arXiv:2608.18374v1 [stat.ML] 18 Aug 2026

Inference and Uncertainty Quantification for Streaming r-PCA

Haoshu Xu
University of Pennsylvania
haoshuxu@sas.upenn.edu

Hongzhe Li
University of Pennsylvania
hongzhe@upenn.edu

Abstract

We address two open questions in streaming PCA via Oja's algorithm: sharp operator-norm convergence for general rank under sub-Gaussian data, and distributional inference for the resulting subspace estimator. Existing convergence analyses, even in the rank-one case, either assume bounded data or leave non-vanishing remainder terms that prevent adaptation to a polynomially vanishing tail spectrum, while existing distributional results are confined to the rank-one case. Our convergence theory removes these remainder terms and yields a sharp rate. In the dense-tail spiked covariance regime, this rate matches the minimax rate up to logarithmic factors. More generally, we prove a matching lower bound, up to logarithmic factors, across both dense-tail and sparse-tail regimes under a mild non-degeneracy condition. The analysis yields a linearization of Oja's iterates, which in turn enables a high-dimensional Gaussian approximation for the general-rank subspace estimation error with an explicit limiting covariance. We also establish a row-wise Gaussian approximation over convex sets for the aligned difference, recovering prior rank-one results as special cases. For practical inference, we develop an online multiplier bootstrap algorithm and prove its consistency. Beyond streaming PCA, our techniques contribute to Gaussian approximation and bootstrap inference for nonconvex stochastic approximation.

1 Introduction

Streaming statistical methods are increasingly important in modern data analysis [9, 11, 16, 21]. In some applications, data arrive sequentially and updates must be performed online [5]; in others, the full dataset is available but too large to store in memory or to process efficiently with batch methods [18]. Streaming principal component analysis (PCA) is a canonical problem in this setting: the goal is to estimate the leading eigenspace of the population covariance matrix from a single pass over the data. Beyond its classical role in online dimension reduction, streaming PCA has recently appeared in memory-efficient LLM training, where low-rank structure is used to compress optimizer states [22]. Oja's algorithm [27], a standard method for streaming PCA, has also motivated differentially private variants of PCA [25].

Starting from random initialization $U_0 \sim \mathsf{Haar}(\mathbb{O}_{d,r})$, Oja's iteration for streaming PCA is given by

$$U_t = \mathsf{QR}(U_{t-1} + \eta X_t X_t^\top U_{t-1}), \tag{1}$$

where $\mathsf{QR}(\cdot)$ denotes the QR factorization. Although Oja's algorithm dates to 1982, non-asymptotic theory for its finite-sample behavior has emerged only recently [2, 13, 14, 20, 23, 24, 26], and key questions about rate optimality and uncertainty quantification remain open.

Prior convergence analysis of Oja's algorithm. Most existing analyses of Oja's algorithm [2, 13, 14, 26] yield error bounds that do not vanish even when $\Sigma = \mathbb{E}XX^\top$ is exactly rank-r. For example, one seminal work [14] derived a bound at least of order $\frac{\lambda_1}{\Delta_{\min}\sqrt{n}}$ in the rank-one case; see Table 1. This rate is tight for a generic perturbation argument based on matrix Bernstein and Wedin's theorem, but it is not statistically sharp for PCA. Indeed, fluctuations within the leading eigenspace do not affect subspace estimation; the statistical difficulty comes from the interaction between the signal and tail eigenspaces. This distinction is reflected in the statistical minimax rates, both in Frobenius norm [30] and in operator norm [33].

Motivated by this observation, several recent works [20, 23, 24] attempted to match the minimax rate under sub-Gaussian assumptions. [20] focuses on the rank-one case, while [23, 24] consider the general-rank case in Frobenius norm. However, all these analyses still contain remainder terms that do not vanish in the small-tail regime. For example, [20, Theorem 3] requires λ_2/λ_1 to be bounded away from zero, which conflicts with the goal of obtaining rates adaptive to a vanishing tail spectrum. Moreover, although [24] proves a Haar-initialization theorem for the constant-step-size case, its bound still contains a non-vanishing remainder term in the small-tail regime; in particular, the constants may depend implicitly on λ_d^{-1}. In addition, the argument does not establish the required no-exit/stability guarantee ensuring that the iterates remain in the contractive region throughout the run. Closing this gap appears to require additional sample-size or stepsize control. Later, [23] considered the varying-step-size case and circumvented the stability issue by invoking an initialization procedure from [13], but the associated burn-in period again imposes a strong implicit requirement on the sample size, and the resulting guarantee still contains a non-vanishing component in the small-tail regime. In view of these issues, it remains an open question whether the minimax rate can be achieved without any non-vanishing remainder term, even in the rank-one case.

Uncertainty quantification for SGD and PCA. This paper is also motivated by recent advances in uncertainty quantification for SGD and high-dimensional PCA. In the strongly convex setting, [31] established Gaussian approximation results for the last iterate of SGD. This naturally raises the question of whether analogous distributional guarantees can be obtained for nonconvex problems, with Oja's algorithm for streaming PCA serving as a canonical example. For Oja's algorithm, [26] established a high-dimensional Gaussian approximation for the rank-one case based on a Hoeffding decomposition of the unnormalized matrix product. Later, [17] extended this approach to entrywise Gaussian approximation for the rank-one iterates. However, due to the noncommutative nature of matrix multiplication, this Hoeffding-decomposition approach does not readily extend to the general-rank case. In the offline PCA setting, [32] established distributional guarantees for heteroscedastic PCA under a spiked covariance model. These developments motivate the present work: can one obtain inference guarantees for streaming PCA in the general-rank case, beyond the spiked covariance model?

1.1 Contributions

Motivated by these open questions, we develop a refined analysis of Oja's dynamics. This analysis yields sharp operator-norm convergence rates and enables Gaussian approximation and inference.

Sharp operator-norm convergence. Table 1 summarizes our rate alongside prior work, where $\nu_2 := (\lambda_{r+1\sim d}/\lambda_{1\sim r})^{1/2}$ and $\nu_\infty := (\lambda_{r+1}/\lambda_1)^{1/2}$ denote tail-to-signal ratios (formally defined in Section 2). Under sub-Gaussian data and Haar initialization, our bound contains no additional non-vanishing remainder term, in contrast to [20, 23, 24], whose residuals prevent adaptivity to a vanishing tail spectrum. The leading factor $\nu_\infty \vee \nu_2$ separates two qualitatively different regimes: the dense-tail regime (ν_2 dominates), which covers the spiked covariance model, and the sparse-tail regime (ν_∞ dominates), arising when the tail spectrum is highly concentrated. This dichotomy is sharp: a matching lower bound shows the dependence on $\nu_\infty \vee \nu_2$ is unimprovable. As corollaries, our rate matches the minimax rate of [33] for the spiked covariance model up to a logarithmic factor, and reduces to the sharpest known rank-one rates of [20] when $r = 1$ (where $\nu_\infty \leq \nu_2$).

The improvement rests on two ingredients: a refined one-step decomposition, sharper than [24]; and matrix Freedman's inequality in place of the matrix Azuma used by [23, 24], which ignores conditional variance.

Gaussian approximation and inference. Building on a linearization result, we establish a high-dimensional Gaussian approximation for Oja's error with an explicit limiting covariance. Unlike

Reference	Rank	Tail assumption	Operator-norm error rate
[14]	$r=1$	Bounded	$\frac{\lambda_1(1\vee\nu_2)}{\Delta_{\min}}\sqrt{\frac{1}{n}}$
[20]	$r=1$	Sub-Gaussian	$\frac{\lambda_1\nu_2}{\Delta_{\min}}\sqrt{\frac{1}{n}}$
[26]	$r=1$	Finite moments	$\frac{\lambda_1(1\vee\nu_2^2)}{\Delta_{\min}}\sqrt{\frac{1}{n}}$
[13]	$r\geq 1$	Bounded	$\frac{\lambda_1(1\vee\nu_2^2)}{\Delta_{\min}}\sqrt{\frac{r^2}{n}}$
Ours	$r\geq 1$	Sub-Gaussian	$\frac{\lambda_1(\nu_\infty\vee\nu_2)}{\Delta_{\min}}\sqrt{\frac{r}{n}}$

Table 1: Comparison of operator-norm guarantees for $\|\sin\Theta(U_n,U_*)\|$. Existing results are translated into our notation under Assumption 1; for bounds involving auxiliary distribution-dependent or boundedness parameters, we display their sharp sub-Gaussian specialization justified in Appendix A. Logarithmic factors are suppressed. The displayed rate for our method concerns the nonzero-tail regime of Theorem 1. When $\operatorname{rank}(\Sigma)=r$, Theorem 2 gives geometric convergence to zero.

[26], whose Hoeffding-decomposition approach does not extend beyond rank one, our proof handles general $r\geq 1$ and recovers the rank-one result of [26]. We further establish a row-wise Gaussian approximation over convex sets for the aligned difference, which in the spiked covariance model recovers the covariance structure of [32] up to a logarithmic factor. To make these results operational, we develop an online multiplier bootstrap procedure that consistently estimates the limiting law.

1.2 Notation.

For a positive integer n, we write $[n]=\{1,\ldots,n\}$. For any $a,b\in\mathbb{R}$, let $a\wedge b:=\min\{a,b\}$ and $a\vee b:=\max\{a,b\}$. We write $f(n)\lesssim g(n)$ if there exists a constant $C>0$ such that $|f(n)|\leq Cg(n)$ for all sufficiently large n. The sub-Gaussian norm of a random variable $X\in\mathbb{R}$ is defined as $\|X\|_{\psi_2}:=\inf\{t>0:\mathbb{E}\exp(X^2/t^2)\leq 2\}$. For a random vector $X\in\mathbb{R}^d$, we define $\|X\|_{\psi_2}:=\sup_{\|v\|=1}\|v^\top X\|_{\psi_2}$. For any matrix M, let $\|M\|=\|M\|_{\mathrm{op}}$ denote the operator norm, and let $\|M\|_{\mathrm{F}}$ denote the Frobenius norm. We denote by $\mathbb{O}_{d,r}$ the set of $d\times r$ matrices with orthonormal columns, and by $\mathsf{Haar}(\mathbb{O}_{d,r})$ the Haar measure on $\mathbb{O}_{d,r}$. For $U,U_*\in\mathbb{O}_{d,r}$, let $\Theta(U,U_*)=\operatorname{diag}(\theta_i)_{i\in[r]}$, $\theta_i=\arccos(\sigma_i(U^\top U_*))$ denote the principal angles, where $\sigma_i(\cdot)$ is the i-th largest singular value, and set $\sin\Theta(U,U_*):=\operatorname{diag}(\sin\theta_i)_{i\in[r]}$. Finally, for any set $\mathcal{A}\subseteq\mathbb{R}^r$, denote $\mathcal{N}(\mu,\Sigma)\{\mathcal{A}\}:=\mathbb{P}(W\in\mathcal{A})$ where $W\sim\mathcal{N}(\mu,\Sigma)$.

2 Problem formulation and preliminaries

This section sets up the streaming PCA model and introduces the notation used in Oja's dynamics and limiting distributions. We impose the following standard sub-Gaussian assumption, as in [20, 23, 24, 33].

Assumption 1. *$X_1,\ldots,X_n$ are i.i.d. copies of $X=\Sigma^{1/2}\widetilde{X}$ where $\Sigma\in\mathbb{R}^{d\times d}$ is positive semi-definite, $\widetilde{X}\in\mathbb{R}^d$ satisfies $\mathbb{E}\widetilde{X}=0$, $\mathbb{E}\widetilde{X}\widetilde{X}^\top=I_d$ and $\|\widetilde{X}\|_{\psi_2}\leq C_\psi\lesssim 1$.*

Let $\Sigma=U_*\Lambda_*U_*^\top+U_{*,\perp}\Lambda_{*,\perp}U_{*,\perp}^\top$ be the eigendecomposition with $U_*\in\mathbb{O}_{d,r}$, $U_{*,\perp}\in\mathbb{O}_{d,d-r}$, $\Lambda_*=\operatorname{diag}(\lambda_1,\ldots,\lambda_r)$, $\Lambda_{*,\perp}=\operatorname{diag}(\lambda_{r+1},\ldots,\lambda_d)$, and $\lambda_1\geq\cdots\geq\lambda_d\geq 0$. The target of estimation is the column space of U_*. We use the shorthand $\lambda_{a\sim b}:=\sum_{i=a}^b\lambda_i$. Define the eigengaps $\Delta_{lj}:=\lambda_j-\lambda_{r+l}$ for $l\in[d-r]$, $j\in[r]$ with extremes $\Delta_{\max}:=\lambda_1-\lambda_d$ and $\Delta_{\min}:=\lambda_r-\lambda_{r+1}$. Throughout, we assume $\Delta_{\min}>0$.

Scale-invariant quantities. A useful structural property of Oja's iteration is its invariance under the following simultaneous rescaling:

$$\text{scale transformation:}\qquad X\leftarrow cX,\quad \eta\leftarrow c^{-2}\eta,\quad c>0. \tag{2}$$

Indeed, under (2), $\eta X_t X_t^\top$ is unchanged, and hence the entire Oja trajectory $\{U_t\}_{t\ge 0}$. Consequently, any subspace estimation error bound should depend on the problem only through quantities that are invariant under (2). We refer to such quantities as *scale invariant*. The *scale-invariant* quantities used throughout the paper are

$$\widetilde{\eta} := \eta\Delta_{\min}, \quad \nu_2 := \left[\frac{\lambda_{r+1\sim d}}{\lambda_{1\sim r}}\right]^{1/2}, \quad \nu_\infty := \left[\frac{\lambda_{r+1}}{\lambda_1}\right]^{1/2}, \quad \bar{\nu} := \max(\nu_2, \nu_\infty), \quad \kappa_* := \frac{\lambda_1}{\Delta_{\min}}. \tag{3}$$

Here ν_2 and ν_∞ are respectively the Schatten-2 and Schatten-∞ cases of the general tail-to-signal ratio $\nu_p := \|\Lambda_{*,\perp}^{1/2}\|_{\mathcal{S}_p}\|\Lambda_*^{1/2}\|_{\mathcal{S}_p}^{-1}$ where $\|\cdot\|_{\mathcal{S}_p}$ denotes the Schatten-p norm. We have $\nu_\infty \le 1$ and $\nu_\infty/\sqrt{r} \le \nu_2 \le \nu_\infty\sqrt{d-r}$ directly from the definitions. All these quantities are unchanged under (2), since eigenvalues of Σ would scale by c^2.

Tangent matrices. Our analysis tracks the iterates U_t through an associated *tangent matrix* $\mathcal{T}_t$, which serves both to bound the subspace estimation error and to characterize the row-wise aligned difference. For any $U \in \mathbb{O}_{d,r}$, define the *cosine*, *sine*, and *tangent matrix of* U *(relative to* U_**)* by

$$\mathcal{C}_U := U_*^\top U \in \mathbb{R}^{r\times r}, \quad \mathcal{S}_U := U_{*,\perp}^\top U \in \mathbb{R}^{(d-r)\times r}, \quad \mathcal{T}_U := \mathcal{S}_U\mathcal{C}_U^{-1} \text{ (when } \mathcal{C}_U \text{ is invertible)}. \tag{4}$$

When $r=1$, these reduce to the usual cosine, sine, and tangent of the angle between U and U_*. For iterates U_t, we use the shorthand $\mathcal{T}_t := \mathcal{T}_{U_t}$. Let $\mathbb{H} := \mathbb{R}^{(d-r)\times r}$ equipped with the Frobenius inner product $\langle\cdot,\cdot\rangle_{\mathrm{F}}$, write $\mathcal{I}$ for the identity on $\mathbb{H}$, and define the *gap operator* $\mathcal{L} : \mathbb{H} \to \mathbb{H}$ by

$$[\mathcal{L}M]_{lj} := \Delta_{lj}M_{lj}, \qquad \text{for any } M \in \mathbb{H}. \tag{5}$$

The operator $\mathcal{L}$ enters the analysis through the contraction $\mathcal{I} - \eta\mathcal{L}$ in the recursion $\mathcal{T}_{t+1} = (\mathcal{I} - \eta\mathcal{L})\mathcal{T}_t + \text{noise} + \text{remainder}$.

Signal–noise coordinates and covariance of the outer product. For each observation $X = \Sigma^{1/2}\widetilde{X}$, define the signal/noise coordinates and their whitened counterparts by

$$Y := U_*^\top X, \quad Z := U_{*,\perp}^\top X, \quad \widetilde{Y} := U_*^\top \widetilde{X}, \quad \widetilde{Z} := U_{*,\perp}^\top \widetilde{X}. \tag{6}$$

Then $Y = \Lambda_*^{1/2}\widetilde{Y}$, $Z = \Lambda_{*,\perp}^{1/2}\widetilde{Z}$, $\mathbb{E}\widetilde{Y}\widetilde{Y}^\top = I_r$, $\mathbb{E}\widetilde{Z}\widetilde{Z}^\top = I_{d-r}$ and $\mathbb{E}ZY^\top = \mathbb{E}\widetilde{Z}\widetilde{Y}^\top = 0$.

The random elements $ZY^\top, \widetilde{Z}\widetilde{Y}^\top \in \mathbb{H} = \mathbb{R}^{(d-r)\times r}$ play a vital role in the Gaussian approximation. We therefore define their covariance operators $\mathfrak{C}, \widetilde{\mathfrak{C}} : \mathbb{H} \to \mathbb{H}$ by

$$\mathfrak{C}(B) := \mathbb{E}[\langle B,\, ZY^\top\rangle_{\mathrm{F}} ZY^\top], \quad \widetilde{\mathfrak{C}}(B) := \mathbb{E}[\langle B, \widetilde{Z}\widetilde{Y}^\top\rangle_{\mathrm{F}}\widetilde{Z}\widetilde{Y}^\top], \qquad \forall B \in \mathbb{H}. \tag{7}$$

Equivalently, $\mathfrak{C}$ and $\widetilde{\mathfrak{C}}$ can be identified with the covariance matrices of the vectorizations $\mathrm{vec}(ZY^\top)$ and $\mathrm{vec}(\widetilde{Z}\widetilde{Y}^\top)$. We adopt this linear-operator formulation to emphasize the Hilbert-space geometry of $\mathbb{H}$ and to keep the notation coordinate-free. This perspective is also standard in functional data analysis [12]. For a linear operator $A : \mathbb{H} \to \mathbb{H}$, $\|A\|_{\mathrm{HS}}$ denotes its Hilbert–Schmidt norm, equivalently the Frobenius norm of its matrix representation under the standard basis of $\mathbb{H}$.

Example 1 (Spiked covariance model)**.** *Consider the spiked covariance model* $\Sigma = U_*S^2U_*^\top + \sigma^2 I_d$ *where* $U_* \in \mathbb{O}_{d,r}$*,* $S = \mathrm{diag}(s_1,\dots,s_r)$ *with* $s_1 \ge \dots \ge s_r > 0$*,* $\sigma > 0$*. Then* $\lambda_1 = s_1^2 + \sigma^2$*,* $\lambda_{1\sim r} = \sum_{i=1}^r (s_i^2+\sigma^2)$*,* $\lambda_{r+1} = \sigma^2$ *and* $\lambda_{r+1\sim d} = (d-r)\sigma^2$*. Under Gaussian* X*, we have* $Y \perp Z$ *and* $\widetilde{\mathfrak{C}} = \mathcal{I}$*.*

3 Main results

This section presents our main theoretical results. Section 3.1 establishes a sharp operator-norm convergence rate, a nearly matching lower bound (up to logarithmic factors, under a mild nondegeneracy condition), and a linearization of the final iterate. Section 3.2 uses this linearization to establish Gaussian approximations for the subspace error, both globally in Frobenius norm and row-wise. Section 3.3 develops an online multiplier bootstrap and proves its consistency.

3.1 Sharp convergence and linearization

Recall $\widetilde{\eta}$, $\bar{\nu}$ and κ_* from (3). Throughout this section, define the noise floor $\tau_* = \kappa_* \bar{\nu}\sqrt{\widetilde{\eta}\, r \log n}$ Under the stepsize $\widetilde{\eta} = c_\eta (\log n)/n$, this becomes

$$\tau_* = \frac{\lambda_1}{\Delta_{\min}}(\nu_\infty \vee \nu_2)\sqrt{\frac{c_\eta r \log^2 n}{n}}. \tag{8}$$

Theorem 1 (Sharp operator-norm convergence)**.** *Under Assumption 1, assume* $U_0 \sim \mathsf{Haar}(\mathbb{O}_{d,r})$ *and*

$$\nu_\infty \geq c_\nu n^{-C_\nu} \quad \textit{for constants } C_\nu \geq 0, c_\nu > 0. \tag{9}$$

Fix $\varepsilon \in [0, 1/2)$ *and set*

$$\widetilde{\eta} = c_\eta \frac{\log n}{n}, \quad \textit{where} \quad c_\eta \geq c_\eta^* := \left(1 - \varepsilon + \tfrac{3C_\nu}{2}\right)\left(1 + \tfrac{\log^{-3} n + \log 2}{\varepsilon \log n + \log\log n}\right). \tag{10}$$

For any $\delta_p \in (0, 1/e)$*, if*

$$n^{1-2\varepsilon} \geq C_1 c_\eta \kappa_*^2 r^3 (1 + \nu_\infty d)\,(\log n)^4 \frac{\log(1/\delta_p)}{\delta_p^2} \tag{11}$$

for some constant $C_1 > 0$ *sufficiently large, then with probability at least* $1 - \delta_p$*,*

$$\|\sin\Theta(U_k, U_*)\| \lesssim \tau_* = \frac{\lambda_1}{\Delta_{\min}}(\nu_\infty \vee \nu_2)\sqrt{\frac{c_\eta r \log^2 n}{n}}, \qquad \textit{for all } \left\lfloor \frac{c_\eta^*}{c_\eta} n \right\rfloor \leq k \leq n. \tag{12}$$

We highlight several implications of Theorem 1.

