(t)u_j^*(t),
\end{aligned} \tag{144}$$

where the second-to-last equality holds since all terms $\mathrm{d}B_{ij}(t)\mathrm{d}B_{i\ell}^*(t)$ with $j \neq \ell$ in the sum $\sum_{j\neq i}\sum_{\ell\neq i} \frac{\mathrm{d}B_{ij}(t)\mathrm{d}B_{i\ell}^*(t)}{(\gamma_i(t)-\gamma_j(t))(\gamma_i(t)-\gamma_\ell(t))} u_j(t)u_\ell^*(t)$ vanish by Ito's Lemma 6 since they have mean 0 and are $O(\mathrm{d}B_{ij}(t)\mathrm{d}B_{i\ell}^*(t))$; we are therefore left only with the terms $j = \ell$ in the sum which have differential terms $\mathrm{d}B_{ij}(t)\mathrm{d}B_{ij}^*(t)$ which have mean $\mathrm{d}t$ plus higher-order terms which vanish by Ito's Lemma 6. Therefore (144) implies that

$$\mathrm{d}(u_i(t)u_i^*(t)) = \sum_{j\neq i} \frac{1}{\gamma_i(t)-\gamma_j(t)}(u_i(t)u_j^*(t)\mathrm{d}B_{ij}(t) + u_j(t)u_i^*(t)\mathrm{d}B_{ij}^*(t)) - \sum_{j\neq i} \frac{\mathrm{d}t}{(\gamma_i(t)-\gamma_j(t))^2}(u_i(t)u_i^*(t) - u_j(t)u_j^*(t)).$$

■

Appendix G. Tightness of the upper bound in Theorem 2

In this section, we show that the upper bound in Theorem 2 is tight up to lower-order terms.

The case when $k = d$: To see why our bound in Theorem 2 is tight when $k = d$, one can plug in $\sigma_{d+1} = 0$ into the r.h.s. of our utility bound which gives a bound of $\sqrt{\mathbb{E}[\|\hat{M} - M\|_F^2]} \leq \tilde{O}(d\sqrt{T})$. Since $\hat{M} - M = (G + G^*) \times \sqrt{T}$ where G has complex Gaussian entries, we have that $\|\hat{M} - M\|_F = \Theta(d\sqrt{T})$ w.h.p. from standard matrix concentration bounds.

The case when $k = 1$: To see why our bound is tight when $k = 1$, consider the case when M has top eigenvalue σ_1 and all other eigenvalues $\sigma_2 = \cdots = \sigma_d$ where σ_1 is very large ($\sigma_1 \to \infty$) and $\frac{\sigma_1}{\sigma_1 - \sigma_2} = c$ for any constant $c > 0$. In this case, the eigenvalue repulsion terms $\frac{1}{\gamma_i(t)-\gamma_j(t)}$ in the eigenvalue evolution equations (4) are higher-order which scale as $\frac{1}{\sigma_1}$ as $\sigma_1 \to \infty$. Thus, from (4) we have that

$$\hat{\sigma}_1 - \sigma_1 + g_1 \tag{145}$$

with probability 1 as $\sigma_1 \to \infty$, where $g_1 \sim N(0, T)$.

In a similar manner, we have that the terms $\frac{1}{(\gamma_i(t)-\gamma_j(t))^2}$ in the eigenvector evolution equations (5) are higher-order terms which scale as $\frac{1}{\sigma_1^2}$ as $\sigma_1 \to \infty$. Denote by $v_1, \ldots, v_d$ the eigenvectors of M and $\hat{v}_1$ the top eigenvector of $\hat{M}$. Thus, we have from (5) that

$$\sigma_1(\hat{v}_1 - v_1) \to \sum_{i=2}^{d} g_i \times c \times v_i \tag{146}$$

with probability 1 as $\sigma_1 \to \infty$ where $g_2, \ldots, g_d \sim N(0, T)$. Thus, from (145) and (146) we have that

$$\|\hat{M}_1 - M_1\|_F = \|\hat{\sigma}_1\hat{v}_1\hat{v}_1^* - \sigma_1 v_1 v_1^*\|_F \to \sqrt{\sum_{i=2}^{d} g_i^2 \times c^2} = \Theta(\sqrt{d} \times c\sqrt{T})$$

with probability 1 as $\sigma_1 \to \infty$. In other words, for σ_1 large enough we have that $\|\hat{M}_1 - M_1\|_F = \Theta(\sqrt{d}\frac{\sigma_1}{\sigma_1 - \sigma_2}\sqrt{T})$ w.h.p.