- **Rate optimality.** The factor $\bar{\nu} = \nu_\infty \vee \nu_2$ in (12) captures the worst-case contribution of the tail spectrum across two regimes. The *dense-tail regime*, where ν_2 dominates, includes the spiked covariance model (Example 1) commonly studied in the statistical literature [32, 33]; here $\nu_2 \gtrsim \nu_\infty$ whenever $\lambda_1 \lesssim \lambda_r$ and $r \lesssim d - r$. The *sparse-tail regime*, where ν_∞ dominates, arises when the tail spectrum is highly concentrated, e.g. $\lambda_1 = \cdots = \lambda_r > \lambda_{r+1} > 0 = \lambda_{r+2} = \cdots = \lambda_d$, in which case $\nu_\infty = \sqrt{r}\,\nu_2$. Similar regime-dependent behavior has also been observed in heteroscedastic PCA [33] and high-dimensional linear regression [4]. The dependence on $\bar{\nu}$ is unimprovable: Theorem 4 establishes a matching lower bound, up to a factor of $\sqrt{\log n}$, under a mild nondegeneracy condition. In the spiked covariance setting (Example 1) with $\lambda_1 \lesssim \lambda_r$ and $r \lesssim d - r$, our rate matches the minimax rate $\sqrt{\frac{d}{n}}\left(\frac{\sigma}{s_r} + \frac{\sigma^2}{s_r^2}\right)$ from [33] to a logarithmic factor. Since $\nu_\infty \leq \sqrt{r}\nu_2$ in general, the regime distinction collapses when $r = O(1)$ and yields $\bar{\nu} \asymp \nu_2$; in particular, the rank-one case ($r = 1$) is governed entirely by ν_2 (see Table 1).
- **Dependence on** C_ν**.** The step size requirement (10) depends on C_ν through the lower bound (9) on ν_∞. This is intuitive: Haar initialization yields $\|\mathcal{T}_0\| \lesssim \delta_p^{-1}\sqrt{rd\log(1/\delta_p)}$ with probability $\geq 1 - \delta_p$, so reaching the target rate within n samples requires a larger step size when ν_∞ is smaller. In the limit $\nu_\infty = \nu_2 = 0$, every update lies in the signal subspace, yet Haar initialization always leaves nonzero error in U_n for any finite n; this is reflected in $c_\eta^* \to \infty$ as $C_\nu \to \infty$. This C_ν dependence is implicit in prior work but not made explicit there: [20, Theorem 3] requires $\nu_\infty \gtrsim 1$ with $c_\eta = 2$ at $r = 1$, and [24, Remark 4.4] notes that the remainder term is non-negligible and the constant C^* in their Theorem 4.6 with $c_\eta = 3/2$ depends on λ_d^{-1}. Earlier works [13, 14, 26] avoid this dependence because their target rate effectively sets $C_\nu = 0$ for $\nu_\infty \asymp 1$ (see Table 1), and hence are not adaptive to a small tail spectrum.
- **Dependence on** c_η**.** The error bound (12) is monotone in c_η, matching the standard intuition that smaller step size yields smaller stochastic error. The step size lower bound (10) $c_\eta^* \to 1 - \epsilon + \frac{3C_\nu}{2}$ as $n \to \infty$. When $C_\nu = 0$, as assumed in [20], one can take $c_\eta = 1$.
- **Weaker sample-size requirement.** The condition (11) is weaker than the corresponding rank-one condition in [24]: the dependence on d enters only through $d\nu_\infty$, which can be much smaller than d when ν_∞ is small. Ignoring logarithmic factors, when $r, \kappa_*, c_\eta \lesssim 1$, our condition reduces to $n^{1-2\varepsilon} \gtrsim (1 + d\nu_\infty)\delta_p^{-2}$, an improvement over $n^{1-2\varepsilon} \gtrsim d\,\delta_p^{-2}$ in [24] when ν_∞ is small.

When $\operatorname{rank}(\Sigma) = r$, we have $\tau_* = 0$. However, for any well-defined $\mathcal{T}_0 \neq 0$, the error remains nonzero at every finite iteration. Thus, Theorem 1 cannot extend to the exact-rank case by simply setting $\nu_2 = \nu_\infty = 0$. Instead, the error converges geometrically to zero.

Theorem 2 (Exact-rank geometric convergence)**.** *Under Assumption 1, suppose that* $\operatorname{rank}(\Sigma) = r$ *and* $U_0 \sim \operatorname{Haar}(\mathbb{O}_{d,r})$. *Set* $\eta = \frac{c_\eta \log n}{n\lambda_r}$ *for* $c_\eta \geq \frac{1}{2}$. *For any* $\varepsilon \in [0, 1/2)$ *and* $\delta_p \in (0, 1/e)$, *if*

$$n^{1-2\varepsilon} \geq C_2 c_\eta \kappa_*^2 r^2 (\log n)^4 \frac{\log(1/\delta_p)}{\delta_p^2} \tag{13}$$

for some constant $C_2 > 0$ *sufficiently large, then, with probability at least* $1 - \delta_p$,

$$\|\sin\Theta(U_k, U_*)\| \lesssim \frac{\sqrt{d}\, n^{1/2-\varepsilon}}{\sqrt{c_\eta}\,\kappa_*\sqrt{r}\,(\log n)^2} \exp\big\{ -(1 - \tfrac{1}{n^\varepsilon \log n}) k\widetilde{\eta}\big\}, \qquad k = 0, 1, \ldots, n,$$

where $\widetilde{\eta} = \eta\lambda_r = c_\eta(\log n)/n$. *In particular,*

$$\|\sin\Theta(U_n, U_*)\| \lesssim \frac{\sqrt{d}}{\sqrt{c_\eta}\,\kappa_*\sqrt{r}\,(\log n)^2} n^{1/2-\varepsilon-c_\eta\left(1-(n^\varepsilon \log n)^{-1}\right)}.$$

Remark 1 (No finite-time exact recovery under exact rank). The distinction between Theorem 1 and Theorem 2 is intrinsic. When $\operatorname{rank}(\Sigma) = r$, we have $Z_k = 0$ almost surely, and the tangent recursion derived in Appendix C gives $\mathcal{T}_{k+1} = \mathcal{T}_k(I_r + \eta Y_{k+1}Y_{k+1}^\top)^{-1}$. Since $I_r + \eta Y_{k+1}Y_{k+1}^\top \succ 0$ for $\eta > 0$, $\mathcal{T}_k = 0$ if and only if $\mathcal{T}_0 = 0$ for every finite k. Thus, under a generic initialization, the error cannot reach zero in finitely many iterations even though $\tau_* = 0$. Theorem 2 therefore gives the appropriate form of the result: geometric convergence of the error to zero.

We now return to the nonzero-tail setting of Theorem 1 and establish a linearization of the final tangent iterate. This representation underlies the distributional results in Section 3.2–3.3. Recall from (4), (5) and (6) the tangent matrix $\mathcal{T}_n$, the signal–noise coordinates $Y_i = U_*^\top X_i$, $Z_i = U_{*,\perp}^\top X_i$, and the gap operator $\mathcal{L}$ given by $[\mathcal{L}M]_{\ell j} = (\lambda_j - \lambda_{r+\ell})M_{\ell j}$.

Proposition 3 (Linearization)**.** *Assume the conditions in Theorem 1 hold with* $c_\eta \geq c_\eta^* + \frac{1}{2}$. *Then, with probability at least* $1 - \delta_p$,

$$\left\|\mathcal{T}_n - \eta \textstyle\sum_{i=1}^n (\mathcal{I} - \eta\mathcal{L})^{n-i}(Z_i Y_i^\top)\right\| \lesssim \tau_* \left[\kappa_* \sqrt{r\widetilde{\eta}\log n}\ (1 + \bar{\nu})\right] \lesssim \frac{\tau_*}{n^\varepsilon \log n}.$$

Under the additional conditions stated below, the linearization in Proposition 3 also yields a nearly matching lower bound.

Proposition 4 (Lower bound)**.** *Assume the assumptions in Theorem 3 hold. If* $\kappa_* \lesssim 1$ *and* $Y \perp Z$, *then with probability at least* $\frac{1}{12}[1 + o(1)] - \delta_p$, $\|\sin\Theta(U_n, U_*)\| \gtrsim \frac{\tau_*}{\sqrt{\log n}}$.

Theorem 4 matches Theorem 1 up to logarithmic factors, and shows the dependence on $\bar{\nu}$ is unimprovable. The independence condition can be relaxed to a weaker nondegeneracy condition $\lambda_{\min}(\widetilde{\mathfrak{C}}) \gtrsim 1$ where $\lambda_{\min}(\widetilde{\mathfrak{C}})$ is the minimum eigenvalue of $\widetilde{\mathfrak{C}}$ (7).

3.2 Gaussian approximation

We now establish high-dimensional Gaussian approximations for the global subspace estimation error. Recall the *sine* and *tangent* matrices $\mathcal{S}_n$ and $\mathcal{T}_n$ from (4). The intuition is simple. In the rank-one case, $\sin\theta \approx \tan\theta$ when $|\tan\theta|$ is small. An analogous approximation holds in the rank-r case: $\|\mathcal{S}_n\|_{\mathrm{F}} \approx \|\mathcal{T}_n\|_{\mathrm{F}}$ when $\|\mathcal{T}_n\|$ is small. Together with the identity $\|\sin\Theta(U_n, U_*)\|_{\mathrm{F}} = \|\mathcal{S}_n\|_{\mathrm{F}}$ and the linearization in Theorem 3, this gives the heuristic approximation

$$\|\sin\Theta(U_n, U_*)\|_{\mathrm{F}} = \|\mathcal{S}_n\|_{\mathrm{F}} \approx \|\mathcal{T}_n\|_{\mathrm{F}} \approx \left\|\eta \textstyle\sum_{i=1}^n (\mathcal{I} - \eta\mathcal{L})^{n-i}(Z_i Y_i^\top)\right\|_{\mathrm{F}}. \tag{14}$$

Theorem 5 makes this intuition rigorous by applying a high-dimensional Gaussian approximation to the normalized linearized error $\eta^{1/2}\sum_{i=1}^n (\mathcal{I} - \eta\mathcal{L})^{n-i}(Z_i Y_i^\top)$.

Recall $\lambda_{a\sim b} = \sum_{i=a}^b \lambda_i$, the covariance operator $\mathfrak{C} = \operatorname{cov}(ZY^\top)$ from (7) and its Hilbert-Schmidt norm $\|\mathfrak{C}\|_{\mathrm{HS}}$. Define the cross-spectral term $\lambda_\times := [\lambda_{1\sim r}\lambda_{r+1\sim d}]^{1/2}$ and effective dimension $d_{\mathfrak{C}} := \frac{\lambda_\times^4}{\|\mathfrak{C}\|_{\mathrm{HS}}^2}$. Under Assumption 1, $d_{\mathcal{C}} \gtrsim 1$; see Lemma 38.

Theorem 5 (Gaussian approximation for the subspace error)**.** *Assume* $\|\mathfrak{C}\|_{\mathrm{HS}} > 0$ *and the conditions of Theorem 3 hold with* $\delta_p \to 0$. *Let* $\mathsf{G} \sim \mathcal{N}(0, \Gamma)$ *be a Gaussian random element in* $\mathbb{R}^{(d-r)\times r}$ *with* $\Gamma = \int_0^\infty e^{-t\mathcal{L}}\mathfrak{C}e^{-t\mathcal{L}}dt$. *If*

$$\text{(i)}\quad \widetilde{\eta}\,\kappa_*^4 r^3 d_{\mathfrak{C}}^{3/2}(\log n)^3 \to 0, \qquad \text{(ii)}\quad \widetilde{\eta}\,\kappa_*^4 r^3 d_{\mathfrak{C}}\bar{\nu}^2(\log n)^3 \to 0, \tag{15}$$

then $\sup_{t\in\mathbb{R}}\left|\mathbb{P}\big\{\eta^{-1}\|\sin\Theta(U_n, U_*)\|_{\mathrm{F}}^2 \leq t\big\} - \mathbb{P}\left\{\|\mathsf{G}\|_{\mathrm{F}}^2 \leq t\right\}\right| \to 0.$

Remark 2 (Gaussian case). When $Y \perp Z$ (in particular for Gaussian X), $\mathfrak{C} = \Lambda_* \otimes \Lambda_{*,\perp}$, $\|\mathfrak{C}\|_{\mathrm{HS}} = \|\Lambda_*\|_{\mathrm{F}} \|\Lambda_{*,\perp}\|_{\mathrm{F}}$, and $d_{\mathfrak{C}} = \frac{[\operatorname{tr}\Lambda_*]^2}{\|\Lambda_*\|_{\mathrm{F}}^2} \frac{[\operatorname{tr}\Lambda_{*,\perp}]^2}{\|\Lambda_{*,\perp}\|_{\mathrm{F}}^2} \le r(d-r)$. Thus, when $c_\eta, \kappa_*, r \lesssim 1$, condition (15) allows $d^{3/2} = o(n)$ up to logarithmic factors.

Remark 3 (Comparison with the rank-one result of [26]). When $r = 1$, $\mathfrak{C}$ reduces to the matrix $\mathfrak{C} = \mathbb{E}\left[ZZ^\top(Y^2)\right] \in \mathbb{R}^{(d-1)\times(d-1)}$, which coincides with the matrix $\mathbb{M}$ in [26, Theorem 1]. Three differences are worth highlighting.

- *Explicit limiting distribution*. In contrast to the n-dependent covariance $\bar{\mathbb{V}}_n$ in [26], our limit $\Gamma = \int_0^\infty e^{-t\mathcal{L}} \mathfrak{C} e^{-t\mathcal{L}} dt$ is explicit and is the unique solution to the Lyapunov equation $\mathcal{L}\Gamma + \Gamma\mathcal{L} = \mathfrak{C}$. The same covariance structure appears in constant-step-size SGD for convex optimization [31]; here it arises from the nonconvex streaming PCA dynamics.
- *Scale-invariant assumptions*. One key assumption [26, Eq. (9)] reads, in our notation, $\mathbb{E}\|ZY^\top\|_{\mathrm{F}}^6 / \|\mathfrak{C}\|_{\mathrm{HS}}^3 = o(n)$. Under Assumption 1, the moment estimates in Appendix I give $\mathbb{E}\|ZY^\top\|_{\mathrm{F}}^6 \lesssim \lambda_\times^6$, with equality up to constants in the Gaussian case. Since $d_C^{3/2} = \lambda_\times^6 / \|\mathfrak{C}\|_{\mathrm{HS}}^3$, this condition requires $d_C^{3/2} = o(n)$. When $r = 1$ and $\kappa_* \lesssim 1$, our condition (15)(i), with $\widetilde{\eta} \asymp (\log n)/n$, requires $d_{\mathfrak{C}}^{3/2} (\log n)^4 = o(n)$, matching the dependence on $d_{\mathfrak{C}}$ up to logarithmic factors. Another assumption in [26, Eq. (8)] is $\|\mathfrak{C}\|_{\mathrm{F}} \gtrsim 1$, which is not scale invariant under transformation (2); our conditions are stated entirely in scale-invariant terms. The remaining condition (15)(ii), which couples $d_{\mathfrak{C}}$ and $\bar{\nu}$, is implied by the assumptions of [26, Theorem 1].
- *Proof technique*. [26] relies on a Hoeffding decomposition of the unnormalized matrix product $[\prod_{i=1}^n (I_d + \eta X_i X_i^\top)] U_0$. Our proof linearizes the normalized iteration via Theorem 3, which keeps the QR step inside the analysis. For general $r > 1$, the unnormalized Hoeffding approach can formally lead to the same Gaussian limit, but making it rigorous appears difficult: without normalization at the level of the expansion, the argument produces exponentially growing terms that are hard to control.

We next turn to row-wise Gaussian approximation. To go beyond the Frobenius norm, we must account for the rotational non-identifiability of the leading eigenspace: for any $R \in \mathbb{O}_{r,r}$, the matrices U and UR span the same subspace, so without further eigenvalue separation within the signal block one can only estimate U_* up to a global rotation. We therefore study the Procrustes-aligned error. For any nonsingular matrix $H \in \mathbb{R}^{r\times r}$ with SVD $H = V_1 \Lambda_H V_2^\top$, denote $\operatorname{sgn}(H) := V_1 V_2^\top \in \mathbb{O}_{r,r}$. Then $\operatorname{sgn}(U^\top U_*) = \operatorname{argmin}_{R \in \mathbb{O}_{r,r}} \|UR - U_*\|_{\mathrm{F}}^2$, and the aligned difference $U \operatorname{sgn}(U^\top U_*) - U_*$ is the natural object for row-wise inference [1, 6, 7].

The intuition is again based on the tangent linearization (Theorem 3). Recall the *cosine*, *sine* and *tangent* matrices $\mathcal{C}_n$, $\mathcal{S}_n$ and $\mathcal{T}_n$ from (4). By definition, $U_n^\top U_* = \mathcal{C}_n^\top$, $U_n^\top U_{*,\perp} = \mathcal{S}_n^\top$ and $\mathcal{S}_n = \mathcal{T}_n \mathcal{C}_n$, which gives the exact decomposition

$$U_n \operatorname{sgn}(U_n^\top U_*) - U_* = U_{*,\perp} \mathcal{T}_n - [U_* + U_{*,\perp}\mathcal{T}_n][I_r - \mathcal{C}_n \mathcal{C}_n^\top] + U_n[\operatorname{sgn}(\mathcal{C}_n^\top) - \mathcal{C}_n^\top]$$

In the rank-one case, $|1 - \cos\theta| \lesssim \tan^2\theta$ when $|\tan\theta|$ is small. Analogously, in the rank-r case: $\|I_r - \mathcal{C}_n \mathcal{C}_n^\top\| \lesssim \|\mathcal{T}_n\|^2$ and $\|\operatorname{sgn}(\mathcal{C}_n^\top) - \mathcal{C}_n^\top\| \lesssim \|\mathcal{T}_n\|^2$ when $\|\mathcal{T}_n\|$ is small. Thus the aligned difference is heuristically $U_{*,\perp}\mathcal{T}_n$. Together with the linearization in Theorem 3, this gives the heuristic approximation

$$U_n \operatorname{sgn}(U_n^\top U_*) - U_* \approx U_{*,\perp}\mathcal{T}_n \approx U_{*,\perp} \eta \sum_{i=1}^n (\mathcal{I} - \eta\mathcal{L})^{n-i} (Z_i Y_i^\top).$$

This motivates the following limiting row-wise covariance.

Limiting covariance. For $u \in \mathbb{R}^{d-r}$, let $\Gamma_u \in \mathbb{R}^{r\times r}$ denote the u-projection of Γ, with entries

$$[\Gamma_u]_{j'j} = \sum_{l,l'=1}^{d-r} \frac{u_{l'} u_l \, \mathbb{E}\left[[Z]_l \, [Y]_j \, [Z]_{l'} \, [Y]_{j'}\right]}{\Delta_{lj} + \Delta_{l'j'}}, \qquad \forall j, j' \in [r]. \tag{16}$$

Here $[v]_k$ denotes the k-th coordinate of vector v; see Appendix F for properties of Γ_u. Let $\lambda_{\min}(\widetilde{\mathfrak{C}})$ denote the minimum eigenvalue of $\widetilde{\mathfrak{C}}$ from (7). For $m \in [d]$, set $u_m := U_{*,\perp}^\top e_m \in \mathbb{R}^{d-r}$ and $\Sigma_{U,m}^* := \eta\Gamma_{u_m}$.

Theorem 6 (Row-wise Gaussian approximation). *Assume the conditions of Theorem 3 hold with $\delta_p \to 0$. Fix $m \in [d]$. If $\Lambda_{*,\perp}^{1/2} u_m \neq 0$, $\lambda_{\min}(\widetilde{\mathfrak{C}}) > 0$, $\widetilde{\eta}\kappa_*^6 r^{7/2} \lambda_{\min}^{-3}(\widetilde{\mathfrak{C}}) \to 0$ and*

$$\widetilde{\eta}\kappa_*^4 r^{5/2} (\log n)^2 \bar{\nu}^2 \; (1+\bar{\nu})^2 \frac{\lambda_1 \max(\|u_m\|^2, \|U_*^\top e_m\|^2)}{\|\Lambda_{*,\perp}^{1/2} u_m\|^2} \lambda_{\min}^{-1}(\widetilde{\mathfrak{C}}) \to 0, \tag{17}$$

then $\sup_{\mathcal{A} \in \mathscr{A}^r} \left| \mathbb{P}\left\{ [U_n \operatorname{sgn}(U_n^\top U_*) - U_*]_{m,\cdot} \in \mathcal{A} \right\} - \mathcal{N}(0, \Sigma_{U,m}^*)\{\mathcal{A}\} \right| \to 0$ *where $\mathscr{A}^r$ is the family of convex sets in $\mathbb{R}^r$.*

Remark 4 (Comparison with existing row-wise inference results).

- *Nondegeneracy.* Condition (17), involving $\lambda_{\min}^{-1}(\widetilde{\mathfrak{C}})$, is scale invariant under (2) since $\widetilde{\mathfrak{C}} = \operatorname{cov}(\widetilde{Z}\widetilde{Y}^\top)$ is defined in whitened coordinates. This condition is mild in standard models: for Gaussian data, $\widetilde{Y} \perp \widetilde{Z}$ so $\widetilde{\mathfrak{C}} = \mathcal{I}$ and $\lambda_{\min}(\widetilde{\mathfrak{C}}) = 1$. It ensures that the leading term dominates higher-order remainders; related lower bound conditions appear in [3, 7, 17]. In [32, Remark 2], an analogous nondegeneracy is encoded by assuming the noise variances are bounded away from zero, which prevents the limiting Gaussian covariance from degenerating.
- *Rank-one case.* When $r = 1$, $\mathfrak{C}$ and Γ reduce to $(d-1) \times (d-1)$ matrices, with entries $\Gamma_{ll'} = \mathfrak{C}_{ll'}/(\Delta_{1l} + \Delta_{1l'})$, and (16) collapses to $\Sigma_{U,m}^* = \eta e_m^\top U_{*,\perp} \Gamma U_{*,\perp}^\top e_m \in \mathbb{R}$, recovering the limit of [17, Proposition 1].
- *Spiked covariance.* Under Example 1 with Gaussian design, so that $\widetilde{Y} \perp \widetilde{Z}$, the independence formula (Section F) gives, for $\eta = \frac{c_\eta \log n}{n \Delta_{\min}}$,

  $$\Sigma_{U,m}^* = \tfrac{c_\eta (\log n) \|u_m\|^2}{2n} \cdot \tfrac{\sigma^2}{s_r^2} \left[I_r + \sigma^2 S^{-2} \right].$$

  The row-wise covariance in [32] is $n^{-1}\sigma^2 S^{-2}[I_r + \sigma^2 S^{-2}]$. Thus, for the jth coordinate, the ratio of our covariance to that of [32] is $\frac{c_\eta \log n}{2} \|u_m\|^2 \frac{s_j^2}{s_r^2}$. Since $\|u_m\| \leq 1$ and $\frac{s_j^2}{s_r^2} \leq \kappa_\Delta$, this ratio is at most $\frac{c_\eta \log n}{2} \kappa_\Delta$. Here, the logarithmic factor comes from the constant step size, while κ_Δ arises from calibrating the step size to the smallest eigengap $\Delta_{\min} = s_r^2$.

3.3 Multiplier bootstrap

The Gaussian approximations in Section 3.2 depend on unknown covariance operators. We now give an online multiplier bootstrap algorithm that estimates the laws without explicitly estimating the covariances. Algorithm 1 runs, in parallel with Oja's algorithm, perturbed Oja recursions driven by independent multiplier weights $w_t^{(b)}$. This construction is different from the rank-one bootstrap of [26], and is motivated by linearization (Theorem 3) as follows.

Defining *cosine*, *sine* and *tangent* matrices for the bootstrap iterates relative to U_* by $\mathcal{C}_t^{(b)} = U_*^\top U_t^{(b)}$, $\mathcal{S}_t^{(b)} = U_{*,\perp}^\top U_t^{(b)}$ and $\mathcal{T}_t^{(b)} = \mathcal{S}_t^{(b)} (\mathcal{C}_t^{(b)})^{-1}$, a bootstrap analogue of Theorem 3 gives $\mathcal{T}_n^{(b)} \approx \eta \sum_{i=1}^n w_i^{(b)} (\mathcal{I} - \eta \mathcal{L})^{n-i} (Z_i Y_i^\top)$. Since both U_n and $U_n^{(b)}$ are close to U_*, their relative subspace error is controlled by the difference of their tangent matrices:

$$\|\sin\Theta(U_n^{(b)}, U_n)\|_{\mathrm{F}} \approx \|\mathcal{T}_n^{(b)} - \mathcal{T}_n\|_{\mathrm{F}} \approx \|\eta \textstyle\sum_{i=1}^n (w_i^{(b)} - 1)(\mathcal{I} - \eta\mathcal{L})^{n-i}(Z_i Y_i^\top)\|_{\mathrm{F}}.$$

Thus, conditionally on the data, the multiplier weights reproduce the same linearized fluctuation that appears in (14) (Theorem 7). An analogous argument for the aligned row-wise difference yields the row-wise bootstrap result (Theorem 8).

Assumption 2. *The multiplier weights $\{w_t^{(b)}\}_{t \in [n], b \in [B]}$ are i.i.d., independent of $\{X_i\}_{i \in [n]}$, and satisfy $\mathbb{E}(w_t^{(b)}) = 1$, $\operatorname{Var}(w_t^{(b)}) = 1$ and $w_t^{(b)} \in [0, C_w]$ almost surely for some constant $C_w > 0$.*

For example, the two-point multiplier $w_t^{(b)} \sim 2\,\mathrm{Bernoulli}(1/2)$ satisfies Assumption 2. Our theory is stated for bounded multipliers. Unbounded choices such as standard exponential weights can also be handled by a truncation argument, at the cost of additional logarithmic factors. In practice, we found that bounded multipliers performed better.

Algorithm 1: Online multiplier bootstrap for Oja's algorithm

Input : $\{X_t\}_{t\in[n]}$, stepsize η, number of bootstrap replicates B, weights $\{w_t^{(b)}\}_{t\in[n],b\in[B]}$.
Output : Oja's solution U and bootstrapped versions $\{U^{(b)}\}_{b\in[B]}$.
Draw $U_0 \sim \mathsf{Haar}(\mathbb{O}_{d,r})$
Initialize $U, U^{(1)}, \dots, U^{(B)} \leftarrow U_0$
for $t = 1, \dots, n$ **do**
  Update $U \leftarrow \mathsf{QR}(U + \eta X_t(X_t^\top U))$
  for $b = 1, \dots, B$ **do**
    Update $U^{(b)} \leftarrow \mathsf{QR}(U^{(b)} + w_t^{(b)} \eta X_t(X_t^\top U^{(b)}))$

Let $\mathcal{F}_X := \sigma(U_0, X_1, \dots, X_n)$, and let $\mathbb{P}_X$ denote the law of $(U_0, X_1, \dots, X_n)$. Conditional on $\mathcal{F}_X$, let $\mathbb{P}^*$ and $\mathbb{E}^*$ denote probability and expectation with respect to the bootstrap multipliers. Let U_n^{bs} denote a generic bootstrap iterate generated by Algorithm 1 with an independent multiplier sequence $\{w_t^{\mathsf{bs}}\}_{t\in[n]}$ satisfying Assumption 2.

Theorem 7 (Bootstrap consistency for the subspace error)**.** *Assume Assumption 2 holds and suppose the conditions in Theorem 5 are satisfied. Then*

$$\sup_{t\in\mathbb{R}} \left| \mathbb{P}^*\left\{ \eta^{-1} \| \sin\Theta(U_n^{\mathsf{bs}}, U_n) \|_{\mathrm{F}}^2 \le t \right\} - \mathbb{P}\left\{ \eta^{-1} \| \sin\Theta(U_n, U_*) \|_{\mathrm{F}}^2 \le t \right\} \right| \xrightarrow{\mathbb{P}_X} 0.$$

Theorem 8 (Row-wise bootstrap consistency)**.** *Assume Assumption 2 holds and suppose the conditions in Theorem 6 are satisfied. Denote $R_n := \operatorname{sgn}(U_n^\top U_*)$. Then for any fixed $m \in [d]$,*

$$\sup_{\mathcal{A}\in\mathscr{A}^r} \left| \mathbb{P}^*\left\{ \left[\left(U_n^{\mathsf{bs}} \operatorname{sgn}(U_n^{\mathsf{bs}\top} U_n) - U_n \right) R_n \right]_{m,\cdot} \in \mathcal{A} \right\} - \mathbb{P}\left\{ [U_n R_n - U_*]_{m,\cdot} \in \mathcal{A} \right\} \right| \xrightarrow{\mathbb{P}_X} 0,$$

where $\mathscr{A}^r$ denotes the set of all convex sets in $\mathbb{R}^r$.

Remark 5 (Non-asymptotic Berry–Esseen bound). Although Theorem 8 is stated asymptotically, its proof can be strengthened to yield a non-asymptotic Berry–Esseen-type bound over convex sets. The Berry–Esseen step in Appendix H.3 is already non-asymptotic; the $o_{\mathbb{P}_X}(1)$ terms arise mainly from empirical moment and covariance bounds obtained via Markov's inequality. Replacing these with suitable Bernstein-type concentration bounds would give explicit finite-sample errors, at the cost of additional logarithmic factors in the conditions.

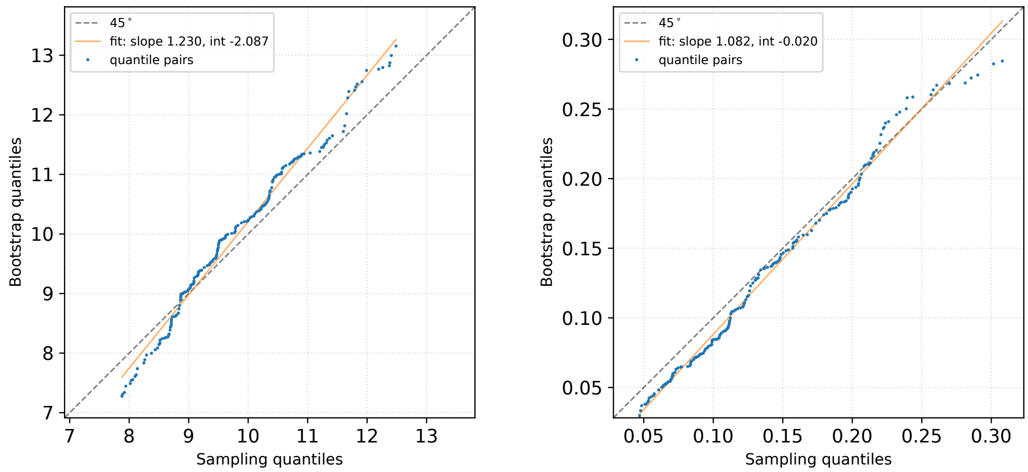
4 Main proof ideas

The proofs of our main results are built around a common analysis of Oja's dynamics in tangent coordinates. The key ideas include a refined convergence argument based on variance-sensitive concentration, a linearization of the final iterate that underlies both the lower bound and Gaussian approximation, and a corresponding linearization for the multiplier bootstrap. We provide a high-level overview of these arguments and their connections in Appendix B.

5 Experiments

We draw d-dimensional Gaussian random vectors $X_1, \dots, X_n \overset{iid}{\sim} \mathcal{N}(0, \Sigma)$ where $(U_*, U_{*,\perp}) \sim \mathsf{Haar}(\mathbb{O}_{d,d})$, $\lambda_1 = \ldots = \lambda_r = 1$ and $\lambda_{r+l} = (l+1)^{-\beta}$ for $l \in [d-r]$, for some $\beta > 0$. We let the multiplier weights be i.i.d. from the two-point distribution $2\text{Bernoulli}(1/2)$ and set $\eta = \frac{c_\eta \log n}{n \Delta_{\min}}$ with $c_\eta = 2$.

Validation of Theorem 7. We set $n = 5000$, $d = 500$, $r = 3$, and consider $\beta \in \{1, 4\}$. Following [26], we fix the Haar initialization U_0. For the sampling distribution, we generate 200 independent datasets and compute $\eta^{-1} \| \sin\Theta(U_n, U_*) \|_{\mathrm{F}}^2$. For the bootstrap distribution, we fix a dataset $\{X_i : i \in [n]\}$ and run Algorithm 1 with $B = 200$ independent bootstrap sequences, computing $\eta^{-1} \| \sin\Theta(U_n^{(b)}, U_n) \|_{\mathrm{F}}^2$. Figure 1 compares the corresponding empirical quantiles through Q–Q plots. A larger β yields faster tail decay and hence smaller values of $\eta^{-1} \| \sin\Theta(U_n, U_*) \|_{\mathrm{F}}^2$. The bootstrap approximation is visibly closer for the faster-decaying tail.



(a) $\beta = 1$. (b) $\beta = 4$.

Figure 1: Q–Q plots comparing the bootstrap and sampling distributions for $n = 5000$, $d = 500$, $r = 3$, $c_\eta = 2$ and $\beta \in \{1, 4\}$.

Validation of Theorem 8. We set $n = 2000$, $d = 100$, $r = 3$ and $\beta \in \{0.5, 1, 4\}$. We follow the protocol of [32]. For each of 200 independent Monte Carlo trials, we run Oja's algorithm together with $B = 200$ multiplier-bootstrap trajectories. For each row $m \in [d]$, we align the bootstrap iterates to U_n, form the bootstrap residuals via $[U_n^{(b)} \operatorname{sgn}(U_n^{(b)\top} U_n) - U_n]_{m,\cdot}$, and estimate their covariance by the sample covariance $\widehat{\Sigma}_m^{\mathsf{bs}}$ across bootstrap replicates. We then construct the row-wise ellipsoidal confidence region $\mathrm{CR}_{m,1-\alpha}^{\mathsf{bs}} = \{z \in \mathbb{R}^r : (z - [U_n]_{m,\cdot})^\top [\widehat{\Sigma}_m^{\mathsf{bs}}]^{-1}(z - [U_n]_{m,\cdot}) \le \chi^2_{r,1-\alpha}\}$, with $\alpha = 0.05$. We report the empirical probability that $[U_* \operatorname{sgn}(U_*^\top U_n)]_{m,\cdot}$ is covered by $\mathrm{CR}_{m,1-\alpha}^{\mathsf{bs}}$. Table 2 reports the average coverage rates; row-specific coverage and Q–Q diagnostics are provided in Appendix J.

Table 2: Empirical coverage rates for different β's over 200 Monte Carlo trials.

	$\beta = 0.5$	$\beta = 1$	$\beta = 4$
Mean±Std coverage	0.914 ± 0.027	0.930 ± 0.028	0.948 ± 0.021

6 Discussion

We established sharp operator-norm convergence guarantees for streaming r-PCA via Oja's algorithm under sub-Gaussian data. The rate adapts to both dense-tail and sparse-tail regimes, and is matched by a lower bound up to logarithmic factors. A key ingredient is a linearization of the tangent matrix $\mathcal{T}_n$, which enables high-dimensional Gaussian approximation for the global subspace error and a row-wise Gaussian approximation over convex sets for the aligned difference. To make these distributional results operational, we developed an online multiplier bootstrap procedure that consistently approximates the limiting laws without explicitly estimating the covariance operators.

Several directions remain open. First, our current theory is limited to the constant-step-size case; it would be useful to extend the theory to broader stepsize schedules. Second, the linearization and bootstrap techniques developed here may be applicable to streaming PCA under additional structural constraints, such as sparsity [16], Markovian dependence [15], or quantization [9]. Finally, it would be interesting to develop analogous inference procedures for other nonconvex stochastic approximation problems, such as online tensor learning [21].

References

[1] Emmanuel Abbe, Jianqing Fan, Kaizheng Wang, and Yiqiao Zhong. Entrywise eigenvector analysis of random matrices with low expected rank. *The Annals of Statistics*, 48(3):1452–1474, 2020. doi: 10.1214/19-AOS1854.

[2] Zeyuan Allen-Zhu and Yuanzhi Li. First efficient convergence for streaming k-pca: a global, gap-free, and near-optimal rate, 2017. URL `https://arxiv.org/abs/1607.07837`.

[3] Changxiao Cai, H. Vincent Poor, and Yuxin Chen. Uncertainty quantification for nonconvex tensor completion: Confidence intervals, heteroscedasticity and optimality. In Hal Daumé III and Aarti Singh, editors, *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pages 1271–1282. PMLR, 2020. URL `https://proceedings.mlr.press/v119/cai20c.html`.

[4] T. Tony Cai and Zijian Guo. Confidence intervals for high-dimensional linear regression: Minimax rates and adaptivity. *The Annals of Statistics*, 45(2):615–646, 2017. doi: 10.1214/16-AOS1461.

[5] Shiyu Chang, Yang Zhang, Jiliang Tang, Dawei Yin, Yi Chang, Mark A. Hasegawa-Johnson, and Thomas S. Huang. Streaming recommender systems. In *Proceedings of the 26th International Conference on World Wide Web*, WWW '17, pages 381–389, Republic and Canton of Geneva, Switzerland, 2017. International World Wide Web Conferences Steering Committee. doi: 10.1145/3038912.3052627.

[6] Ji Chen, Dekai Liu, and Xiaodong Li. Nonconvex rectangular matrix completion via gradient descent without $\ell_{2,\infty}$ regularization. *IEEE Transactions on Information Theory*, 66(9):5806–5841, 2020. doi: 10.1109/TIT.2020.2992234.

[7] Yuxin Chen, Jianqing Fan, Cong Ma, and Yuling Yan. Inference and uncertainty quantification for noisy matrix completion. *Proceedings of the National Academy of Sciences*, 116(46):22931–22937, 2019. doi: 10.1073/pnas.1910053116.

[8] Yuxin Chen, Yuejie Chi, Jianqing Fan, and Cong Ma. Spectral methods for data science: A statistical perspective. *Foundations and Trends in Machine Learning*, 14(5):566–806, 2021. doi: 10.1561/2200000079.

[9] Sanjoy Dasgupta, Syamantak Kumar, Shourya Pandey, and Purnamrita Sarkar. Low-precision streaming PCA. In *Advances in Neural Information Processing Systems*, 2025. URL `https://openreview.net/forum?id=GYLMYTYqZg`. NeurIPS 2025 poster.

[10] Kenneth R. Davidson and Stanislaw J. Szarek. Local operator theory, random matrices and banach spaces. In *Handbook of the Geometry of Banach Spaces*, volume 1, chapter 8, pages 317–366. North-Holland, Amsterdam, 2001. doi: 10.1016/S1874-5849(01)80010-3.

[11] Jianqing Fan, Wenyan Gong, Chris Junchi Li, and Qiang Sun. Statistical sparse online regression: A diffusion approximation perspective. In Amos Storkey and Fernando Perez-Cruz, editors, *Proceedings of the Twenty-First International Conference on Artificial Intelligence and Statistics*, volume 84 of *Proceedings of Machine Learning Research*, pages 1017–1026. PMLR, 2018. URL `https://proceedings.mlr.press/v84/fan18a.html`.

[12] Tailen Hsing and Randall Eubank. *Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators*. Wiley Series in Probability and Statistics. John Wiley & Sons, Hoboken, NJ, 2015. ISBN 9781118762547. doi: 10.1002/9781118762547.

[13] De Huang, Jonathan Niles-Weed, and Rachel Ward. Streaming k-PCA: Efficient guarantees for Oja's algorithm, beyond rank-one updates. In Mikhail Belkin and Samory Kpotufe, editors, *Proceedings of the Thirty Fourth Conference on Learning Theory*, volume 134 of *Proceedings of Machine Learning Research*, pages 2463–2498. PMLR, 2021. URL `https://proceedings.mlr.press/v134/huang21a.html`.

[14] Prateek Jain, Chi Jin, Sham M. Kakade, Praneeth Netrapalli, and Aaron Sidford. Streaming PCA: Matching matrix bernstein and near-optimal finite sample guarantees for Oja's algorithm. In Vitaly Feldman, Alexander Rakhlin, and Ohad Shamir, editors, *Proceedings of the 29th Conference on Learning Theory*, volume 49 of *Proceedings of Machine Learning Research*, pages 1147–1164. PMLR, 2016. URL `https://proceedings.mlr.press/v49/jain16.html`.

[15] Syamantak Kumar and Purnamrita Sarkar. Streaming PCA for markovian data. In *Advances in Neural Information Processing Systems*, volume 36, 2023. URL `https://openreview.net/forum?id=FQGRkwmRzm`.

[16] Syamantak Kumar and Purnamrita Sarkar. Oja's algorithm for streaming sparse PCA. In *Advances in Neural Information Processing Systems*, 2024. URL `https://openreview.net/forum?id=clQdPtooRD`. NeurIPS 2024 poster.

[17] Syamantak Kumar, Shourya Pandey, and Purnamrita Sarkar. Beyond sin-squared error: Linear time entrywise uncertainty quantification for streaming PCA. In Silvia Chiappa and Sara Magliacane, editors, *Proceedings of the Forty-first Conference on Uncertainty in Artificial Intelligence*, volume 286 of *Proceedings of Machine Learning Research*, pages 2396–2430. PMLR, 2025. URL `https://proceedings.mlr.press/v286/kumar25a.html`.

[18] John Langford, Lihong Li, and Tong Zhang. Sparse online learning via truncated gradient. In *Advances in Neural Information Processing Systems*, volume 21, pages 905–912, 2008. URL `https://proceedings.neurips.cc/paper/2008/hash/92c8c96e4c37100777c7190b76d28233-Abstract.html`.

[19] Rafał Latała. Estimation of moments of sums of independent real random variables. *The Annals of Probability*, 25(3):1502–1513, 1997. doi: 10.1214/aop/1024404522.

[20] Chris Junchi Li, Mengdi Wang, Han Liu, and Tong Zhang. Near-optimal stochastic approximation for online principal component estimation. *Mathematical Programming*, 167(1):75–97, 2018. doi: 10.1007/s10107-017-1182-z.

[21] Jingyang Li, Jian-Feng Cai, Yang Chen, and Dong Xia. Online tensor learning: Computational and statistical trade-offs, adaptivity and optimal regret, 2023.

[22] Kaizhao Liang, Bo Liu, Lizhang Chen, and Qiang Liu. Memory-efficient LLM training with online subspace descent. In *Advances in Neural Information Processing Systems*, 2024. URL `https://openreview.net/forum?id=P8rTCT6g45`.

[23] Xin Liang. On the optimality of Oja's algorithm for online PCA. *Statistics and Computing*, 33(3):62, 2023. doi: 10.1007/s11222-023-10229-z.

[24] Xin Liang, Zhen-Chen Guo, Li Wang, Ren-Cang Li, and Wen-Wei Lin. Nearly optimal stochastic approximation for online principal subspace estimation. *Science China Mathematics*, 66(5):1087–1122, 2023. doi: 10.1007/s11425-021-1972-5.

[25] Xiyang Liu, Weihao Kong, Prateek Jain, and Sewoong Oh. DP-PCA: Statistically optimal and differentially private PCA. In *Advances in Neural Information Processing Systems*, volume 35, pages 29929–29943. Curran Associates, Inc., 2022. URL `https://proceedings.neurips.cc/paper_files/paper/2022/hash/c150ebe1b9d1ca0eb61502bf979fa87d-Abstract-Conference.html`.

[26] Robert Lunde, Purnamrita Sarkar, and Rachel Ward. Bootstrapping the error of Oja's algorithm. In *Advances in Neural Information Processing Systems*, volume 34, 2021. URL `https://proceedings.neurips.cc/paper_files/paper/2021/hash/3152e3b1e52e2cb123363787d5f76c95-Abstract.html`.

[27] Erkki Oja. Simplified neuron model as a principal component analyzer. *Journal of Mathematical Biology*, 15(3):267–273, 1982. doi: 10.1007/BF00275687.

[28] Joel A. Tropp. Freedman's inequality for matrix martingales. *Electronic Communications in Probability*, 16:262–270, 2011. doi: 10.1214/ECP.v16-1624.

[29] Roman Vershynin. *High-Dimensional Probability: An Introduction with Applications in Data Science*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2 edition, 2026. ISBN 9781009490672. doi: 10.1017/9781009490672.

[30] Vincent Q. Vu and Jing Lei. Minimax sparse principal subspace estimation in high dimensions. *The Annals of Statistics*, 41(6):2905–2947, 2013. doi: 10.1214/13-AOS1151.

[31] Ziyang Wei, Jiaqi Li, Zhipeng Lou, and Wei Biao Wu. Gaussian approximation and concentration of constant learning-rate stochastic gradient descent. In *Advances in Neural Information Processing Systems*, 2025. URL `https://openreview.net/forum?id=aCPFvEg22L`.

[32] Yuling Yan, Yuxin Chen, and Jianqing Fan. Inference for heteroskedastic PCA with missing data. *The Annals of Statistics*, 52(2):729–756, 2024. doi: 10.1214/24-AOS2366.

[33] Anru R. Zhang, T. Tony Cai, and Yihong Wu. Heteroskedastic PCA: Algorithm, optimality, and applications. *The Annals of Statistics*, 50(1):53–80, 2022. doi: 10.1214/21-AOS2074.

A Additional comparison results

A.1 Error bounds from previous work under sub-Gaussian assumption

In this section, we justify the entries in Table 1. The results in [13, 14, 20, 26] are stated in terms of different problem-dependent quantities. We first translate these quantities into the notation of this paper, and then derive tight sub-Gaussian upper bounds for them under Assumption 1. These bounds are matched, up to constants, in the Gaussian case.

Lemma 1. *Suppose Assumption 1 holds and let $\lambda_1 \geq \ldots \geq \lambda_d \geq 0$ be the eigenvalues of Σ.*

(i) *Denote $\mathcal{V} = \|\mathbb{E}[(XX^\top - \Sigma)^2]\|$. Then $\mathcal{V} \lesssim C_\psi^4 \lambda_1 \operatorname{tr} \Sigma$. The bound is tight for Gaussian distribution.*

(ii) *For any $\alpha \geq 1$, denote $\mathcal{M}_\alpha = \mathbb{E}[\|XX^\top - \Sigma\|^\alpha]^{1/\alpha}$. Then $\mathcal{M}_1 \leq \mathcal{M}_2 \lesssim C_\psi^2 [\operatorname{tr} \Sigma]$. The bound is tight for Gaussian distribution.*

Proof. See Appendix A.2. □

[26] Using the bound for $\mathcal{M}_2$ from Lemma 1, the bound in [26, Corollary 1] when $r = 1$ is

$$\|\sin \Theta(U_n, U_*)\| \lesssim \frac{\mathcal{M}_2}{\Delta_{\min}} \sqrt{\frac{\log n}{n}} \leq \frac{\lambda_1(1 \vee \nu_2^2)}{\Delta_{\min}} \sqrt{\frac{\log n}{n}}.$$

[14] Using the bound for $\mathcal{V}$ from Lemma 1, their error bound is

$$\|\sin \Theta(U_n, U_*)\| \lesssim \frac{\mathcal{V}^{1/2}}{\Delta_{\min}} \sqrt{\frac{1}{n}} \asymp \frac{\lambda_1(1 \vee \nu_2^2)^{1/2}}{\Delta_{\min}} \sqrt{\frac{1}{n}} = \frac{\lambda_1(1 \vee \nu_2)}{\Delta_{\min}} \sqrt{\frac{1}{n}}.$$

[13] Let $\widetilde{\mathcal{M}}$ denote the almost-sure boundedness parameter in [13], namely

$$\widetilde{\mathcal{M}} \geq \sup_{P \in \mathbb{O}_{d,r}} \left\|P^\top (XX^\top - \Sigma)\right\|_{\mathrm{F}} \qquad \text{a.s.}$$

Their operator-norm guarantee, suppressing logarithmic factors, takes the form

$$\|\sin \Theta(U_n, U_*)\| \lesssim \frac{\widetilde{\mathcal{M}}}{\Delta_{\min}} \sqrt{\frac{1}{n}}$$

We now relate $\widetilde{\mathcal{M}}$ to the covariance scale. Since $\widetilde{\mathcal{M}} \geq \|XX^\top - \Sigma\|$, we have $\widetilde{\mathcal{M}} \geq \mathcal{M}_1$. Hence their error bound is no better than

$$\frac{\lambda_1(1 \vee \nu_2^2)}{\Delta_{\min}} \sqrt{\frac{r^2}{n}}.$$

[20] The error bound follows directly from their Theorem 3.