The case when $1 \le k < d$**:** The above example, which was given for $k = 1$, can be generalized to any $k < d$ by setting M to have top-k eigenvalues $\sigma_1 = \cdots = \sigma_k$, and the remaining eigenvalues $\sigma_{k+1} = \cdots = \sigma_d$, and taking $\sigma_k \to \infty$ where $\frac{\sigma_k}{\sigma_{k+1} - \sigma_k} = c$ for any constant $c > 0$. In this case, we get that, for σ_k large enough, $\|\hat{M}_k - M_k\|_F = \Theta(\sqrt{k}\sqrt{d}\frac{\sigma_k}{\sigma_k - \sigma_k}\sqrt{T})$ w.h.p. Thus, for any $k \le d$, our bound is tight up to factors of $(\log d)^{\log \log d}$ hidden in the $\tilde{O}$ notation.

Appendix H. Eigengap-free utility bounds in a weaker Frobenius norm metric

The following steps can be used to extend our main result in Theorem 2 to obtain eigengap-free utility bounds on the weaker Frobenius metric $\|\hat{M}_k - M\|_F^2 - \|M_k - M\|_F^2$:

1. **Applying Ito's lemma to the weaker Frobenius norm metric:**

   When bounding the stronger utility metric $\left\|\hat{M}_k - M_k\right\|_F^2$ in the proof of Theorem 2 we apply Ito's lemma to the function $f(Y) = \|Y\|_F^2$. If we only wish to bound the weaker utility metric $\|\hat{M}_k - M\|_F^2 - \|M_k - M\|_F^2$, we can instead apply Ito's Lemma to the function $g(Y) := \|Y - M\|_F^2$. Then we have

$$\begin{aligned} \|\hat{M}_k - M\|_F^2 - \|M_k - M\|_F^2 &= g(\Psi(T)) - g(\Psi(0)) = \int_0^T g(\Psi(t))\mathrm{d}t \\ &= \int_0^T \frac{1}{2}(\mathrm{d}\Psi(t))^\top \nabla^2 g(Y)\mathrm{d}\Psi(t) + (\nabla g(Y))^\top \mathrm{d}\Psi(t)\ \mathrm{d}t, \end{aligned}$$

   where the last equality is by Ito's lemma, and where

$$\nabla g(Y)[ij] = \frac{\partial}{\partial Y_{ij}} g(Y) = 2Y_{ij} - 2M_{ij}, \qquad \text{and} \tag{147}$$

$$\nabla^2 g(Y)[ij, \alpha\beta] = \frac{\partial}{\partial Y_{ij}\partial Y_{\alpha\beta}} g(Y) = \begin{cases} 2 & \text{for } (i,j) = (\alpha, \beta) \\ 0 & \text{otherwise.} \end{cases}.$$

2. **Canceling the eigengap terms:** The extra term $-2M_{ij}$ in the first derivative (147) leads to cancellations of the terms in the utility bound which depend on the eigenvalue gap. More specifically, writing $M = \sum_{i=1}^d \gamma_i(0) u_i(0) u_i^*(0)$, we have