A.2 Proof of Lemma 1

Without loss of generality, assume Σ is diagonal.

(i) Since $XX^\top = \Sigma^{1/2} \widetilde{X} \widetilde{X}^\top \Sigma^{1/2}$, we have $XX^\top - \Sigma = \Sigma^{1/2}(\widetilde{X}\widetilde{X}^\top - I_d)\Sigma^{1/2}$, so

$$\mathbb{E}\big[\big(XX^\top - \Sigma\big)^2\big] = \Sigma^{1/2} \underbrace{\mathbb{E}\big[(\widetilde{X}\widetilde{X}^\top - I_d)\Sigma(\widetilde{X}\widetilde{X}^\top - I_d)\big]}_{=:\tilde{\mathcal{V}}} \Sigma^{1/2}$$

Therefore $\mathcal{V} = \|\Sigma^{1/2} \tilde{\mathcal{V}} \Sigma^{1/2}\|$, and it suffices to analyze $\tilde{\mathcal{V}}$. Expanding $\tilde{\mathcal{V}}$ gives

$$\tilde{\mathcal{V}} = \mathbb{E}\big[\widetilde{X}\widetilde{X}^\top \Sigma \widetilde{X}\widetilde{X}^\top\big] - \Sigma \mathbb{E}\big[\widetilde{X}\widetilde{X}^\top\big] - \mathbb{E}\big[\widetilde{X}\widetilde{X}^\top\big]\Sigma + \Sigma = \mathbb{E}\big[\widetilde{X}\widetilde{X}^\top \Sigma \widetilde{X}\widetilde{X}^\top\big] - \Sigma$$

Sub-Gaussian upper bound. Since $\tilde{\mathcal{V}}$ is symmetric and positive semi-definite, $\|\tilde{\mathcal{V}}\| = \sup_{\|u\|=1} u^\top \tilde{\mathcal{V}} u$, we have for any fixed unit vector $u \in \mathbb{R}^d$,

$$u^\top \tilde{\mathcal{V}} u \le \mathbb{E}\big[(u^\top \widetilde{X})^2(\widetilde{X}^\top \Sigma \widetilde{X})\big] \le \sqrt{\mathbb{E}(u^\top \widetilde{X})^4}\sqrt{\mathbb{E}(\widetilde{X}^\top \Sigma \widetilde{X})^2} \lesssim C_\psi^4 \operatorname{tr}\Sigma$$

Here the last inequality follows since Lemma 37 implies that $\mathbb{E}(\widetilde{X}^\top \Sigma \widetilde{X})^2 = \mathbb{E}\|\Sigma^{1/2}\widetilde{X}\|^4 \lesssim C_\psi^4 \left[\operatorname{tr}\Sigma\right]^2$ and $\mathbb{E}(u^\top \widetilde{X})^4 \lesssim C_\psi^4$. Taking supremum over u gives $\|\tilde{\mathcal{V}}\| \lesssim C_\psi^4 \operatorname{tr}\Sigma$, and thus

$$\mathcal{V} = \big\|\Sigma^{1/2}\tilde{\mathcal{V}}\Sigma^{1/2}\big\| \le \big\|\Sigma^{1/2}\big\|^2 \big\|\tilde{\mathcal{V}}\big\| \lesssim C_\psi^4 \lambda_1 \operatorname{tr}\Sigma$$

Gaussian lower bound. Assume $\widetilde{X} \sim \mathcal{N}(0, I_d)$. Since Σ is diagonal, the (i,l)-entry of $\tilde{\mathcal{V}}$ is

$$[\tilde{\mathcal{V}}]_{il} = \widetilde{X}_i \Big(\sum_{j=1}^d \widetilde{X}_j^2 \Sigma_{jj}\Big)\widetilde{X}_l - \Sigma_{il}$$

By Isserlis' theorem, we have

$$\mathbb{E}\big[\widetilde{X}_j^2 \widetilde{X}_i \widetilde{X}_l\big] = \mathbb{E}\big[\widetilde{X}_j^2\big]\mathbb{E}\big[\widetilde{X}_i \widetilde{X}_l\big] + 2\mathbb{E}\big[\widetilde{X}_j \widetilde{X}_i\big]\mathbb{E}\big[\widetilde{X}_j \widetilde{X}_l\big] = \delta_{il} + 2\delta_{ij}\delta_{jl}$$

Summing over j gives

$$\widetilde{X}_i \Big(\sum_{j=1}^d \widetilde{X}_j^2 \Sigma_{jj}\Big)\widetilde{X}_l = \sum_j \Sigma_{jj}(\delta_{il} + 2\delta_{ij}\delta_{jl}) = \delta_{il}\operatorname{tr}\Sigma + 2\delta_{il}\Sigma_{il}$$

Since Σ is diagonal, we have

$$\tilde{\mathcal{V}} = (\operatorname{tr}\Sigma) I_d + 2\Sigma - \Sigma = (\operatorname{tr}\Sigma) I_d + \Sigma$$

Therefore,

$$\mathcal{V} = \big\|\Sigma^{1/2}\tilde{\mathcal{V}}\Sigma^{1/2}\big\| = \big\|(\operatorname{tr}\Sigma)\Sigma + \Sigma^2\big\| = \lambda_1^2 + \lambda_1 \operatorname{tr}\Sigma \asymp \lambda_1 \operatorname{tr}\Sigma$$

The last equality follows since Σ is diagonal and λ_1 is the largest eigenvalue.

(ii) Cauchy-Schwarz inequality gives $\mathcal{M}_1 \le \mathcal{M}_2$. It suffices to show $\mathcal{M}_2 \lesssim C_\psi^2[\operatorname{tr}\Sigma]$ and in the Gaussian case $\mathcal{M}_1 \gtrsim \operatorname{tr}\Sigma$.

Sub-Gaussian upper bound for $\mathcal{M}_2$.

Since $\|XX^\top - \Sigma\| \le \max\left\{\|XX^\top\|, \|\Sigma\|\right\}$, in view of Lemma 37, we have

$$\mathcal{M}_2^2 \lesssim \mathbb{E}\max\big\{\|X\|^4, \lambda_1^2\big\} \lesssim C_\psi^4 \left[\operatorname{tr}\Sigma\right]^2 + \lambda_1^2 \lesssim C_\psi^4 \left[\operatorname{tr}\Sigma\right]^2$$

Here the last inequality follows since $C_\psi \ge 1$ and $\operatorname{tr}\Sigma \ge \lambda_1$.

Gaussian lower bound for $\mathcal{M}_1$. Assume $X \sim \mathcal{N}(0, \Sigma)$. Without loss of generality, assume Σ is diagonal: $\Sigma = \operatorname{diag}(\lambda_1, \ldots, \lambda_d)$.

- Since $\big\|XX^\top - \Sigma\big\| \ge \|X\|^2 - \|\Sigma\|$, we have

$$\mathbb{E}\,\big\|XX^\top - \Sigma\big\| \ge \mathbb{E}\,\|X\|^2 - \|\Sigma\| = \operatorname{tr}\Sigma - \lambda_1$$

- Let e_1 be the first standard basis vector. Since

$$\big\|XX^\top - \Sigma\big\| \ge \big|e_1^\top(XX^\top - \Sigma)e_1\big| = \big|X_1^2 - \lambda_1\big| = \lambda_1\left|\widetilde{X}_1^2 - 1\right|,$$

  we have

$$\mathbb{E}\,\big\|XX^\top - \Sigma\big\| \ge \lambda_1 \mathbb{E}\left|\widetilde{X}_1^2 - 1\right| \ge c_0\lambda_1$$

  for some constant $c_0 > 0$.

Combining the two bounds above gives

$$\mathbb{E}\,\big\|XX^\top - \Sigma\big\| \ge \max\left\{\operatorname{tr}\Sigma - \lambda_1, c_0\lambda_1\right\}$$

If $\lambda_1 \le \operatorname{tr}\Sigma/2$, then $\operatorname{tr}\Sigma - \lambda_1 \ge \frac{1}{2}\operatorname{tr}\Sigma$. If $\lambda_1 > \operatorname{tr}\Sigma/2$, then $c_0\lambda_1 > c_0 \operatorname{tr}\Sigma/2$. In either case, we have

$$\mathbb{E}\,\big\|XX^\top - \Sigma\big\| \gtrsim \operatorname{tr}\Sigma.$$

B Main proof ideas

This appendix provides a high-level overview for the proofs of the main results in Section 3. The analysis is organized around the tangent matrix $\mathcal{T}$, which converts Oja's subspace dynamics into a recursion consisting of a deterministic contraction, a centered stochastic fluctuation, and a higher-order remainder. Variance-sensitive control of this recursion first yields sharp convergence of the iterates. Once the iterates are localized near U_*, the same recursion admits a first-order linearization driven by the signal–tail interaction $Z_i Y_i^\top$. This linear representation is the main bridge from convergence to the lower bound, Gaussian approximation, and multiplier bootstrap.

We follow the organization of Sections 3.1–3.3. Appendix B.1 explains the sharp convergence and linearization arguments, including the auxiliary bounded-observation analysis, its transfer to the original sub-Gaussian model through joint truncation, and the separate treatment of the exactly rank-r case. Appendix B.2 shows how the linearization yields the global and row-wise Gaussian approximations, while Appendix B.3 develops the corresponding bootstrap linearization and conditional Gaussian approximation. Technical details are deferred to the subsequent appendices: Theorems 1–2 and Propositions 3–4 are proved in Appendix E, Theorems 5–6 in Appendix G, and Theorems 7–8 in Appendix H.

B.1 Sharp convergence and linearization

We first explain the core convergence and linearization arguments in an auxiliary bounded-observation setting, where

$$\|Y\|_2 \le C_{\psi,n}\sqrt{\lambda_{1\sim r}}, \qquad \|Z\|_2 \le C_{\psi,n}\sqrt{\lambda_{r+1\sim d}}$$

almost surely, with $C_{\psi,n} \asymp C_\psi\sqrt{\log n}$ (Assumption 3). We then explain how these arguments are transferred to the original sub-Gaussian model (Assumption 1) through a joint truncation of the signal and tail coordinates.

Convergence under bounded observations. The starting point is the exact tangent recursion (Lemma 2). After separating the conditional mean from the centered fluctuation, one Oja update can be written as

$$\mathcal{T}_{k+1} = (\mathcal{I} - \eta\mathcal{L})\mathcal{T}_k + D_{k+1} + R_{k+1}, \tag{B.1}$$

where $\mathcal{I} - \eta\mathcal{L}$ gives the deterministic contraction, D_{k+1} is conditionally centered, and R_{k+1} is a higher-order drift remainder.

Matrix Freedman's inequality, together with the control of the drift remainder, leads to a quadratic function

$$\varphi(\tau) = \bar{a}\tau^2 + \bar{b}\tau + \bar{c}$$

that describes the effective evolution of the error (Lemmas 3–5) starting from $\|\mathcal{T}_k\|$ of order τ. After sufficiently many iterations the deterministic contraction suppresses the initial error and $\|\mathcal{T}_k\|$ drops to order $\varphi(\tau)$. Repeating this argument drives the error toward the stable fixed-point of φ, which is comparable to the noise floor τ_*. The remaining argument verifies that the total number of iterations required to reach this error level is at most n (Lemmas 6–8), yielding the convergence guarantee in Theorem 1.

Linearization under bounded observations. Once the convergence argument has localized the iterates to $\|\mathcal{T}_k\| \lesssim \tau_*$, we revisit the one-step decomposition B.1 and isolate its leading stochastic term. Among the terms in the centered increment D_{k+1}, $\eta Z_{k+1}Y_{k+1}^\top$ is the only one that contains neither a factor of $\mathcal{T}_k$ nor an additional power of η. This suggests the first-order recursion

$$\mathcal{T}_{k+1} \approx (\mathcal{I} - \eta\mathcal{L})\mathcal{T}_k + \eta Z_{k+1}Y_{k+1}^\top,$$

and hence the linear representation

$$\mathcal{T}_n \approx \eta\sum_{i=1}^{n}(\mathcal{I} - \eta\mathcal{L})^{n-i} Z_i Y_i^\top.$$

Lemma 10 formalizes this approximation by splitting the trajectory at the end of the burn-in period. The pre-burn-in contribution is exponentially damped, while thereafter $\|\mathcal{T}_k\| \lesssim \tau_*$ makes the higher-order terms negligible relative to the leading fluctuation, yielding Proposition 3.

From bounded to sub-Gaussian observations. To transfer the preceding convergence and linearization arguments to Assumption 1, we jointly truncate the whitened signal $\widetilde{Y}$ and tail $\widetilde{Z}$ coordinates (Lemma 11). The truncated coordinates remain centered and sub-Gaussian and satisfy the almost-sure signal and tail bounds required by the bounded-observation analysis. The truncation slightly perturbs their joint covariance, by a matrix H with $\|H\| \ll 1$. Consequently, relative to the original basis $(U_*, U_{*,\perp})$, both diagonal covariance blocks are perturbed and the truncated signal and tail coordinates need no longer be uncorrelated.

Rather than re-diagonalizing the perturbed covariance, we retain the original contraction $\mathcal{I} - \eta\mathcal{L}$ and absorb the induced drift perturbation into the remainder. Lemma 12 shows that the resulting recursion satisfies the same one-step almost-sure, conditional-variance, and drift-remainder bounds as in the bounded setting, up to constant factors. Consequently, both the convergence and linearization arguments carry over to the truncated process; for the latter, the only additional term comes from the nonzero signal–tail cross-covariance, whose post-burn-in contribution is absorbed into the modified remainder and whose pre-burn-in contribution is exponentially damped. This yields Lemmas 13 and 15. Finally, the truncation coupling (Lemma 11) transfers these guarantees to the original sub-Gaussian observations, giving Theorem 1 and Proposition 3.

Exact-rank case. When $\operatorname{rank}(\Sigma) = r$, the tail coordinates Z vanish identically and $\tau_* = 0$, so the analysis above is replaced by a direct contraction argument. The tangent recursion reduces to

$$\mathcal{T}_{k+1} = \mathcal{T}_k (I_r + \eta Y_{k+1} Y_{k+1}^\top)^{-1},$$

which preserves the column space of $\mathcal{T}_k$ and makes $\|\mathcal{T}_k\|$ nonincreasing. This column-space preservation allows the matrix martingale arising over each contraction block to be compressed from dimension $(d-r) \times r$ to at most $r \times r$. Hence the dimension prefactor in matrix Freedman's inequality is at most $2r$, rather than of ambient order d. Combining the resulting one-block contraction with a block-chaining argument yields the geometric decay in Lemmas 9 and 14, and hence Theorem 2.

B.2 Gaussian approximation

Global Gaussian approximation. Proposition 3 reduces the distributional analysis of $\mathcal{T}_n$ to the weighted sum $\eta \sum_{i=1}^{n} (\mathcal{I} - \eta\mathcal{L})^{n-i} Z_i Y_i^\top$. After normalization by $\sqrt{\eta}$, this is a sum of independent random matrices. A high-dimensional Gaussian approximation, together with convergence of its finite-sample covariance, therefore gives a Gaussian approximation for $\eta^{-1/2}\|\mathcal{T}_n\|_{\mathrm{F}}$. Finally, the local equivalence between the tangent and sine errors transfers this approximation to $\eta^{-1}\|\sin\Theta(U_n, U_*)\|_{\mathrm{F}}^2$, yielding Theorem 5.

Row-wise Gaussian approximation. The same linearization also drives the row-wise result. After Procrustes alignment, the leading term of $U_n \operatorname{sgn}(U_n^\top U_*) - U_*$ is $U_{*,\perp}\mathcal{T}_n$. Projecting onto a fixed row therefore reduces the problem to a sum of independent r-dimensional random vectors, to which a multivariate Berry–Esseen bound applies. The convergence and linearization bounds make the remaining alignment terms negligible, yielding Theorem 6.

B.3 Multiplier bootstrap

Bootstrap linearization. The bootstrap tangent recursion has the same contractive structure as the original recursion, with the bounded multiplier weights affecting only the constants in the one-step bounds. The convergence and linearization arguments therefore extend to the bootstrap iterates, giving

$$\mathcal{T}_n^{\mathsf{bs}} - \mathcal{T}_n \approx \eta \sum_{i=1}^{n} (w_i^{\mathsf{bs}} - 1)(\mathcal{I} - \eta\mathcal{L})^{n-i} Z_i Y_i^\top.$$

Thus, conditionally on the data, the leading bootstrap fluctuation is obtained by attaching independent centered multipliers to the same summands that drive the sampling fluctuation in Proposition 3.

Conditional Gaussian approximation. Given the bootstrap linearization above, the remaining argument parallels the Gaussian approximation in Appendix B.2, now conditionally on the data. The linearized bootstrap error is a sum of independent centered multiplier terms, whose empirical covariance converges to the same limiting covariance as the sampling linearization. Conditional high-dimensional Gaussian approximation, together with a multivariate Berry–Esseen bound after row-wise projection, therefore yields the same global and row-wise Gaussian approximations as in Theorems 5 and 6. The same tangent-to-subspace and alignment arguments then transfer these approximations to the bootstrap statistics. Comparing the bootstrap and sampling Gaussian approximations yields Theorems 7 and 8.

C Analysis of Oja's algorithm: bounded observations

This appendix develops the core convergence and linearization analysis under Assumptions 3–5. In particular, the signal and tail coordinates Y and Z are almost surely bounded, the dimension grows at most polynomially with n, and the initial tangent matrix satisfies a deterministic bound. Under Assumptions 3–5, we establish convergence to the noise floor (Lemma 8) and linearization (Lemma 10) with polynomially small failure probability. We also treat the exactly rank-r case separately: when $\operatorname{rank}(\Sigma) = r$, the noise floor vanishes, and under Assumptions 3–4, we prove geometric contraction of the tangent error (Lemma 9). Appendix D extends these results to the original sub-Gaussian observations through truncation, and Appendix E uses them to prove the convergence results in Section 3.1.

Assumption 3 (Bounded signal and tail coordinates)**.** *$X_1, \dots, X_n \in \mathbb{R}^d$ are i.i.d. copies of $X = \Sigma^{1/2}\widetilde{X}$ where $\mathbb{E}\widetilde{X} = 0$, $\mathbb{E}\widetilde{X}\widetilde{X} = I_d$, $\|\widetilde{X}\|_{\psi_2} \le C_\psi \lesssim 1$. Moreover, almost surely,*

$$\left\|U_*^\top X\right\| \le C_{\psi,n}\sqrt{\lambda_{1\sim r}} \quad \text{and} \quad \left\|U_{*,\perp}^\top X\right\| \le C_{\psi,n}\sqrt{\lambda_{r+1\sim d}},$$

where $C_{\psi,n} := C_{\mathsf{noise}} C_\psi \sqrt{\log n}$ for some absolute constant $C_{\mathsf{noise}} > 0$.

Remark 6. It follows from the definition of $\|\cdot\|_{\psi_2}$ and $\mathbb{E}\widetilde{X}\widetilde{X}^\top = I_d$ that $C_\psi \ge 1$. Moreover, recalling from (6) that $Y = U_*^\top X$ and $Z = U_{*,\perp}^\top X$, the two almost-sure conditions in Assumption 3 can equivalently be written as

$$\|Y\| \le C_{\psi,n}\sqrt{\lambda_{1\sim r}} \quad \text{and} \quad \|Z\| \le C_{\psi,n}\sqrt{\lambda_{r+1\sim d}}.$$

Assumption 4 (Dimension)**.** *$d \lesssim n^{C_d}$ for some constant $C_d > 0$.*

Remark 7 (Role of Assumption 4). Assumption 4 is used to absorb the matrix-dimension prefactor in Freedman's inequality into logarithmic factors. For $(d-r) \times r$ matrix martingales, this prefactor is of order d, so $d \lesssim n^{C_d}$ implies $\log d \lesssim \log n$. The assumption could therefore be removed at the cost of replacing the relevant $\log n$ factors by $\log(n+d)$. Moreover, Assumption 4 imposes no additional restriction in Theorem 1. Indeed, conditions (9) and (11) imply

$$n^{1-2\epsilon} \gtrsim \nu_\infty d\, \tfrac{\log(1/\delta_p)}{\delta_p^2} \gtrsim \nu_\infty d \ge c_\nu n^{-C_\nu} d,$$

and hence $d \lesssim c_\nu^{-1} n^{1-2\epsilon+C_\nu}$. Thus the polynomial-growth condition on d follows automatically from the assumptions of Theorem 1.

Assumption 5 (Initialization)**.** *$\mathcal{T}_0$ exists and satisfies $\|\mathcal{T}_0\| \le C_{\mathsf{init}}\delta_p^{-1}\sqrt{rd\log(1/\delta_p)}$ for $\delta_p \in (0, \frac{1}{e})$ and some absolute constant $C_{\mathsf{init}} > 0$.*

Remark 8 (Role of Assumption 5). Assumption 5 is a deterministic good-initialization condition used in the intermediate convergence analysis. Its role is to control the initial tangent radius $\tau_0 = \|\mathcal{T}_0\|$, which enters the stability and epoch-chaining arguments and ensures that the iterates reach the target neighborhood within the available n updates. This assumption imposes no additional restriction in the main theorems under Haar initialization. Indeed, Lemma 41 shows that, for $U_0 \sim \operatorname{Haar}(\mathbb{O}_{d,r})$,

$$\|\mathcal{T}_0\| \lesssim \delta_p^{-1}\sqrt{rd\log(1/\delta_p)}$$

with probability at least $1 - O(\delta_p)$. Thus Assumption 5 is simply the deterministic event on which we carry out the subsequent analysis. In the exactly rank-r case, the geometric contraction itself does not require Assumption 5; it is only used to convert the bound proportional to $\|\mathcal{T}_0\|$ into the explicit rate in d, n, and δ_p.

Recursion and one-step bounds. For brevity, we write $\mathcal{T} = \mathcal{T}_i$, $\mathcal{T}_+ = \mathcal{T}_{i+1}$, $Y_+ = U_*^\top X_{i+1}$ and $Z_+ = U_{*,\perp}^\top X_{i+1}$.