$$\begin{aligned} \mathbb{E}\ &\left[\int_0^T g(\Psi(t))\mathrm{d}t\right] = \mathbb{E}\left[\int_0^T \frac{1}{2}(\mathrm{d}\Psi(t))^\top \nabla^2 g(Y)\mathrm{d}\Psi(t) + (\nabla g(Y))^\top \mathrm{d}\Psi(t)\ \mathrm{d}t\right] \\ = \ &\mathbb{E}\Bigg[\int_{t_0}^T \sum_{i=1}^d \sum_{j \ne i} \frac{(\lambda_i(t) - \lambda_j(t))^2}{(\gamma_i(t) - \gamma_j(t))^2}\mathrm{d}t \\ &\quad -2\sum_{i=1}^d \sum_{j \ne i} \frac{\lambda_i(t) - \lambda_j(t)}{(\gamma_i(t) - \gamma_j(t))^2} \times \sum_{\ell=1}^k \gamma_\ell(t) \left\langle u_i(t)u_i^*(t),\ u_\ell(t)u_\ell^*(t)\right\rangle \mathrm{d}t \\ &\quad +2\sum_{i=1}^d \sum_{j \ne i} \frac{\lambda_i(t) - \lambda_j(t)}{(\gamma_i(t) - \gamma_j(t))^2} \times \sum_{\ell=1}^d \gamma_\ell(0) \left\langle u_i(0)u_i^*(0),\ u_\ell(t)u_\ell^*(t)\right\rangle \mathrm{d}t\Bigg] \\ &+\mathbb{E}\left[\int_0^T (\nabla g(Y))^\top \sum_{i=1}^d (\mathrm{d}\lambda_i(t))(u_i(t)u_i^*(t))\ \mathrm{d}t\right], \end{aligned} \tag{148}$$

where the second equality holds since $\mathbb{E}[M_{ij}\mathrm{d}B_{ij}] = 0$ because M_{ij} is a constant.

The term $\mathbb{E}\left[\int_0^T (\nabla g(Y))^\top \sum_{i=1}^d (\mathrm{d}\lambda_i(t))(u_i(t)u_i^*(t))\ \mathrm{d}t\right] = \tilde{O}(kd)$ in (148) can be bounded using the same steps as (48). Thus, we have

$$
\begin{aligned}
\mathbb{E} \quad & \left[\int_0^T g(\Psi(t))\mathrm{d}t\right] \\
= \quad & \mathbb{E}\Big[\int_{t_0}^T \Big(\sum_{i=1}^d \sum_{j\neq i} \frac{(\lambda_i(t)-\lambda_j(t))^2}{(\gamma_i(t)-\gamma_j(t))^2} \\
- \quad & 2\sum_{i=1}^d \sum_{j\neq i} \frac{\lambda_i(t)-\lambda_j(t)}{(\gamma_i(t)-\gamma_j(t))^2} \times \sum_{\ell=1}^d \gamma_\ell(t)\,\langle u_\ell(t)u_\ell^*(t),\ u_i(t)u_i^*(t)\rangle \\
& \qquad +2\sum_{i=1}^d \sum_{j\neq i} \frac{\lambda_i(t)-\lambda_j(t)}{(\gamma_i(t)-\gamma_j(t))^2} \times \sum_{\ell=1}^d \gamma_\ell(0)\,\langle u_\ell(0)u_\ell^*(0),\ u_i(t)u_i^*(t)\rangle\Big)\mathrm{d}t\Big] + \tilde{O}(kd) \\
= \quad & \mathbb{E}\Big[\int_{t_0}^T \Big(\sum_{i=1}^d \sum_{j\neq i} \frac{(\lambda_i(t)-\lambda_j(t))^2}{(\gamma_i(t)-\gamma_j(t))^2} - 2\sum_{i=k+1}^d \sum_{j\neq i} \frac{(\lambda_j(t)-\lambda_i(t))\times\gamma_i(t)}{(\gamma_i(t)-\gamma_j(t))^2} \\
& \qquad +2\sum_{i=k+1}^d \sum_{j\neq i} \frac{(\lambda_j(t)-\lambda_i(t))\times\gamma_i(t)}{(\gamma_i(t)-\gamma_j(t))^2}\Big)\mathrm{d}t\Big] + \mathbb{E}\left[\int_0^T \mathcal{H}(t)\mathrm{d}t\right] + \tilde{O}(kd) \\
= \quad & \mathbb{E}\Big[\int_{t_0}^T \sum_{i=1}^k \sum_{j\neq i} \frac{(\gamma_i(t)-\gamma_j(t))^2}{(\gamma_i(t)-\gamma_j(t))^2}\mathrm{d}t\Big] + \mathbb{E}\left[\int_0^T \mathcal{H}(t)\mathrm{d}t\right] + \tilde{O}(kd) \\
= \quad & \tilde{O}(kd),
\end{aligned}
$$

where the second equality is obtained by making small-t approximations $\gamma_i(t) \approx \gamma_i(0)$ and $u_i(0) \approx u_i(t)$, and $\mathcal{H}(t)$ are the higher-order terms which remain after making these approximations.

3. **Bounding the higher-order terms:** More specifically, the higher-order terms are

$$
\begin{aligned}
& \mathcal{H}(t) = \sum_{i=1}^d \sum_{j\neq i} \frac{\lambda_i(t)-\lambda_j(t)}{(\gamma_i(t)-\gamma_j(t))^2} \times \sum_{\ell=1}^d (\gamma_\ell(0)-\gamma_\ell(t))\,\langle u_\ell(t)u_\ell^*(t),\ u_i(t)u_i^*(t)\rangle\,\mathrm{d}t \\
+ \quad & 2\sum_{i=1}^d \sum_{j\neq i} \frac{\lambda_i(t)-\lambda_j(t)}{(\gamma_i(t)-\gamma_j(t))^2} \times \sum_{\ell=1}^d \gamma_\ell(0)\,\langle u_\ell(0)u_\ell^*(0) - u_\ell(t)u_\ell^*(t),\ u_i(t)u_i^*(t)\rangle\,\mathrm{d}t \\
= \quad & 2\sum_{i=1}^d \sum_{j\neq i} \frac{\lambda_i(t)-\lambda_j(t)}{(\gamma_i(t)-\gamma_j(t))^2} \times (\gamma_i(0)-\gamma_i(t))\mathrm{d}t \\
+ \quad & 2\sum_{i=1}^d \sum_{j\neq i} \frac{\lambda_i(t)-\lambda_j(t)}{(\gamma_i(t)-\gamma_j(t))^2} \times \sum_{\ell=1}^d \gamma_\ell(0)\,\langle u_\ell(0)u_\ell^*(0) - u_\ell(t)u_\ell^*(t),\ u_i(t)u_i^*(t)\rangle\,\mathrm{d}t,
\end{aligned} \tag{149}
$$

where the first equality holds since $\langle u_i(t)u_i^*(t),\ u_\ell(t)u_\ell^*(t)\rangle = 0$ for $\ell \neq i$, and $\langle u_i(t)u_i^*(t),\ u_i(t)u_i^*(t)\rangle = 1$.

To bound the first term on the r.h.s. of (149), we use the fact that the two-time joint distribution of Dyson Brownian motion, $f(\gamma(0), \gamma(t))$, is symmetric in the sense that it depends only on the quantities $\{|\gamma_i(t) - \gamma_j(0)|\}_{1 \leq i,j \leq d}$ (see e.g. Tao (2012)), which implies that $\mathbb{E}[\sum_{i=1}^d \sum_{j \neq i} \frac{\lambda_i(t) - \lambda_j(t)}{(\gamma_i(t) - \gamma_j(t))^2} \times (\gamma_i(0) - \gamma_i(t)) \mathrm{d}t] = 0$. The second term can be bounded in a similar manner.

After bounding these higher-order terms, one gets the eigenvalue gap-free bound

$$\mathbb{E}[\|\hat{M}_k - M\|_F - \|M_k - M\|_F] \leq \sqrt{\mathbb{E}[\|\hat{M}_k - M\|_F^2 - \|M_k - M\|_F^2]} \leq \tilde{O}(\sqrt{kd}),$$

since $\|\hat{M}_k - M\|_F \geq \|M_k - M\|_F \geq 0$ and since $(a-b)^2 = a^2 + b^2 - 2ab \leq a^2 - b^2$ for any $a \geq b \geq 0$.