Lemma 2 (Recurrence)**.** *If $\mathcal{T}$ is well-defined and $1 + \eta(Y_+^\top + Z_+^\top \mathcal{T})Y_+ \neq 0$, then $\mathcal{T}_+$ is well-defined and satisfies*

$$\mathcal{T}_+ = \mathcal{T} + \eta \frac{(-\mathcal{T}Y_+ + Z_+)(Y_+^\top + Z_+^\top \mathcal{T})}{1 + \eta(Y_+^\top + Z_+^\top \mathcal{T})Y_+} \tag{C.1}$$

Proof. See Appendix C.1. □

Let $\mathcal{F}_i := \sigma(U_0, X_1, \dots, X_i)$ and write $\mathbb{E}'[\,\cdot\,] := \mathbb{E}[\,\cdot\,|\mathcal{F}_i]$. For brevity, denote

$$\zeta_+ := (Y_+^\top + Z_+^\top \mathcal{T})\, Y_+ \in \mathbb{R}, \qquad F_+ := (-\mathcal{T}Y_+ + Z_+)(Y_+^\top + Z_+^\top \mathcal{T}) \in \mathbb{R}^{(d-r)\times r},$$

Then Lemma 2 gives

$$\mathcal{T}_+ - \mathcal{T} = \eta \frac{F_+}{1 + \eta\zeta_+} = \eta F_+ - \eta^2 \frac{\zeta_+ F_+}{1 + \eta\zeta_+}. \tag{C.2}$$

Since $\mathbb{E}'[F_+] = -\mathcal{L}\mathcal{T}$, the conditional drift can be decomposed as

$$\mathbb{E}'[\mathcal{T}_+ - \mathcal{T}] = -\eta\mathcal{L}\mathcal{T} + \underbrace{\eta^2\, \mathbb{E}'\left[-\frac{\zeta_+ F_+}{1 + \eta\zeta_+}\right]}_{=:R_+}. \tag{C.3}$$

Defining the centered one-step noise $D_+ := (\mathcal{T}_+ - \mathcal{T}) - \mathbb{E}'[\mathcal{T}_+ - \mathcal{T}]$, we obtain the one-step decomposition

$$\mathcal{T}_+ = (\mathcal{I} - \eta\mathcal{L})\mathcal{T} + D_+ + R_+. \tag{C.4}$$

The next lemma quantifies the size of D_+ and R_+ given $\|\mathcal{T}\| \le \tau$. The important point is that the martingale term admits a variance scale substantially sharper than its deterministic bound scale; this is what later makes a tight analysis possible.

For the linearization in Lemma 10, we can further decompose D_+ according to the two terms in (C.2):

$$D_+ = \underbrace{\eta\,(F_+ - \mathbb{E}'[F_+])}_{=:\,D_+^{(1)}} + \underbrace{\eta^2\left(-\frac{\zeta_+\,F_+}{1+\eta\,\zeta_+} + \mathbb{E}'\left[\frac{\zeta_+\,F_+}{1+\eta\,\zeta_+}\right]\right)}_{=:\,D_+^{(2)}}. \tag{C.5}$$

Recall $\kappa_* := \frac{\lambda_1}{\Delta_{\min}}$.

Lemma 3 (One step bound)**.** *Assume Assumption 3 holds. Assume $\mathcal{T}$ is well-defined with $\|\mathcal{T}\| = \tau$. If $C_{\psi,n}^2 \widetilde{\eta}\kappa_* r(1 + \tau\nu_2) \le \frac{1}{2}$, then the following hold.*

(I) *Noise D_+: $\|D_+\| \le 4C_{\psi,n}^2 \widetilde{\eta}\kappa_* r(1 + \tau\nu_2)(\tau + \nu_2)$ almost surely and*

$$\max\left\{\left\|\mathbb{E}' D_+ D_+^\top\right\|^{1/2}, \left\|\mathbb{E}' D_+^\top D_+\right\|^{1/2}\right\} \lesssim C_\psi^2 \widetilde{\eta}\kappa_* \sqrt{r}\,(1 + \tau\nu_\infty)\,(\tau + \bar{\nu})\,.$$

(II) *Higher-order noise $D_+^{(2)}$: $\|D_+^{(2)}\| \le 4C_{\psi,n}^4 \widetilde{\eta}^2\kappa_*^2 r^2 (1 + \tau\nu_2)^2(\tau + \nu_2)$ almost surely and*

$$\max\left\{\left\|\mathbb{E}' D_+^{(2)} D_+^{(2)\top}\right\|^{1/2}, \left\|\mathbb{E}' D_+^{(2)\top} D_+^{(2)}\right\|^{1/2}\right\} \lesssim C_\psi^4 \widetilde{\eta}^2\kappa_*^2 r^{3/2}\,(1 + \tau\nu_\infty)^2\,(\tau + \bar{\nu})\,.$$

(III) *Drift remainder R_+: $\|R_+\| \lesssim C_\psi^4 \widetilde{\eta}^2\kappa_*^2 r^{3/2}\,(1 + \tau\nu_\infty)^2\,(\tau + \nu_\infty)$.*

Proof. See Appendix C.2. □

$\mathcal{T}_k = (\mathcal{I} - \eta\mathcal{L})^k \mathcal{T}_0 + \mathcal{D}^{(k)} + \mathcal{R}^{(k)}$ where

$$\mathcal{D}_k = \sum_{i=1}^{k} (\mathcal{I} - \eta\mathcal{L})^{k-i} D_i \quad \text{and} \quad \mathcal{R}_k = \sum_{i=1}^{k} (\mathcal{I} - \eta\mathcal{L})^{k-i} R_i.$$

Define

$$\begin{aligned} \varphi_\sigma(\tau) &= \sqrt{\widetilde{\eta}}\rho_\sigma(1 + \tau\nu_\infty)(\tau + \bar{\nu}), \qquad &\text{where} \quad \rho_\sigma &= C_5 C_\psi^2 \kappa_* \sqrt{r \log n}, \\ \varphi_B(\tau) &= \widetilde{\eta}\rho_B(1 + \tau\nu_2)(\tau + \nu_2), \qquad &\text{where} \quad \rho_B &= C_5 C_\psi^2 \kappa_* r \log^2 n. \end{aligned} \tag{C.6}$$

Here $C_5 > 0$ is some constant to be determined later in Lemma 5.

Let $\varphi_\sigma \vee_{\mathrm{cw}} \varphi_B$ denote the coefficient-wise maximum of two quadratics. Specifically, after expanding φ_σ and φ_B as

$$\begin{aligned} \varphi_\sigma(\tau) &= a_\sigma\tau^2 + b_\sigma\tau + c_\sigma, \quad & a_\sigma &= \sqrt{\widetilde{\eta}}\rho_\sigma\nu_\infty, \quad & b_\sigma &= \sqrt{\widetilde{\eta}}\rho_\sigma(1 + \nu_\infty\bar{\nu}), \quad & c_\sigma &= \sqrt{\widetilde{\eta}}\rho_\sigma\bar{\nu}, \\ \varphi_B(\tau) &= a_B\tau^2 + b_B\tau + c_B, \quad & a_B &= \widetilde{\eta}\rho_B\nu_2, \quad & b_B &= \widetilde{\eta}\rho_B(1 + \nu_2^2), \quad & c_B &= \widetilde{\eta}\rho_B\nu_2. \end{aligned}$$

we let

$$\varphi(\tau) := (\varphi_\sigma \vee_{\mathrm{cw}} \varphi_B)(\tau) := \bar{a}\,\tau^2 + \bar{b}\,\tau + \bar{c}, \qquad \bar{a} = a_\sigma \vee a_B, \quad \bar{b} = b_\sigma \vee b_B, \quad \bar{c} = c_\sigma \vee c_B. \tag{C.7}$$

Since all coefficients are nonnegative, we have $\varphi(\tau) \geq \varphi_\sigma(\tau) \vee \varphi_B(\tau)$ for all $\tau \geq 0$. Define the *noise floor* τ_*

$$\tau_* := 3\bar{c} = 3\max\left(\sqrt{\widetilde{\eta}}\rho_\sigma\bar{\nu}, \widetilde{\eta}\rho_B\nu_2\right) \tag{C.8}$$

Before establishing convergence, we record several properties of the envelope φ that will be used repeatedly.

Lemma 4 (Properties of φ). *Let φ and τ_* be defined as in* (C.7) *and* (C.8). *If $\bar{a}, \bar{b}, \bar{c} > 0$ and $\max(\bar{a}\tau_0, \bar{b}) < 1/3$ for some $\tau_0 \geq \tau_*$, then the following hold.*

(i) (Fixed points) *The equation $\varphi(\tau) = \tau$ has two positive roots $\tau_- < \tau_+$ with*

$$\bar{c} < \tau_- < 3\bar{c}, \qquad \frac{1}{3\bar{a}} < \tau_+ < \frac{1}{\bar{a}}.$$

(ii) (Stability) $\tau_- < \tau_* \leq \tau_0 < \tau_+$.

(iii) (Contraction between $\tau_\pm$) *$\varphi(\tau) < \tau$ for $\tau \in (\tau_-, \tau_+)$.*

(iv) (Quantitative contraction) *For all $\tau \in [\tau_*, \tau_0]$: $\varphi(\tau) \leq \max\left\{(\bar{a}\tau_0 + 2\bar{b})\tau, \tau_*\right\}$.*

Proof. See Appendix C.3. □

Lemma 5 (One epoch bound). *Assume Assumption 3 and 4 hold. Assume $\|\mathcal{T}_0\| = \tau_0$. Let $\widetilde{\tau}_0 = \tau_0 \vee \tau_*$. If $\nu_\infty > 0$ and $\max\left\{\bar{a}\widetilde{\tau}_0, \bar{b}\right\} < 1/3$ and*

$$\widetilde{\eta}\kappa_* \in \left(0, \tfrac{1}{2}\right), \quad \widetilde{\eta}C_{\psi,n}^2\kappa_* r\,(1 + 2\widetilde{\tau}_0\nu_2) \leq \tfrac{1}{2}, \quad C_\psi^2\sqrt{\widetilde{\eta}}\,\kappa_* r\,(1 + 2\widetilde{\tau}_0\nu_\infty) \leq \log^{-3/2} n, \tag{C.9}$$

then there exists a constant $C_5 \geq 1$ large enough (only depending on C_d) such that, defining φ and τ_* as in (C.7) *and* (C.8)*, with probability exceeding $1 - n^{-20}$, the following properties hold simultaneously for all $k \in \{0, \ldots, n\}$:*

$$\text{(Concentration)} \quad \|\mathcal{D}_k\| + \|\mathcal{R}_k\| \leq \frac{\varphi(\widetilde{\tau}_0)}{2}, \tag{C.10}$$

$$\text{(Descent)} \quad \|\mathcal{T}_k\| \leq (1 - \tilde{\eta})^k\widetilde{\tau}_0 + \frac{\varphi(\widetilde{\tau}_0)}{2}. \tag{C.11}$$

$$\text{(Stability)} \quad \|\mathcal{T}_k\| \leq 2\widetilde{\tau}_0, \tag{C.12}$$

Proof. See Appendix C.4. □

Corollary 9. *Under the conditions of Lemma 5, if $\tau_0 > \tau_*$ and $\delta \geq \varphi(\tau_0)$, then, defining $N(\tau_0, \delta) := \lceil \log(2\tau_0/\delta)/\tilde{\eta} \rceil$, with probability at least $1 - n^{-20}$, $\|T_k\| \leq \delta$ for all $k \in [N(\tau_0, \delta), n]$.*

Proof. See Appendix C.5. □

Define the *epoch sequence*: $\bar{\tau}_0 := \|\mathcal{T}_0\| = \tau_0 \vee \tau_*$ and

$$\bar{\tau}_{j+1} := \varphi(\bar{\tau}_j) \vee \tau_* \quad \text{for } j = 0, 1, \dots \tag{C.13}$$

with $\bar{J}_* := \min\{j \geq 0 : \bar{\tau}_j \leq \tau_*\}$, epoch lengths $\bar{n}_j := \lceil \log(2\bar{\tau}_j/\bar{\tau}_{j+1})/\widetilde{\eta} \rceil$, and cumulative lengths $\bar{N}_0 := 0$, $\bar{N}_{j+1} := \bar{N}_j + \bar{n}_j$.

Define epochs to be $[\bar{N}_0, \bar{N}_1), [\bar{N}_1, \bar{N}_2), \dots, [\bar{N}_{\bar{J}_*-1}, \bar{N}_{\bar{J}_*})$ and $[\bar{N}_{\bar{J}_*}, \infty)$. We call $[\bar{N}_j, \bar{N}_{j+1})$ the j-th epoch. An epoch j is called nontruncated if $\varphi(\bar{\tau}_j) > \tau_*$, and truncated otherwise. By definition, the $(\bar{J}_* - 1)$-th epoch is truncated. Lemma 6 shows that all epochs $j \leq \bar{J}_* - 2$ are nontruncated. Moreover, if n falls in the final epoch $[\bar{N}_{\bar{J}_*}, \infty)$ (i.e. $n \geq N_{\bar{J}_*}$), then $\|\mathcal{T}_n\| \leq 2\tau_*$

Lemma 6 (Epoch chaining)**.** *Assume the conditions of Lemma 5. Suppose further that*

$$\max\left\{\bar{a}\bar{\tau}_0, \bar{b}\right\} < \frac{1}{3n^{\widetilde{\epsilon}} \log n}$$

for some constant $\widetilde{\epsilon} \geq 0$, then the following hold.

I (Epoch structure)

(a) $\bar{\tau}_j \downarrow \tau_*$, $\bar{J}_* < \infty$ *and* $\bar{\tau}_{\bar{J}_*} = \tau_*$.

(b) (Uniform progress) *Every nontruncated epoch j with $\varphi(\bar{\tau}_j) \geq \tau_*$ satisfies*

$$\log\left(\frac{\bar{\tau}_j}{\bar{\tau}_{j+1}}\right) \geq \widetilde{\epsilon} \log n + \log\log n.$$

In particular, each nontruncated epoch reduces $\log \bar{\tau}_j$ by at least $\widetilde{\epsilon} \log n + \log\log n$.

(c) (Epoch count) $\bar{J}_* \leq \log_+(\frac{\tau_0}{\tau_*})/(\tilde{\epsilon} \log n + \log\log n) + 1$.

(d) (Total iterations) $\bar{N}_{\bar{J}_*} \leq \widetilde{\eta}^{-1} \log_+(\frac{\tau_0}{\tau_*}) + \bar{J}_*[\widetilde{\eta}^{-1}(\log 2) + 1]$.

II ($\|\mathcal{T}_k\|$ bound) *With probability at least $1 - \bar{J}_* n^{-20}$, the following hold uniformly.*

$$\|\mathcal{T}_k\| \leq 2\bar{\tau}_j \quad \textit{for any } k \in [\bar{N}_j, n], \textit{ for any } j = 0, 1, \dots, \bar{J}_*.$$

Proof. See Appendix C.6. □

Lemma 7. *Assume the conditions of Lemma 5. Let $\eta = \frac{c_\eta \log n}{n\Delta_{\min}}$ for $c_\eta \geq \frac{1}{2}$. If $c_\eta \log n \geq C_{\text{bud}} \log^+(\sqrt{2c_\eta}\tau_0/\tau_*)$, where*

$$C_{\mathsf{bud}} = 1 + \left(\frac{1}{\widetilde{\epsilon}\log n + \log\log n} + \frac{1}{\log_+(\sqrt{2c_\eta}\,\tau_0/\tau_*)}\right)\left(\log 2 + \log^{-3} n\right). \tag{C.14}$$

Then:

(i) $\bar{N}_{\bar{J}_*} \leq n \cdot \frac{C_{\mathsf{bud}} \log_+(\sqrt{2c_\eta}\,\tau_0/\tau_*)}{c_\eta \log n} \leq n$.

(ii) $\tau_* \asymp \sqrt{\widetilde{\eta}}\rho_\sigma \bar{\nu}$.

Proof. See Appendix C.7. □

Lemma 8 (Convergence)**.** *Suppose Assumptions 3–5 hold. Assume*

$$\nu_\infty \geq c_\nu n^{-C_\nu} \quad \textit{for constants } C_\nu \geq 0, c_\nu > 0 \tag{C.15}$$

Let $\varepsilon \in [0, 1/2)$ and set

$$\eta = c_\eta \frac{\log n}{n\Delta_{\min}}, \quad \textit{where} \quad c_\eta \geq c_\eta^* := \left(1 - \varepsilon + \frac{3C_\nu}{2}\right)\left(1 + \frac{\log^{-3} n + \log 2}{\varepsilon \log n + \log\log n}\right). \tag{C.16}$$

If

$$n^{1-2\varepsilon} \geq C_8 c_\eta \kappa_*^2 r^3 (\log n)^4 \frac{\log(1/\delta_p)}{\delta_p^2} \max(1, \nu_\infty d) \tag{C.17}$$

for some constant $C_8 > 0$ large enough (depending only on C_ψ, C_ν, c_ν and C_d). Then with probability at least $1 - n^{-19}$,

$$\|\mathcal{T}_k\| \leq 2\tau_* \asymp \kappa_* \bar{\nu} \sqrt{\widetilde{\eta} r \log n} = \frac{\lambda_1}{\Delta_{\min}} \sqrt{c_\eta r}\, \bar{\nu} \frac{\log n}{\sqrt{n}}, \qquad \textit{for all } \left\lfloor \tfrac{c_\eta^*}{c_\eta} n \right\rfloor \leq k \leq n \tag{C.18}$$

Proof. See Appendix C.8. □

Lemma 9 (Exact-rank contraction). *Suppose Assumption 3 holds,* $\operatorname{rank}(\Sigma) = r$ *and $\mathcal{T}_0$ is well-defined. Fix $\varepsilon \in [0, 1/2)$ and set*

$$\eta = \frac{c_\eta \log n}{n \lambda_r}, \qquad \textit{where} \quad c_\eta \geq 1/2.$$

If

$$n^{1-2\varepsilon} \geq C_{\mathrm{rk}} c_\eta \kappa_*^2 r^2 (\log n)^4 \tag{C.19}$$

for some constant $C_{\mathrm{rk}} > 0$ sufficiently large (depending only on C_ψ), then, conditionally on $\mathcal{F}_0 = \sigma(U_0)$, with probability at least $1 - n^{-19}$,

$$\|\mathcal{T}_k\| \leq 16e \exp\left\{ -\left(1 - \frac{1}{n^\varepsilon \log n}\right) k\widetilde{\eta} \right\} \|\mathcal{T}_0\|, \qquad k = 0, 1, \ldots, n. \tag{C.20}$$

In particular,

$$\|\mathcal{T}_n\| \leq 16e\, n^{-c_\eta(1-(n^\varepsilon \log n)^{-1})} \|\mathcal{T}_0\|.$$

If, in addition, Assumption 5 holds and

$$n^{1-2\varepsilon} \geq C_{\mathrm{rk}} c_\eta \kappa_*^2 r^2 (\log n)^4 \frac{\log(1/\delta_p)}{\delta_p^2}, \tag{C.21}$$

then on the same event,

$$\|\mathcal{T}_k\| \lesssim \frac{\sqrt{d}\, n^{1/2-\varepsilon}}{\sqrt{c_\eta} \kappa_* \sqrt{r} (\log n)^2} \exp\left\{ -\left(1 - \frac{1}{n^\varepsilon \log n}\right) k\widetilde{\eta} \right\}, \qquad k = 0, 1, \ldots, n. \tag{C.22}$$

In particular,

$$\|\mathcal{T}_n\| \lesssim \frac{\sqrt{d}}{\sqrt{c_\eta} \kappa_* \sqrt{r} (\log n)^2} n^{1/2-\varepsilon-c_\eta(1-(n^\varepsilon \log n)^{-1})}.$$

Proof. See Appendix C.9. □

Lemma 10 (Linearization). *Under the assumptions of Lemma 8 with $c_\eta \geq c_\eta^* + \frac{1}{2}$, with probability exceeding $1 - n^{-18}$,*

$$\left\| \mathcal{T}_n - \eta \sum_{i=1}^{n} (\mathcal{I} - \eta \mathcal{L})^{n-i} (Z_i Y_i^\top) \right\| \lesssim \tau_* \left[C_5 C_\psi^2 \kappa_* \sqrt{r \widetilde{\eta} \log n} + \tau_* \right] \lesssim \frac{\tau_*}{n^\varepsilon \log n}.$$

Proof. See Appendix C.10. □

C.1 Proof of Lemma 2

Since $U_+ = \mathsf{QR}(\widetilde{U}_+)$, the ratio $\mathcal{T}_+ = (U_{*,\perp}^\top U_+)(U_*^\top U_+)^{-1}$ is invariant under the right factor, hence $\mathcal{T}_+ = \widetilde{\mathcal{S}}_+ \widetilde{\mathcal{C}}_+^{-1}$ where $\widetilde{\mathcal{C}}_+ = U_*^\top \widetilde{U}_+$ and $\widetilde{\mathcal{S}}_+ = U_{*,\perp}^\top \widetilde{U}_+$.

Since $X_+ = U_* Y_+ + U_{*,\perp} Z_+$ and $U = U_* \mathcal{C} + U_{*,\perp}\mathcal{S}$, we have

$$X_+^\top U = Y_+^\top \mathcal{C} + Z_+^\top \mathcal{S} = \underbrace{(Y_+^\top + Z_+^\top \mathcal{T})}_{=:v^\top \in \mathbb{R}^{1\times r}} \mathcal{C} \tag{C.23}$$

Substituting (C.23) into $\widetilde{U}_+ = (I_d + \eta X_+ X_+^\top)U$, we can obtain

$$\widetilde{\mathcal{C}}_+ = \mathcal{C} + \eta Y_+ X_+^\top U = \left[I_r + \eta Y_+ v^\top\right] \mathcal{C}$$
$$\widetilde{\mathcal{S}}_+ = \mathcal{S} + \eta Z_+ X_+^\top U = \left[\mathcal{T} + \eta Z_+ v^\top\right] \mathcal{C}$$

Hence

$$\mathcal{T}_+ = \widetilde{\mathcal{S}}_+ \widetilde{\mathcal{C}}_+^{-1} = \left[\mathcal{T} + \eta Z_+ v^\top\right] \left[I_r + \eta Y_+ v^\top\right]^{-1}$$

Note that the inverse can be computed by Sherman-Morrison formula. Specifically, if $1 + \eta v^\top Y_+ \neq 0$, then

$$\mathcal{T}_+ = \left[\mathcal{T} + \eta Z_+ v^\top\right] \left[I_r - \eta \frac{Y_+ v^\top}{1 + \eta v^\top Y_+}\right]$$

Expanding the product gives

$$\begin{aligned}
\mathcal{T}_+ &= \mathcal{T} - \eta \frac{\mathcal{T} Y_+ v^\top}{1 + \eta v^\top Y_+} + \eta Z_+ v^\top - \eta^2 \frac{Z_+ v^\top (v^\top Y_+)}{1 + \eta v^\top Y_+} \\
&= \mathcal{T} - \eta \frac{\mathcal{T} Y_+ v^\top}{1 + \eta v^\top Y_+} + \eta Z_+ v^\top \left[1 - \frac{\eta v^\top Y_+}{1 + \eta v^\top Y_+}\right] \\
&= \mathcal{T} - \eta \frac{\mathcal{T} Y_+ v^\top}{1 + \eta v^\top Y_+} + \eta \frac{Z_+ v^\top}{1 + \eta v^\top Y_+} \\
&= \mathcal{T} + \eta \frac{(-\mathcal{T} Y_+ + Z_+) v^\top}{1 + \eta v^\top Y_+}
\end{aligned}$$

Finally, plugging $v^\top$ from (C.23) back in gives the desired recursion formula.

C.2 Proof of Lemma 3

Recall (C.2) that

$$\mathcal{T}_+ - \mathcal{T} = \eta F_+ - \eta^2 \frac{\zeta_+ F_+}{1 + \eta \zeta_+}, \quad \text{where} \quad F_+ = (-\mathcal{T} Y_+ + Z_+)(Y_+^\top + Z_+^\top \mathcal{T}) \in \mathbb{R}^{(d-r)\times r}$$
$$\text{and} \quad \zeta_+ = (Y_+^\top + Z_+^\top \mathcal{T}) Y_+ \in \mathbb{R},$$

Define $V_1 = (-\mathcal{T} Y_+ + Z_+) \in \mathbb{R}^{d-r}$ and $V_2 = (Y_+ + \mathcal{T}^\top Z_+) \in \mathbb{R}^r$, then

$$F_+ = V_1 V_2^\top, \qquad \zeta_+ = V_2^\top Y_+.$$

Recall conditional drift B_+ and noise D_+ from (C.3)

$$B_+ = E'[\mathcal{T}_+ - \mathcal{T}] = -\eta \mathcal{L}\mathcal{T} + \underbrace{\eta^2\, \mathbb{E}'\left[-\frac{\zeta_+ F_+}{1 + \eta \zeta_+}\right]}_{=:R_+},$$

$$D_+ = (\mathcal{T}_+ - \mathcal{T}) - E'[\mathcal{T}_+ - \mathcal{T}] = \underbrace{\eta\,(F_+ - \mathbb{E}'[F_+])}_{=:\,D_+^{(1)}} + \underbrace{\eta^2 \left(-\frac{\zeta_+ F_+}{1 + \eta\,\zeta_+} + \mathbb{E}'\left[\frac{\zeta_+ F_+}{1 + \eta\,\zeta_+}\right]\right)}_{=:\,D_+^{(2)}}.$$

Recall $\nu_2 = \lambda_{r+1\sim d}^{1/2}/\lambda_{1\sim r}^{1/2}$ and $\nu_\infty = \lambda_{r+1}^{1/2}/\lambda_1^{1/2}$.

Assumption 3 yields

$$\|Y_+\|_2^2 \le C_{\psi,n}^2 \lambda_{1\sim r}, \qquad \|Z_+\|_2^2 \le C_{\psi,n}^2 \lambda_{r+1\sim d}. \tag{C.24}$$

which implies

$$\begin{aligned} \|V_1\|_2 &= \|-\mathcal{T}Y_+ + Z_+\|_2 \le \|\mathcal{T}\|_2 \|Y_+\|_2 + \|Z_+\|_2 \le C_{\psi,n}\lambda_{1\sim r}^{1/2}(\tau + \nu_2), \\ \|V_2\|_2 &= \left\|Y_+ + \mathcal{T}^\top Z_+\right\|_2 \le \|Y_+\|_2 + \|\mathcal{T}\|_2 \|Z_+\|_2 \le C_{\psi,n}\lambda_{1\sim r}^{1/2}(1 + \tau\nu_2). \end{aligned} \tag{C.25}$$

Since $|\zeta_+| = |V_2^\top Y_+| \le \|V_2\|_2 \|Y_+\|_2 \le C_{\psi,n}^2 \lambda_{1\sim r}(1+\tau\nu_2)$, the hypothesis $\eta C_{\psi,n}^2 \lambda_{1\sim r}(1+\tau\nu_2) \le 1/2$ gives

$$|\eta\zeta_+| \le \frac{1}{2}, \qquad \text{hence} \quad |1 + \eta\zeta_+| \ge \frac{1}{2}. \tag{C.26}$$

For later use, we also record the sub-Gaussian norms. Since $Y_+ = \Lambda_*^{1/2}\widetilde{Y}_+$ and $Z_+ = \Lambda_{*,\perp}^{1/2}\widetilde{Z}_+$ with $\|(\widetilde{Y}_+^\top, \widetilde{Z}_+^\top)^\top\|_{\psi_2} \le C_\psi$, we have $\|Y_+\|_{\psi_2} \le \lambda_1^{1/2} C_\psi$ and $\|Z_+\|_{\psi_2} \le \lambda_{r+1}^{1/2} C_\psi$, and thus

$$\begin{aligned} \|V_1\|_{\psi_2} &\le \|\mathcal{T}\| \, \|Y_+\|_{\psi_2} + \|Z_+\|_{\psi_2} \le C_\psi \lambda_1^{1/2}(\tau + \nu_\infty), \\ \|V_2\|_{\psi_2} &\le \|Y_+\|_{\psi_2} + \|\mathcal{T}\| \, \|Z_+\|_{\psi_2} \le C_\psi \lambda_1^{1/2}(1 + \tau\nu_\infty). \end{aligned} \tag{C.27}$$

Preliminary $[\mathbb{E}'\zeta^4]^{1/4}$**:** Since $\zeta_+ = \|Y_+\|^2 + Z_+^\top \mathcal{T} Y_+$, triangle inequality of L_4 norm gives

$$\left[E'\zeta_+^4\right]^{1/4} \le \left[E' \|Y_+\|^8\right]^{1/4} + \left[E'(Z_+^\top \mathcal{T} Y_+)^4\right]^{1/4} \lesssim C_\psi^2 \lambda_{1\sim r} + \left[E'(Z_+^\top \mathcal{T} Y_+)^4\right]^{1/4}$$

where the last inequality follows from Lemma 37. For the second term, Cauchy–Schwarz gives

$$\begin{aligned} \left[\mathbb{E}'(Z_+^\top \mathcal{T} Y_+)^4\right]^{1/4} &\le \left[\mathbb{E}' \left\|\mathcal{T}^\top Z_+\right\|^8\right]^{1/8} \cdot \left[\mathbb{E}' \|Y_+\|^8\right]^{1/8} \\ &\lesssim C_\psi \left[\operatorname{tr}(\mathcal{T}^\top \Lambda_{*,\perp} \mathcal{T})\right]^{1/2} \cdot C_\psi \left[\operatorname{tr} \Lambda_*\right]^{1/2} \\ &\overset{(i)}{\lesssim} C_\psi \lambda_{r+1}^{1/2} \sqrt{r}\tau \cdot C_\psi \lambda_{1\sim r}^{1/2} \\ &\le C_\psi^2 r \lambda_1 \tau \nu_\infty \end{aligned}$$

where (i) follows from Lemma 37. Therefore, we have

$$[\mathbb{E}'\zeta_+^4]^{1/4} \lesssim C_\psi^2 r \lambda_1 (1 + \tau\nu_\infty). \tag{C.28}$$

(I) Noise D_+**.** Recall definition

$$D_+ = (\mathcal{T}_+ - \mathcal{T}) - E'[\mathcal{T}_+ - \mathcal{T}], \quad \text{where} \quad \mathcal{T}_+ - \mathcal{T} = \eta\frac{F_+}{1 + \eta\zeta_+}, \quad F_+ = V_1 V_2^\top.$$

Deterministic noise bound: Equation (C.25) and (C.26) give

$$\|\mathcal{T}_+ - \mathcal{T}\| = \eta\|F_+\|/|1 + \eta\zeta_+| \le 2\eta\|F_+\| = 2\eta\|V_1\|\|V_2\| \le 2C_{\psi,n}^2 \eta\lambda_{1\sim r}(\tau + \nu_2)(1 + \tau\nu_2).$$

Since $\|D_+\| = \|(\mathcal{T}_+ - \mathcal{T}) - E'[\mathcal{T}_+ - \mathcal{T}]\| \le \|\mathcal{T}_+ - \mathcal{T}\| + \|\mathbb{E}'[\mathcal{T}_+ - \mathcal{T}]\|$, we obtain

$$\|D_+\| \le 4C_{\psi,n}^2 \eta\lambda_{1\sim r}(1 + \tau\nu_2)(\tau + \nu_2) \le 4C_{\psi,n}^2 \widetilde{\eta}\kappa_* r(1 + \tau\nu_2)(\tau + \nu_2).$$

Variance bound for D_+**:** By definition,

$$D_+ = (\mathcal{T}_+ - \mathcal{T}) - \mathbb{E}'(\mathcal{T}_+ - \mathcal{T}), \quad \text{with} \quad \mathcal{T}_+ - \mathcal{T} = \eta\frac{F_+}{1 + \eta\zeta_+}, \quad F_+ = V_1 V_2^\top.$$

The second-moment domination of the variance and (C.26) give

$$\mathbb{E}' D_+ D_+^\top \preceq \mathbb{E}'(\mathcal{T}_+ - \mathcal{T})(\mathcal{T}_+ - \mathcal{T})^\top \preceq 4\eta^2 \mathbb{E}' F_+ F_+^\top$$

$$\mathbb{E}' D_+^\top D_+ \preceq \mathbb{E}'(\mathcal{T}_+ - \mathcal{T})^\top(\mathcal{T}_+ - \mathcal{T}) \preceq 4\eta^2 \mathbb{E}' F_+^\top F_+$$

Row quadratic variation $\|\mathbb{E}' D_+ D_+^\top\|^{1/2}$. Since $F_+ F_+^\top = V_1 V_1^\top \|V_2\|^2$, for any unit vector $u \in \mathbb{R}^{d-r}$, Cauchy-Schwarz inequality yields

$$\begin{aligned}\left[u^\top E'(F_+ F_+^\top) u\right]^{1/2} = \left[E'\left((u^\top V_1)^2 \|V_2\|^2\right)\right]^{1/2} &\le \left[E'(u^\top V_1)^4\right]^{1/4} \left[E' \|V_2\|^4\right]^{1/4} \\ &\lesssim C_\psi \lambda_1^{1/2} (\tau + \nu_\infty) \underbrace{\left[E' \|V_2\|^4\right]^{1/4}}_{=:\widetilde{v}_{2,4}}\end{aligned}$$

where the last inequality follows from (C.27). To bound $\widetilde{v}_{2,4}$, note that $V_2 = Y_+ + \mathcal{T}^\top Z_+$ is a linear transformation of $\widetilde{X}_+$, hence Lemma 37 implies that $\widetilde{v}_{2,4} \lesssim [\operatorname{tr} \operatorname{cov}'(V_2)]^{1/2} \cdot \|\widetilde{X}_+\|_{\psi_2}$ where cov' denotes the covariance conditional on previous filtration (similar to $\mathbb{E}'$). Using inequality $\operatorname{cov}'(A+B) \preceq 2\operatorname{cov}'(A) + 2\operatorname{cov}'(B)$ and $(a+b)^{1/2} \le a^{1/2} + b^{1/2}$ for $a, b \ge 0$ then gives

$$\begin{aligned}[\operatorname{tr} \operatorname{cov}'(V_2)]^{1/2} &\lesssim \left[\operatorname{tr} \Lambda_* + \operatorname{tr} \mathcal{T}^\top \Lambda_{*,\perp} \mathcal{T}\right]^{1/2} = \left[\lambda_{1\sim r} + \left\|\Lambda_{*,\perp}^{1/2} \mathcal{T}\right\|_{\mathrm{F}}^2\right]^{1/2} \\ &\le \left[r\lambda_1 + \left\|\Lambda_{*,\perp}^{1/2}\right\|^2 \|\mathcal{T}\|_{\mathrm{F}}^2\right]^{1/2} \le \left[r\lambda_1 + \lambda_{r+1} \cdot r\tau^2\right]^{1/2} \\ &\le (r\lambda_1)^{1/2} \left[1 + \nu_\infty^2 \tau^2\right]^{1/2} \le (r\lambda_1)^{1/2} (1 + \nu_\infty \tau)\end{aligned}$$

Thus

$$\widetilde{v}_{2,4} \lesssim C_\psi r^{1/2} \lambda_1^{1/2} (1 + \tau\nu_\infty). \tag{C.29}$$

which implies

$$\|\mathbb{E}' F_+ F_+^\top\|^{1/2} \lesssim C_\psi^2 r^{1/2} \lambda_1 (\tau + \nu_\infty)(1 + \tau\nu_\infty)$$

and

$$\|\mathbb{E}' D_+ D_+^\top\|^{1/2} \lesssim C_\psi^2 \widetilde{\eta} \kappa_* \sqrt{r} (1 + \tau\nu_\infty)(\tau + \nu_\infty). \tag{C.30}$$

Column quadratic variation $\|\mathbb{E}' D_+^\top D_+\|^{1/2}$. Since $F_+^\top F_+ = \|V_1\|^2 V_2 V_2^\top$. For any unit vector $v \in \mathbb{R}^r$, we have

$$\left[v^\top E'(F_+^\top F_+) v\right]^{1/2} \le \left[E'(v^\top V_2)^4\right]^{1/4} \left[E' \|V_1\|^4\right]^{1/4} \lesssim C_\psi \lambda_1^{1/2} (1 + \tau\nu_\infty) \underbrace{\left[E' \|V_1\|^4\right]^{1/4}}_{=:\widetilde{v}_{1,4}}$$

where the last inequality follows from (C.27). Recall $V_1 = -\mathcal{T} Y_+ + Z_+$, similar arguments as row quadratic moment give $\widetilde{v}_{1,4} \lesssim C_\psi [\operatorname{tr} \operatorname{cov}'(V_1)]^{1/2}$ with

$$\begin{aligned}[\operatorname{tr} \operatorname{cov}'(V_1)]^{1/2} &\lesssim \left[\operatorname{tr} \mathcal{T} \Lambda_* \mathcal{T}^\top + \operatorname{tr} \Lambda_{*,\perp}\right]^{1/2} = \left[\left\|\Lambda_*^{1/2} \mathcal{T}^\top\right\|_{\mathrm{F}}^2 + \operatorname{tr} \Lambda_{*,\perp}\right]^{1/2} \\ &\le \left[\left\|\Lambda_*^{1/2}\right\|_{\mathrm{F}}^2 \|\mathcal{T}\|^2 + \lambda_{r+1\sim d}\right]^{1/2} \le \left[\lambda_{1\sim r} \tau^2 + \lambda_{r+1\sim d}\right]^{1/2} \\ &\le \lambda_{1\sim r}^{1/2} (\tau^2 + \nu_2^2)^{1/2} \le \lambda_{1\sim r}^{1/2} (\tau + \nu_2)\end{aligned}$$

which implies

$$\widetilde{v}_{1,4} \lesssim C_\psi \lambda_{1\sim r}^{1/2} (\tau + \nu_2).$$

Thus

$$\|\mathbb{E}' F_+^\top F_+\|^{1/2} \lesssim C_\psi^2 \lambda_1^{1/2} \lambda_{1\sim r}^{1/2} (1 + \tau\nu_\infty)(\tau + \nu_2) \le C_\psi^2 \lambda_1 r^{1/2} (1 + \tau\nu_\infty)(\tau + \nu_2)$$

and

$$\|\mathbb{E}' D_+^\top D_+\|^{1/2} \lesssim C_\psi^2 \widetilde{\eta} \kappa_* \sqrt{r} (1 + \tau\nu_\infty)(\tau + \nu_2). \tag{C.31}$$

Combining (C.30) and (C.31) gives the desired variance bound for D_+ with $\bar{\nu} = \max(\nu_\infty, \nu_2)$.

(II) Higher-order noise $D_+^{(2)}$. Recall definition

$$D_+^{(2)} = -\eta^2\left(\frac{\zeta_+ F_+}{1+\eta\,\zeta_+} - \mathbb{E}'\left[\frac{\zeta_+ F_+}{1+\eta\,\zeta_+}\right]\right), \quad \text{where} \quad F_+ = V_1 V_2^\top, \quad \zeta_+ = V_2^\top Y_+.$$

Deterministic bound: Equation (C.25) and (C.26) give

$$\begin{aligned}\left\|\frac{\eta^2\zeta_+ F_+}{1+\eta\,\zeta_+}\right\| &\le 2\eta^2\,|\zeta_+|\,\|F_+\| \le 2\eta^2\,|\zeta_+|\,\|V_1\|\,\|V_2\| \\ &\le 2\eta^2 C_{\psi,n}^2\;\lambda_{1\sim r}\,(1+\tau\nu_2)\cdot C_{\psi,n}^2\;\lambda_{1\sim r}(\tau+\nu_2)(1+\tau\nu_2) \\ &= 2C_{\psi,n}^4\eta^2\lambda_{1\sim r}^2(1+\tau\nu_2)^2(\tau+\nu_2)\end{aligned}$$

which implies

$$\|D_+^{(2)}\| \le 4C_{\psi,n}^4\eta^2\lambda_{1\sim r}^2(1+\tau\nu_2)^2(\tau+\nu_2) \le 4C_{\psi,n}^4\widetilde{\eta}^2\kappa_*^2 r^2(1+\tau\nu_2)^2(\tau+\nu_2).$$

Variance bound: By the second-moment domination of the variance and (C.26), we have

$$\begin{aligned}\mathbb{E}'\big[D_+^{(2)}(D_+^{(2)})^\top\big] &\preceq \mathbb{E}'\Big[\eta^4\frac{\zeta_+^2}{(1+\eta\zeta_+)^2}F_+F_+^\top\Big] \preceq 4\eta^4\mathbb{E}'\left[\zeta_+^2F_+F_+^\top\right] \\ \mathbb{E}'\big[(D_+^{(2)})^\top D_+^{(2)}\big] &\preceq \mathbb{E}'\Big[\eta^4\frac{\zeta_+^2}{(1+\eta\zeta_+)^2}F_+^\top F_+\Big] \preceq 4\eta^4\mathbb{E}'\left[\zeta_+^2F_+^\top F_+\right]\end{aligned}$$

Row quadratic variation $\|\mathbb{E}'D_+^{(2)}(D_+^{(2)})^\top\|^{1/2}$. Since $F_+F_+^\top = \|V_2\|^2\,V_1V_1^\top$, for any unit vector $u\in\mathbb{R}^{d-r}$, Cauchy-Schwarz inequality yields

$$\begin{aligned}\left[u^\top E'(\zeta_+^2F_+F_+^\top)u\right]^{1/2} &= \left[E'\zeta_+^2\|V_2\|^2(V_2^\top Y_+)^2(u^\top V_1)^2\right]^{1/2} \\ &\le \left[E'\zeta_+^4\right]^{1/4}\;\left[E'\left[\|V_2\|^4(u^\top V_1)^4\right]\right]^{1/4} \\ &\le \left[E'\zeta_+^4\right]^{1/4}\;[E'\|V_2\|^8]^{1/8}[E'(u^\top V_1)^8]^{1/8} \\ &\lesssim C_\psi^2 r\lambda_1(1+\tau\nu_\infty)\underbrace{[E'\|V_2\|^8]^{1/8}}_{=:\widetilde{v}_{2,8}}C_\psi\lambda_1^{1/2}(\tau+\nu_\infty)\end{aligned}$$

where the last inequality follows from (C.27) and (C.28). A similar argument as $\widetilde{v}_{2,4}$ (C.29) gives

$$\widetilde{v}_{2,8} \lesssim C_\psi r^{1/2}\lambda_1^{1/2}\left(1+\tau\nu_\infty\right). \tag{C.32}$$

which implies

$$\|\mathbb{E}'\left[\zeta_+^2F_+F_+^\top\right]\|^{1/2} \lesssim C_\psi^4\;r^{3/2}\lambda_1^2(1+\tau\nu_\infty)^2(\tau+\nu_\infty).$$

Thus

$$\|\mathbb{E}'D_+^{(2)}(D_+^{(2)})^\top\|^{1/2} \lesssim C_\psi^4\widetilde{\eta}^2\kappa_*^2 r^{3/2}(1+\tau\nu_\infty)^2(\tau+\nu_\infty).$$

Column quadratic variation $\|\mathbb{E}'(D_+^{(2)})^\top D_+^{(2)}\|^{1/2}$. Similar to the row quadratic variation, we have

$$\|\mathbb{E}'(D_+^{(2)})^\top D_+^{(2)}\|^{1/2} \lesssim C_\psi^4\widetilde{\eta}^2\kappa_*^2 r^{3/2}(1+\tau\nu_\infty)^2(\tau+\nu_2).$$

The proof is omitted.

Combining the bounds above gives the desired variance bound for $D_+^{(2)}$ with $\bar{\nu} = \max(\nu_\infty,\nu_2)$.

(III) Drift remainder R_+. By definition,

$$R_+ = -\eta^2\mathbb{E}'\left[\frac{\zeta_+F_+}{1+\eta\zeta_+}\right]$$

By Lemma 36, we have $\|E'W\| \le \|E'WW^\top\|^{1/2}$ for any random matrix W. Applying this with $W = \zeta_+ F_+/(1+\eta\zeta_+)$ and using $|1+\eta\zeta_+| \ge 1/2$ give

$$\|R_+\| \le \eta^2 \Big\| E'\Big[\frac{\zeta_+^2}{(1+\eta\zeta_+)^2} F_+ F_+^\top\Big]\Big\|^{1/2} \le 2\eta^2 \left\| E'[\zeta_+^2 F_+ F_+^\top]\right\|^{1/2}.$$

Since $F_+F_+^\top = \|V_2\|^2 V_1 V_1^\top$, for any unit $u \in \mathbb{R}^{d-r}$, we can obtain

$$\begin{aligned}
\left[u^\top E'(\zeta_+^2 F_+ F_+^\top) u\right]^{1/2} &\le \left[E'\zeta_+^4\right]^{1/4} \ \left[E'\left[\|V_2\|^4 (u^\top V_1)^4\right]\right]^{1/4} \\
&\le \left[E'\zeta_+^4\right]^{1/4} \ [E'\|V_2\|^8]^{1/8} [E'(u^\top V_1)^8]^{1/8} \\
&\lesssim C_\psi^2 r \lambda_1 (1+\tau\nu_\infty) \cdot C_\psi r^{1/2} \lambda_1^{1/2} (1+\tau\nu_\infty) \cdot C_\psi \lambda_1^{1/2} (\tau+\nu_\infty) \\
&= C_\psi^4 r^{3/2} \lambda_1^2 (1+\tau\nu_\infty)^2 (\tau+\nu_\infty)
\end{aligned}$$

where the last inequality follows from (C.28), (C.32) and (C.27). Therefore,

$$\|R_+\| \lesssim C_\psi^4 \widetilde{\eta}^2 \kappa_*^2 r^{3/2} (1+\tau\nu_\infty)^2 (\tau+\nu_\infty).$$

C.3 Proof of Lemma 4

(i) The fixed-point equation reads $\bar{a}\tau^2 - (1-\bar{b})\tau + \bar{c} = 0$. Since $\tau_0 \ge \tau_* = 3\bar{c}$: $9\bar{a}\bar{c} = 3\bar{a} \cdot \tau_* \le 3\bar{a}\tau_0 < 1$. Hence $4\bar{a}\bar{c} < 4/9 = (2/3)^2 < (1-\bar{b})^2$ (using $\bar{b} < 1/3$), giving positive discriminant. Since $\bar{a}, \bar{c} > 0$ and $1 - \bar{b} > 0$, both roots are positive. Vieta's formulas then give

$$\tau_+ + \tau_- = \frac{1-\bar{b}}{\bar{a}} \quad \text{and} \quad \tau_+\tau_- = \frac{\bar{c}}{\bar{a}}.$$

Since $\tau_+ \ge \tau_- \ge 0$, we then have

$$\frac{1-\bar{b}}{2\bar{a}} \le \tau_+ \le \frac{1-\bar{b}}{\bar{a}}$$

which combined with the equality $\tau_+\tau_- = \bar{c}/\bar{a}$ gives

$$\frac{\bar{c}}{1-\bar{b}} \le \tau_- \le \frac{2\bar{c}}{1-\bar{b}}$$

Plugging in $2/3 < 1 - \bar{b} < 1$ gives

$$\frac{1}{3\bar{a}} < \tau_+ < \frac{1}{\bar{a}}, \qquad \bar{c} < \tau_- < 3\bar{c}.$$

(ii) By assumption, $\tau_0 \ge \tau_*$. It suffices to prove $\tau_0 < \tau_+$ and $\tau_* > \tau_-$.

$\tau_* > \tau_-$ is implied by $\tau_- < 3\bar{c}$ from (i) since $\tau_* := 3\bar{c}$.

$\tau_0 < \tau_+$ follows from the assumption $\bar{a}\tau_0 < 1/3$ and $\tau_+ > 1/(3\bar{a})$ from (i).

(iii) $h(\tau) := \varphi(\tau) - \tau$ is a convex quadratic vanishing at $\tau_\pm$, hence $h < 0$ on (τ_-, τ_+).

(iv) Set $t_0 = \bar{c}/\bar{b}$. We have $t_0 > \tau_*$ since $\bar{c}/\bar{b} > 3\bar{c} = \tau_*$. We consider two cases:

Case 1 ($\tau \in [t_0, \tau_0]$): $\varphi(\tau)/\tau = \bar{a}\tau + \bar{b} + \bar{c}/\tau \le \bar{a}\tau + 2\bar{b}$ (using $\bar{c}/\tau \le \bar{b}$).

Case 2 ($\tau \in [\tau_*, t_0]$): we have $\varphi(\tau) \le 3\bar{c} = \tau_*$. This follows from an analysis of the quadratic and linear term in $\varphi(\tau) = \bar{a}\tau^2 + \bar{b}\tau + \bar{c}$. Linear term: $\bar{b}\tau \le \bar{c}$ (using $\bar{b} \le \bar{c}/\tau$). Quadratic term: $\bar{a}\tau^2 \le \bar{a}\bar{c}^2/\bar{b}^2 < \bar{c}$ where the last inequality follows from

$$\bar{b}^2 > \bar{a}\bar{c}$$

which we now prove. Without loss of generality, assume $C_5 = 1$. Note that $\nu_2 > 0$ and $\nu_\infty > 0$ otherwise $\bar{a} = 0$. Since $\bar{a}\bar{c} = \max_i a_i \cdot \max_j c_j$ is one of four products $a_i c_j$, it suffices to show $\bar{b}^2 > a_i c_j$ for all $(i,j) \in \{\sigma, B\}^2$.

Same-index: $\bar{b}^2 \ge b_i^2 \ge 4a_i c_i > a_i c_i$ (by AM–GM: $(1+x)^2 \ge 4x$).

Cross-index ($a_\sigma c_B$): $(1+\nu_\infty\bar{\nu})(1+\nu_2^2) \ge 2\nu_2 > \nu_\infty\nu_2$ since $\nu_\infty \le 1$.

Cross-index ($a_B c_\sigma$): note that $\bar{b}^2 \ge b_\sigma b_B$, hence it suffices to prove

$$(1+\nu_\infty\bar{\nu})(1+\nu_2^2) > \nu_2\bar{\nu}$$

- when $\bar{\nu} = \nu_\infty$ (so $\nu_\infty \ge \nu_2$, both ≤ 1), LHS $> 1 \ge \nu_\infty\nu_2 = \nu_2\bar{\nu}$;
- when $\bar{\nu} = \nu_2$, LHS $= 1 + \nu_2^2 + \nu_\infty\nu_2 + \nu_\infty\nu_2^3 > \nu_2^2 = \nu_2\bar{\nu}$.

C.4 Proof of Lemma 5

Note that $\nu_\infty > 0$ implies that $\bar{a}, \bar{b}, \bar{c} > 0$. Also, $\eta\lambda_1 = \widetilde{\eta}\kappa_*$ which is $\in (0, \frac{1}{2})$ by assumption, hence Lemma 36 can be applied.

The proof has two steps: Step 1 shows uniform concentration of the stopped accumulated noise and remainder, establishing the existence of C_5 via matrix Freedman's inequality. Step 2 converts this into (C.10)-(C.12) via induction.

Step 1: Multi-step concentration and construction of C_5. Recall the decomposition

$$\mathcal{T}_k = (\mathcal{I} - \eta\mathcal{L})^k \mathcal{T}_0 + \mathcal{D}_k + \mathcal{R}_k,$$
$$\text{where} \quad \mathcal{D}_k = \sum_{i=1}^{k} (\mathcal{I} - \eta\mathcal{L})^{k-i} D_i, \quad \mathcal{R}_k = \sum_{i=1}^{k} (\mathcal{I} - \eta\mathcal{L})^{k-i} R_i$$

Define the good event

$$\mathscr{G}_k(\widetilde{\tau}_0, 2) := \{\|\mathcal{T}_0\| \le \widetilde{\tau}_0\} \cap \bigcap_{i=1}^{k} \{\|\mathcal{T}_i\| \le 2\widetilde{\tau}_0\}, \qquad k \ge 0,$$

with the convention $\mathscr{G}_0(\widetilde{\tau}_0, 2) = \{\|\mathcal{T}_0\| \le \widetilde{\tau}_0\}$, and let $\mathbb{1}_k(\widetilde{\tau}_0)$ denote its indicator. Define the *stopped* accumulated noise and remainder for $k \in [n]$ as

$$\mathcal{D}_k^*(\widetilde{\tau}_0) := \sum_{i=1}^{k} (\mathcal{I} - \eta\,\mathcal{L})^{k-i} \mathbb{1}_{i-1}(\widetilde{\tau}_0) D_i,$$
$$\mathcal{R}_k^*(\widetilde{\tau}_0) := \sum_{i=1}^{k} (\mathcal{I} - \eta\,\mathcal{L})^{k-i}\, \mathbb{1}_{i-1}(\widetilde{\tau}_0)\, R_i\,,$$

and $\mathcal{D}_k^*(\widetilde{\tau}_0) = \mathcal{R}_k^*(\widetilde{\tau}_0) = 0$. On $\mathscr{G}_{k-1}(\widetilde{\tau}_0, 2)$ every indicator $\mathbb{1}_{i-1} = 1$ for $i \le k$, so the stopped processes coincide with their unstopped counterparts:

$$\mathcal{D}_k^*(\widetilde{\tau}_0) = \mathcal{D}_k \quad \text{and} \quad \mathcal{R}_k^*(\widetilde{\tau}_0) = \mathcal{R}_k \qquad \text{on } \mathscr{G}_{k-1}(\widetilde{\tau}_0, 2).$$

Claim 1 (Multi-step concentration)**.** *Assume assumptions of Lemma 5 hold. Then there exists a constant $C_5 > 0$ (only depending on C_d) such that*

$$\mathbb{P}\Big\{\|\mathcal{D}_k^*\| + \|\mathcal{R}_k^*\| \le \frac{\varphi(\widetilde{\tau}_0)}{2} \textit{ for all } k \in [n]\Big\} \ge 1 - n^{-20}$$

Proof. See Appendix C.4.1. ☐

Define event

$$\mathscr{E}_* := \left\{\|\mathcal{D}_k^*(\widetilde{\tau}_0, 2)\| + \|\mathcal{R}_k^*(\widetilde{\tau}_0, 2)\| \le \frac{\varphi(\widetilde{\tau}_0)}{2} \text{ for all } k \in [n]\right\}.$$

Claim 1 implies $\mathbb{P}(\mathscr{E}_*) \ge 1 - n^{-20}$.

Step 2: Induction. We show that on $\mathscr{E}_*$, (C.10)-(C.12) hold for every $0 \le k \le n$ by induction.

Base case ($k = 0$). Since $\mathcal{D}_0 = \mathcal{R}_0 = 0$ and $\|\mathcal{T}_0\| \le \widetilde{\tau}_0 \le 2\widetilde{\tau}_0$, (C.10)-(C.12) hold trivially.

Inductive step. Suppose (C.10)-(C.12) hold for all indices $m = 0, 1, \ldots, k-1$ on $\mathscr{E}_*$. Then event $\mathscr{G}_{k-1}(\widetilde{\tau}_0, 2)$ occurs. Thus, by definition of the stopped processes, we have

$$\mathcal{D}_k^* = \mathcal{D}_k, \qquad \mathcal{R}_k^* = \mathcal{R}_k \qquad \text{on } \mathscr{E}_*.$$

which implies (C.10) holds for $m = k$ on $\mathscr{E}_*$.

Triangle inequality then gives

$$\big\|\mathcal{T}_k - (\mathcal{I} - \eta\mathcal{L})^k \mathcal{T}_0\big\| \le \|\mathcal{D}_k\| + \|\mathcal{R}_k\| \le \|\mathcal{D}_k^*\| + \|\mathcal{R}_k^*\| \le \frac{\varphi(\widetilde{\tau}_0)}{2} \qquad \text{on } \mathscr{E}_*.$$

which is (C.11) for $m = k$.

Moreover, since $\|(\mathcal{I}-\eta\mathcal{L})^k\mathcal{T}_0\| \le (1-\eta\Delta_{\min})^k\tau_0 \le \tau_0 \vee \tau_* = \widetilde{\tau}_0$ by Lemma 36 and $\varphi(\widetilde{\tau}_0) \le \widetilde{\tau}_0$ (Property 4(iii)), we have

$$\|\mathcal{T}_k\| \le \widetilde{\tau}_0 + \frac{\varphi(\widetilde{\tau}_0)}{2} \le 2\widetilde{\tau}_0.$$

which is (C.12) for $m = k$.

This completes the induction and hence the proof.

C.4.1 Proof of Claim 1

For brevity, denote $\tau = 2\widetilde{\tau}_0$ and

$$\mathcal{D}_k^* = \mathcal{D}_k^*(\widetilde{\tau}_0), \qquad R_k^* = R_k^*(\widetilde{\tau}_0), \qquad \mathbb{1}_k = \mathbb{1}_k(\widetilde{\tau}_0).$$

(1) Bounds for $\mathcal{D}_k^*$. For $i \in [k]$, define

$$A_i^{(k)} := (\mathcal{I}-\eta\mathcal{L})^{k-i}\mathbb{1}_{i-1}D_i.$$

Then $\mathcal{D}_k^* = \sum_{i=1}^k A_i^{(k)}$, and $\big\{\sum_{i=1}^j A_i^{(k)}\big\}_{j=0}^k$ is an $\mathbb{R}^{(d-r)\times r}$-valued martingale. We can calculate

$$\begin{aligned}\max_{i\in[k]}\big\|A_i^{(k)}\big\| &= \max_{i\in[k]}\big\|(\mathcal{I}-\eta\mathcal{L})^{k-i}\mathbb{1}_{i-1}D_i\big\| \\ &\overset{(i)}{\le} \max_{i\in[k]}\|\mathbb{1}_{i-1}D_i\| \overset{(ii)}{\lesssim} C_{\psi,n}^2\widetilde{\eta}\kappa_* r(1+\tau\nu_2)(\tau+\nu_2) =: B_{\mathcal{D}}^*(\tau) \end{aligned} \tag{C.33}$$

and

$$\begin{aligned}&\max\Big\{\Big\|\sum_{i=1}^k \mathbb{E}_{i-1}A_i^{(k)}(A_i^{(k)})^\top\Big\|^{1/2}, \Big\|\sum_{i=1}^k \mathbb{E}_{i-1}(A_i^{(k)})^\top A_i^{(k)}\Big\|^{1/2}\Big\} \\ &\overset{(iii)}{\le} \Big[\sum_{i=1}^k (1-\widetilde{\eta})^{2(k-i)}\max\big\{\big\|\mathbb{E}_{i-1}\mathbb{1}_{i-1}D_iD_i^\top\big\|, \big\|\mathbb{E}_{i-1}\mathbb{1}_{i-1}D_i^\top D_i\big\|\big\}\Big]^{1/2} \\ &\overset{(iv)}{\lesssim} \Big[\sum_{i=1}^k (1-\widetilde{\eta})^{2(k-i)}\Big]^{1/2} C_\psi^2\widetilde{\eta}\kappa_*\sqrt{r}(1+\tau\nu_\infty)(\tau+\bar{\nu}) \\ &\lesssim C_\psi^2\widetilde{\eta}^{1/2}\kappa_*\sqrt{r}(1+\tau\nu_\infty)(\tau+\bar{\nu}) =: \sigma_{\mathcal{D}}^*(\tau)\end{aligned} \tag{C.34}$$

where (i), (iii) follows from Lemma 36; (ii), (iv) follows from Lemma 3.

In view of the matrix Freedman's inequality (Theorem 12), with probability exceeding $1 - n^{-21}$

$$\|\mathcal{D}_k^*\| \lesssim \sigma_{\mathcal{D}}^*(\tau)\sqrt{\log n} + B_{\mathcal{D}}^*(\tau)\log n \tag{C.35}$$

where the last line follows from the definition of φ_σ and φ_B in (C.6).

(2) Bound for $\mathcal{R}_k^*$. Applying Lemma 3 and Lemma 36, we can obtain

$$\begin{aligned}\|\mathcal{R}_k^*\| &\le \sum_{i=1}^k (1-\widetilde{\eta})^{k-i}\,\|\mathbb{1}_{i-1}R_i\| \\ &\lesssim \sum_{i=1}^k (1-\widetilde{\eta})^{k-i} C_\psi^4\widetilde{\eta}^2\kappa_*^2 r^{3/2}(\tau+\nu_\infty)(1+\tau\nu_\infty)^2 \\ &\le C_\psi^4\widetilde{\eta}\kappa_*^2 r^{3/2}(\tau+\nu_\infty)(1+\tau\nu_\infty)^2 =: B_{\mathcal{R}}^*(\tau)\end{aligned} \tag{C.36}$$

(3) Variance domination. Combining the above the bounds (C.35) and (C.36), we have with probability exceeding $1-n^{-21}$,

$$\|\mathcal{D}_k^*\| + \|\mathcal{R}_k^*\| \lesssim \sigma_{\mathcal{D}}^*(\tau)\sqrt{\log n} + B_{\mathcal{D}}^*(\tau)\log n + B_{\mathcal{R}}^*(\tau)$$

We next show that the remainder term $B_{\mathcal{R}}^*(\tau)$ is dominated by $\sigma_{\mathcal{D}}^*(\tau)\sqrt{\log n}$. To this end, note that

$$\sigma_{\mathcal{D}}^*(\tau)\sqrt{\log n} \gtrsim B_{\mathcal{R}}^*(\tau)\sqrt{\log n} \iff C_\psi^2\ \sqrt{\widetilde{\eta}}\,\kappa_* r\frac{(1+\tau\nu_\infty)(\tau+\nu_\infty)}{\tau+\bar{\nu}} \lesssim 1$$

For $\tau > 0$, we have $\tau + \nu_\infty \le \tau + \bar{\nu}$, thus

$$\frac{(1+\tau\nu_\infty)(\tau+\nu_\infty)}{\tau+\bar{\nu}} \le 1 + \tau\nu_\infty$$

Hence, a sufficient condition for $\sigma_{\mathcal{D}}^*(\tau)\sqrt{\log n} \gg B_{\mathcal{R}}^*(\tau)$ is

$$C_\psi^2\sqrt{\widetilde{\eta}}\kappa_* r(1+\tau\nu_\infty) \lesssim 1$$

which holds under (C.9).

(4) Defining φ. Combining the bounds for $\mathcal{D}_k^*$ and $\mathcal{R}_k^*$ and the conditions for variance domination, we have with probability exceeding $1-n^{-21}$,

$$\|\mathcal{D}_k^*\| + \|\mathcal{R}_k^*\| \lesssim \sigma_{\mathcal{D}}^*(\tau)\sqrt{\log n} + B_{\mathcal{D}}^*(\tau)\log n$$

Thus, if we take

$$\varphi(\widetilde{\tau}_0) := \left[C_5\sigma_{\mathcal{D}}^*(\tau)\sqrt{\log n}\right] \vee_{\mathrm{cw}} \left[C_5 B_{\mathcal{D}}^*(\tau)\log n\right]$$

for some sufficiently large constant $C_5 > 0$, then we have

$$\mathbb{P}\Big\{\|\mathcal{D}_k^*\| + \|\mathcal{R}_k^*\| \le \frac{\varphi(\widetilde{\tau}_0)}{2}\Big\} \ge 1 - n^{-21}$$

After plugging in the definition of $B_{\mathcal{D}}^*$ and $\sigma_{\mathcal{D}}^*$ from (C.33) and (C.34), we see that φ is exactly the function defined in (C.7).

Finally, a union bound over $k \in [n]$ gives the desired result.

C.5 Proof of Corollary 9

By Lemma 5, with probability at least $1-n^{-20}$,

$$\|\mathcal{T}_k\| \le (1-\widetilde{\eta})^k\tau_0 + \frac{\varphi(\tau_0)}{2}, \qquad k = 0, 1, \ldots, n.$$

For all $k \ge N(\tau_0,\delta)$, $(1-\widetilde{\eta})^k\tau_0 \le \delta/2$, while $\varphi(\tau_0)/2 \le \delta/2$ by assumption, hence $\|\mathcal{T}_k\| \le \delta$.

C.6 Proof of Lemma 6

(I) Epoch structure. Since $\nu_\infty > 0$, we have $\bar{a}, \bar{b}, \bar{c} > 0$. Moreover,

$$\max\left\{\bar{a}\bar{\tau}_0, \bar{b}\right\} < \frac{1}{3n^{\widetilde{\epsilon}}\log^+ n} < \frac{1}{3}$$

Hence the assumption in Lemma 4 is satisfied.

(a) Let $\varphi^{\circ j}$ be the j-th iterate of φ. Lemma 4(iii) implies that τ_- is a stable fixed point over (τ_-, τ_+) and the sequence $\{\varphi^{\circ j}(\bar{\tau}_0)\}_{j\ge 0}$ decreases monotonically to τ_-. Moreover, we have $\tau_- < \tau_* \le \bar{\tau}_0 < \tau_+$ from Lemma 4(ii), thus (a) follows.

(b) For any nontruncated epoch j with $\varphi(\bar{\tau}_j) \ge \tau_*$, Lemma 4(iv) gives

$$\bar{\tau}_{j+1} = \varphi(\bar{\tau}_j) \le (\bar{a}\bar{\tau}_0 + 2\bar{b})\bar{\tau}_j < \frac{1}{n^{\varepsilon}\log n}\bar{\tau}_j$$

Thus (b) follows.

(c) Note that

$$\sum_{j=0}^{\bar{J}_*-1} \log(\bar{\tau}_j/\bar{\tau}_{j+1}) = \log(\bar{\tau}_0/\bar{\tau}_{\bar{J}_*}) = \log(\tau_0/\tau_*) \leq \log_+(\tau_0/\tau_*)$$

Note that the j-th epoch must be nontruncated if $j \leq \bar{J}_* - 2$, thereby each contributes $\geq \widetilde{\epsilon}\log n + \log\log n$ to the sum above by (b). Thus, we have

$$\log_+(\tau_0/\tau_*) \geq \sum_{j=0}^{\bar{J}_*-2} \log(\bar{\tau}_j/\bar{\tau}_{j+1}) \geq (\bar{J}_* - 1)(\widetilde{\epsilon}\log n + \log\log n)$$

which implies

$$\bar{J}_* - 1 \leq \frac{\log_+(\tau_0/\tau_*)}{\widetilde{\epsilon}\log n + \log\log n}$$

and hence (c) follows.

(d) Since Corollary 9 implies that

$$n_j \leq \widetilde{\eta}^{-1} \log\left(\frac{\bar{\tau}_j}{\bar{\tau}_{j+1}}\right) + \widetilde{\eta}^{-1}\log 2 + 1$$

for $j = 0, ..., \bar{J}_* - 1$. Summing over j gives the bound on N_{J_*}.

(II) $\|\mathcal{T}_k\|$ **bound** follows from iteratively applying Lemma 5 and Corollary 9 over epochs $j = 0, 1, \ldots, \bar{J}_*$, and taking a union bound over the failure probabilities of each epoch.

C.7 Proof of Lemma 7

Under assumption $C_\psi^2 \sqrt{\widetilde{\eta}}\kappa_* r\,(1 + 2\widetilde{\tau}_0\nu_\infty) \leq \log^{-3/2} n$, we have

$$\widetilde{\eta} \leq \left[\frac{1}{(\log n)^{3/2} r C_\psi^2 \kappa_*}\right]^2 \overset{(i)}{\leq} \frac{1}{r^2\log^3 n} \tag{C.37}$$

where (i) follows from $\kappa_* \geq 1$ and $C_\psi \geq 1$.

(i) ($\bar{N}_{\bar{J}_*} \leq n$) Note that

$$n\widetilde{\eta} = c_\eta \log n = \frac{c_\eta \log n}{C_{\text{bud}} \log^+(\sqrt{2c_\eta}\,\tau_0/\tau_*)} C_{\text{bud}} \log_+(\sqrt{2c_\eta}\,\tau_0/\tau_*) \tag{C.38}$$

Then, by Lemma 6, we can obtain

$$\begin{aligned}
\bar{N}_{\bar{J}_*} &\leq \frac{\log_+(\tau_0/\tau_*)}{\widetilde{\eta}} + \left(\frac{\log_+(\tau_0/\tau_*)}{\widetilde{\epsilon}\log n + \log\log n} + 1\right)\left(\frac{\log 2}{\widetilde{\eta}} + 1\right) \\
&\overset{(a)}{\leq} \frac{\log_+(\sqrt{2c_\eta}\,\tau_0/\tau_*)}{\widetilde{\eta}} + \left(\frac{\log_+(\sqrt{2c_\eta}\,\tau_0/\tau_*)}{\widetilde{\epsilon}\log n + \log\log n} + 1\right)\left(\frac{\log 2}{\widetilde{\eta}} + 1\right) \\
&= n\frac{\log_+(\sqrt{2c_\eta}\,\tau_0/\tau_*)}{n\widetilde{\eta}}\left[1 + \left(\frac{1}{\widetilde{\epsilon}\log n + \log\log n} + \frac{1}{\log_+(\sqrt{2c_\eta}\,\tau_0/\tau_*)}\right)(\log 2 + \widetilde{\eta})\right] \\
&\overset{(b)}{\leq} n\frac{\log_+(\sqrt{2c_\eta}\,\tau_0/\tau_*)}{n\widetilde{\eta}}\left[1 + \left(\frac{1}{\widetilde{\epsilon}\log n + \log\log n} + \frac{1}{\log_+(\sqrt{2c_\eta}\,\tau_0/\tau_*)}\right)\left(\log 2 + \log^{-3} n\right)\right] \\
&\overset{(c)}{=} n\frac{\log_+(\sqrt{2c_\eta}\,\tau_0/\tau_*)}{c_\eta \log n} C_{\text{bud}}
\end{aligned}$$

where (a) is due to $2c_\eta \geq 1$, (b) results from (C.38) and (c) follows from the definition of C_{bud} (C.14).

(ii) ($\tau_* \asymp \sqrt{\widetilde{\eta}}\rho_\sigma\bar{\nu}$) By definition (C.8), $\tau_* := 3\max\{\sqrt{\widetilde{\eta}}\rho_\sigma\bar{\nu}, \widetilde{\eta}\rho_B\nu_2\}$. Thus it suffices to show $\widetilde{\eta}\rho_B\nu_2 \lesssim \sqrt{\widetilde{\eta}}\rho_\sigma\bar{\nu}$. Straightforward computation gives

$$\frac{\widetilde{\eta}\rho_B\nu_2}{\sqrt{\widetilde{\eta}}\rho_\sigma\bar{\nu}} = \sqrt{\widetilde{\eta}r}\log^{3/2} n \leq \frac{1}{\sqrt{r}}$$

where the last inequality follows from (C.37). The proof is then complete.

C.8 Proof of Lemma 8

In view of Lemma 7, it suffices to verify the conditions therein are satisfied. Note that by definition, $c_\eta \geq c_\eta^* \geq \frac{1}{2}$. The conditions to be verified are listed below:

(C1) $\widetilde{\eta}\kappa_* \in (0, 1/2)$.

(C2) $\max(\bar{a}\tau_0, \bar{b}) \leq 1/(3n^{\tilde{\epsilon}} \log n)$ for some $\tilde{\epsilon} \geq 0$.

(C3) $\widetilde{\eta}C_{\psi,n}^2 \kappa_* r(1 + 2\tau_0\nu_2) \leq 1/2$.

(C4) $C_\psi^2 \widetilde{\eta}^{1/2}\kappa_* r(1 + \tau_0\nu_\infty) \leq \log^{-3/2} n$.

(C5) $c_\eta \log n \geq C_{\text{bud}} \log^+(\sqrt{2c_\eta}\tau_0/\tau_*)$, where

$$C_{\mathsf{bud}} = 1+\Big(\frac{1}{\widetilde{\epsilon}\log n + \log\log n} + \frac{1}{\log_+(\sqrt{2c_\eta}\tau_0/\tau_*)}\Big)\left(\log 2 + \log^{-3} n\right)$$

Define

$$\begin{aligned}
\Phi &:= \frac{\sqrt{c_\eta}\kappa_* r^{3/2}\max(1, \sqrt{\nu_\infty d})\log^2 n}{\sqrt{n}} \frac{\sqrt{\log(1/\delta_p)}}{\delta_p} \\
&= \sqrt{\widetilde{\eta}}\kappa_* r^{3/2}\max(1, \sqrt{\nu_\infty d})(\log n)^{3/2}\frac{\sqrt{\log(1/\delta_p)}}{\delta_p}
\end{aligned}$$

Assumption C.17 then reads

$$\Phi \leq \frac{1}{C_8^{1/2} n^\varepsilon}. \tag{C.39}$$

(C1). Since $\eta\lambda_1 = \widetilde{\eta}\kappa_* = \frac{c_\eta \log n}{n}\kappa_*$, we have

$$\frac{\eta\lambda_1}{\Phi^2} \leq \frac{1}{\log^3 n}\frac{\delta_p^2}{\log(1/\delta_p)} \leq 1$$

where we have used $\kappa_* \geq 1$ and $\delta_p \in (0, 1/e)$. Thus,

$$\begin{aligned}
\eta\lambda_1 &\leq \frac{1}{C_8} \\
&< \frac{1}{2} \qquad \text{for } C_8 \text{ large enough.}
\end{aligned}$$

(C2). By definition, $\max(\bar{a}\tau_0, \bar{b}) = \max\{a_\sigma\tau_0, a_B\tau_0, b_\sigma, b_B\}$. We consider each term separately.

Term b_σ. By definition, $b_\sigma = \sqrt{\widetilde{\eta}}C_5 C_\psi^2 \kappa_* \sqrt{r}(1 + \nu_\infty\bar{\nu})\sqrt{\log n}$. Note that $\bar{\nu} = \max(\nu_2, \nu_\infty) \asymp \nu_2 + \nu_\infty$, we have

$$1 + \nu_\infty\bar{\nu} \lesssim 1 + \nu_\infty(\nu_2 + \nu_\infty) \leq 2 + \nu_\infty\sqrt{d}\nu_\infty \lesssim 1 \vee \sqrt{\nu_\infty d}$$

where the last two inequalities exploit $\nu_\infty \leq 1$ and $\nu_2 \leq \sqrt{d}\nu_\infty$. Thus, we have

$$\frac{b_\sigma}{\Phi} \lesssim C_5 C_\psi^2 \frac{1}{r\log n}\frac{\sqrt{\log(1/\delta_p)}}{\delta_p}$$

thus

$$\begin{aligned}
b_\sigma &\lesssim \frac{C_5 C_\psi^2}{C_8^{1/2}}\frac{1}{n^\varepsilon \log n} \\
&< \frac{1}{3n^\varepsilon \log n} \qquad \text{for } C_8 \text{ large enough.}
\end{aligned}$$

Term b_B. By definition, $b_B = \widetilde{\eta}C_5 C_\psi^2 \kappa_* r(\log n)^2(1 + \nu_2^2)$. Since

$$1 + \nu_2^2 \leq 1 + d\nu_\infty^2 \lesssim \max(1, d\nu_\infty)$$

we have

$$\frac{b_B}{\Phi^2} \lesssim C_5 C_\psi^2 \frac{1}{\log n}$$

which implies

$$\begin{aligned} b_B &\lesssim \frac{C_5 C_\psi^2}{C_8} \frac{1}{n^{2\varepsilon} \log n} \\ &< \frac{1}{3n^\varepsilon \log n} \qquad \text{for } C_8 \text{ large enough.} \end{aligned}$$

Term $a_\sigma \tau_0$. By definition,

$$\begin{aligned} a_\sigma \tau_0 &\lesssim \sqrt{\widetilde{\eta}} C_5 C_\psi^2 \kappa_* \sqrt{r \log n} \cdot \sqrt{rd \log(1/\delta_p)}\, \delta_p^{-1} \\ &\lesssim C_5 C_\psi^2 \Phi \log^{-1} n \\ &< \frac{1}{3n^\varepsilon \log n} \qquad \text{for } C_8 \text{ large enough.} \end{aligned}$$

Term $a_B \tau_0$. By definition and Assumption 5, we can obtain

$$\begin{aligned} a_B \tau_0 &\lesssim \widetilde{\eta} C_5 C_\psi^2 \kappa_* r (\log n)^2 \sqrt{d} \nu_\infty \cdot \sqrt{rd \log(1/\delta_p)}\, \delta_p^{-1} \\ &\lesssim C_5 C_\psi^2 \Phi^2 \log^{-1} n \lesssim \frac{C_5 C_\psi^2}{C_8} \frac{1}{n^{2\varepsilon} \log n} \\ &< \frac{1}{3n^\varepsilon \log n} \qquad \text{for } C_8 \text{ large enough.} \end{aligned}$$

(C3). It suffices to show

$$\widetilde{\eta} C_{\psi,n}^2 \kappa_* r \le 1/4 \qquad \text{and} \qquad \widetilde{\eta} C_{\psi,n}^2 \kappa_* r \cdot 2\tau_0 \nu_2 \le 1/4$$

First term.

$$\widetilde{\eta} C_{\psi,n}^2 \kappa_* r \lesssim \widetilde{\eta} C_\psi^2 (\log n) \kappa_* r \le \frac{1}{\Phi^2}$$

Thus $\widetilde{\eta} C_{\psi,n}^2 \kappa_* r < 1/4$ for C_8 large enough.

Second term. By Assumption 5, we have

$$\begin{aligned} \widetilde{\eta} C_{\psi,n}^2 \kappa_* r \cdot 2\tau_0 \nu_2 &\lesssim \widetilde{\eta} C_\psi^2 (\log n) \kappa_* r \cdot \sqrt{rd \log(1/\delta_p)}\, \delta_p^{-1} \sqrt{d} \nu_\infty \\ &= C_\psi^2 \widetilde{\eta} \kappa_* r^{3/2} d \nu_\infty (\log n) \sqrt{\log(1/\delta_p)}\, \delta_p^{-1} \\ &\le \Phi^2 \end{aligned}$$

Hence $\widetilde{\eta} C_{\psi,n}^2 \kappa_* r \cdot 2\tau_0 \nu_2 < 1/4$ for C_8 large enough.

(C4). It suffices to show

$$C_\psi^2\, \widetilde{\eta}^{1/2} \kappa_* r \le \frac{1}{2 \log^{3/2} n} \quad \text{and} \quad C_\psi^2\, \widetilde{\eta}^{1/2} \kappa_* r \tau_0 \nu_\infty \le \frac{1}{2 \log^{3/2} n}$$

First term. $C_\psi^2 \widetilde{\eta}^{1/2} \kappa_* r \log^{3/2} n \le \Phi < 1/2$ for C_8 large enough.

Second term.

$$\begin{aligned} C_\psi^2\, \widetilde{\eta}^{1/2} \kappa_* r \tau_0 \nu_\infty &\lesssim C_\psi^2\, \widetilde{\eta}^{1/2} \kappa_* r \sqrt{rd \log(1/\delta_p)}\, \delta_p^{-1} \nu_\infty \\ &\le C_\psi^2\, \widetilde{\eta}^{1/2} \kappa_* r^{3/2} \sqrt{d \nu_\infty} \sqrt{\log(1/\delta_p)}\, \delta_p^{-1} \le \Phi \log^{-3/2} n \\ &< \frac{1}{2 \log^{3/2} n} \qquad \text{for } C_8 \text{ large enough.} \end{aligned}$$

(C5). It suffices to show

$$\begin{aligned} c_\eta &\geq \frac{1}{\log n} C_{\mathrm{bud}} \log^+(\sqrt{2c_\eta}\tau_0/\tau_*) \\ &= \frac{1}{\log n}\left[\left(1+\frac{\log 2+\log^{-3} n}{\epsilon \log n+\log\log n}\right)\log^+(\sqrt{2c_\eta}\tau_0/\tau_*)+\log 2+\log^{-3} n\right] \end{aligned} \tag{C.40}$$

where the last line follows from the definition of C_{bud}.

By definition of τ_* (C.8), we have

$$\tau_* \geq 3\sqrt{\widetilde{\eta}}\rho_\sigma \nu_\infty = 3\sqrt{\widetilde{\eta}}\, C_5 C_\psi^2 \kappa_* \sqrt{r\log n}\,\nu_\infty$$

By (C.15), we have

$$\frac{\tau_*}{\sqrt{2c_\eta}} \geq \frac{3}{\sqrt{2}} C_5 C_\psi^2 \frac{\log n}{\sqrt{n}} \kappa_* \sqrt{r}\cdot c_\nu n^{-C_\nu} \tag{C.41}$$

By Assumption 5,

$$\begin{aligned} \tau_0 &\leq C_{\mathsf{init}} \delta_p^{-1}\sqrt{rd\log(1/\delta_p)} \\ &\leq C_{\mathsf{init}}\sqrt{r} n^{\frac{1}{2}-\varepsilon+\frac{C_\nu}{2}} \frac{1}{\kappa_* r(\log n)^2\sqrt{C_8 c_\eta c_\nu}} \end{aligned} \tag{C.42}$$

Here the last line follows from assumption (C.17) and (C.15), which together imply

$$n^{1-2\varepsilon} \geq C_8 c_\eta \kappa_*^2 r^2 \nu_\infty d(\log n)^4 \frac{\log(1/\delta_p)}{\delta_p^2} \geq C_8 c_\eta \kappa_*^2 r^2 (c_\nu n^{-C_\nu}) d(\log n)^4 \frac{\log(1/\delta_p)}{\delta_p^2}$$

giving, after rearrangement,

$$d\frac{\log(1/\delta_p)}{\delta_p^2} \leq n^{1-2\varepsilon+C_\nu}\frac{1}{C_8 c_\eta \kappa_*^2 r^2 c_\nu (\log n)^4}$$

Combining (C.41) and (C.42) gives

$$\frac{\sqrt{2c_\eta}\tau_0}{\tau_*} \lesssim C_8^{-1/2} n^{1-\varepsilon+\frac{3C_\nu}{2}} \log^{-3} n$$

Thus for C_8 large enough, we have

$$\log_+(\sqrt{2c_\eta}\tau_0/\tau_*) \leq \left(1+\frac{3C_\nu}{2}-\varepsilon\right)\log n - 3\log\log n$$

Hence

$$\begin{aligned} \text{RHS of (C.40)} &\leq \left[\left(1+\frac{3C_\nu}{2}-\varepsilon\right)-\frac{3\log\log n}{\log n}\right]\left(1+\frac{\log 2+\log^{-3} n}{\epsilon\log n+\log\log n}\right)+\frac{\log 2+\log^{-3} n}{\log n} \\ &\leq \left(1+\frac{3C_\nu}{2}-\varepsilon\right)\left(1+\frac{\log 2+\log^{-3} n}{\epsilon\log n+\log\log n}\right) =: c_\eta^*, \qquad \text{for } n\geq 4. \end{aligned}$$

where the last line follows from the fact that

$$\frac{3\log\log n}{\log n}\left(1+\frac{\log 2+\log^{-3} n}{\frac{1}{2}\log n+\log\log n}\right) \geq \frac{\log 2+\log^{-3} n}{\log n} \qquad \text{for } n\geq 4.$$

Therefore, we have

$$\frac{C_{\mathrm{bud}}\log^+(\sqrt{2c_\eta}\tau_0/\tau_*)}{c_\eta \log n} \leq \frac{c_\eta^*}{c_\eta} \leq 1$$

for $n \geq 4$ and C_8 large enough.

The proof is then complete.

C.9 Proof of Lemma 9

Step 0: Preliminaries. Since $\operatorname{rank}(\Sigma) = r$, we have $\Lambda_{*,\perp} = 0$. Therefore $Z_i = 0$ almost surely for all $i \in [n]$,

$$\nu_2 = \nu_\infty = \bar{\nu} = 0, \qquad \Delta_{\min} = \lambda_r, \qquad \widetilde{\eta} = \eta\lambda_r = c_\eta \frac{\log n}{n}, \qquad \mathcal{L}M = M\Lambda_*.$$

- *Monotonicity.* By Lemma 2, the tangent recursion reduces to

$$\mathcal{T}_{k+1} = \mathcal{T}_k - \frac{\eta \mathcal{T}_k Y_{k+1} Y_{k+1}^\top}{1 + \eta\|Y_{k+1}\|^2} = \mathcal{T}_k (I_r + \eta Y_{k+1} Y_{k+1}^\top)^{-1}, \tag{C.43}$$

  where the last equality follows from the Sherman-Morrison formula. Since $I_r + \eta Y_{k+1} Y_{k+1}^\top \succ 0$ and $\|(I_r + \eta Y_{k+1} Y_{k+1}^\top)^{-1}\| \le 1$, we have

$$\operatorname{col}(\mathcal{T}_{k+1}) = \operatorname{col}(\mathcal{T}_k), \qquad \|\mathcal{T}_{k+1}\| \le \|\mathcal{T}_k\|, \qquad k \ge 0. \tag{C.44}$$

- *Sample-size condition.* Define

$$\Phi_{\mathrm{rk}} := \frac{\sqrt{c_\eta}\, \kappa_* r (\log n)^2}{\sqrt{n}}.$$

  The sample-size condition is equivalent to

$$\Phi_{\mathrm{rk}} \le \frac{1}{\sqrt{C_{\mathrm{rk}}}\, n^{\varepsilon}}. \tag{C.45}$$

  For C_{rk} sufficiently large, this implies

$$C_{\psi,n}^2 \widetilde{\eta} \kappa_* r \le \frac{1}{2}, \qquad \eta\lambda_1 = \widetilde{\eta}\kappa_* \le \frac{1}{2}. \tag{C.46}$$

  Since $\nu_2 = 0$, the first inequality ensures that Lemma 3 applies at every step.

Step 1: Contraction over one block. For any start time $s \in \{0, \dots, n-1\}$ and every $t \ge 1$ such that $s + t \le n$, iterating (C.4) gives

$$\mathcal{T}_{s+t} = (\mathcal{I} - \eta\mathcal{L})^t \mathcal{T}_s + \sum_{i=s+1}^{s+t} (\mathcal{I} - \eta\mathcal{L})^{s+t-i} D_i + \sum_{i=s+1}^{s+t} (\mathcal{I} - \eta\mathcal{L})^{s+t-i} R_i. \tag{C.47}$$

Specializing Lemma 3 to $\nu_2 = \nu_\infty = \bar{\nu} = 0$ and using contraction $\|\mathcal{T}_{i-1}\| \le \tau_s$, where $\tau_s := \|\mathcal{T}_s\|$, gives

$$\begin{aligned} \|D_i\| &\lesssim C_{\psi,n}^2 \widetilde{\eta} \kappa_* r \tau_s, \\ \max\big\{\|\mathbb{E}_{i-1} D_i D_i^\top\|^{1/2}, \|\mathbb{E}_{i-1} D_i^\top D_i\|^{1/2}\big\} &\lesssim C_\psi^2 \widetilde{\eta} \kappa_* \sqrt{r} \tau_s, \\ \|R_i\| &\lesssim C_\psi^4 \widetilde{\eta}^2 \kappa_*^2 r^{3/2} \tau_s. \end{aligned} \tag{C.48}$$

With these bounds, we can bound the three terms in (C.47) as follows.

- Linear term. By Lemma 36, $\|(\mathcal{I} - \eta\mathcal{L})^j\| \le (1 - \widetilde{\eta})^j$, $j \ge 0$. Thus $\|(\mathcal{I} - \eta\mathcal{L})^t \mathcal{T}_s\| \le (1 - \widetilde{\eta})^t \tau_s$.
- Martingale term. Define $A_i^{(t)} := (\mathcal{I} - \eta\mathcal{L})^{s+t-i} D_i$ for $i = s+1, \dots, s+t$. Then $\max_i \|A_i^{(t)}\| \lesssim C_{\psi,n}^2 \widetilde{\eta} \kappa_* r \tau_s$. Moreover,

$$\begin{aligned} &\max\Big\{\Big\|\sum_{i=s+1}^{s+t} \mathbb{E}_{i-1} A_i^{(t)} A_i^{(t)\top}\Big\|^{1/2}, \Big\|\sum_{i=s+1}^{s+t} \mathbb{E}_{i-1} A_i^{(t)\top} A_i^{(t)}\Big\|^{1/2}\Big\} \\ &\lesssim C_\psi^2 \widetilde{\eta} \kappa_* \sqrt{r}\, \tau_s \Big[\sum_{j=0}^{\infty} (1 - \widetilde{\eta})^{2j}\Big]^{1/2} \\ &\lesssim C_\psi^2 \kappa_* \sqrt{r \widetilde{\eta}}\, \tau_s. \end{aligned}$$

We now exploit the multiplicative structure in (C.43), which is specific to the exact-rank case, to reduce the dimension factor in matrix Freedman's inequality from the ambient dimension d to at most $2r$. This removes the need for Assumption 4. Writing

$$K_i := \frac{Y_i Y_i^\top}{1+\eta\|Y_i\|^2}, \qquad i \in [n],$$

we have $\mathcal{T}_i - \mathcal{T}_{i-1} = -\eta \mathcal{T}_{i-1} K_i$, and hence $D_i = -\eta \mathcal{T}_{i-1}\big(K_i - \mathbb{E}K_i\big)$, which implies $\mathrm{col}(D_i) \subseteq \mathrm{col}(\mathcal{T}_s)$. Moreover, since $\mathcal{L}D_i = D_i\Lambda_*$ in the exact rank-r case, $(\mathcal{I}-\eta\mathcal{L})^j D_i = D_i(I_r - \eta\Lambda_*)^j$. Right multiplication cannot enlarge the column space, so

$$\mathrm{col}(A_i^{(t)}) \subseteq \mathrm{col}(\mathcal{T}_s), \qquad i = s+1, \dots, s+t.$$

Let

$$q_s := \mathrm{rank}(\mathcal{T}_s) \le r.$$

If $q_s = 0$, then $\mathcal{T}_s = 0$, and monotonicity implies $\mathcal{T}_k = 0$ for all $k \ge s$, so the claimed bound is immediate. Suppose henceforth that $q_s \ge 1$, and let $V_s \in \mathbb{O}_{d-r,q_s}$ have columns spanning $\mathrm{col}(\mathcal{T}_s)$. Conditionally on $\mathcal{F}_s$, define

$$\widetilde{A}_i^{(t)} := V_s^\top A_i^{(t)} \in \mathbb{R}^{q_s \times r}.$$

Then

$$A_i^{(t)} = V_s \widetilde{A}_i^{(t)}, \qquad \|A_i^{(t)}\| = \|\widetilde{A}_i^{(t)}\|,$$

and

$$\Big\|\sum_{i=s+1}^{s+t} A_i^{(t)}\Big\| = \Big\|\sum_{i=s+1}^{s+t} \widetilde{A}_i^{(t)}\Big\|.$$

Furthermore,

$$\begin{aligned}
\Big\|\sum_{i=s+1}^{s+t} \mathbb{E}_{i-1}\widetilde{A}_i^{(t)}\widetilde{A}_i^{(t)\top}\Big\| &= \Big\|V_s^\top\big[\sum_{i=s+1}^{s+t} \mathbb{E}_{i-1}A_i^{(t)}A_i^{(t)\top}\big]V_s\Big\| \\
&\le \Big\|\sum_{i=s+1}^{s+t} \mathbb{E}_{i-1}A_i^{(t)}A_i^{(t)\top}\Big\|,
\end{aligned}$$

while, because $\mathrm{col}(A_i^{(t)}) \subseteq \mathrm{col}(V_s)$,

$$\widetilde{A}_i^{(t)\top}\widetilde{A}_i^{(t)} = A_i^{(t)\top}A_i^{(t)}.$$

Thus the same increment and quadratic-variation bounds hold for the $q_s \times r$ martingale $\{\sum_{i=s+1}^{j}\widetilde{A}_i^{(t)}\}_{j=s}^{s+t}$. Moreover, for C_{rk} sufficiently large, (C.19) implies $r \le n$, and hence $q_s + r \le 2r \le 2n$. Therefore, matrix Freedman's inequality gives, with probability at least $1 - n^{-20}$,

$$\sup_{1\le t\le n-s}\Big\|\sum_{i=s+1}^{s+t} A_i^{(t)}\Big\| = \sup_{1\le t\le n-s}\Big\|\sum_{i=s+1}^{s+t} \widetilde{A}_i^{(t)}\Big\| \lesssim \big[C_\psi^2\kappa_*\sqrt{r\widetilde{\eta}\log n} + C_{\psi,n}^2\widetilde{\eta}\kappa_* r\log n\big]\tau_s.$$

- Remainder term. (C.48) gives

$$\sup_{1\le t\le n-s}\Big\|\sum_{i=s+1}^{s+t}(\mathcal{I}-\eta\mathcal{L})^{s+t-i}R_i\Big\| \lesssim C_\psi^4\widetilde{\eta}^2\kappa_*^2 r^{3/2}\tau_s\sum_{j=0}^{\infty}(1-\widetilde{\eta})^j \lesssim C_\psi^4\widetilde{\eta}\kappa_*^2 r^{3/2}\tau_s.$$

Combining the three bounds, we obtain that, conditionally on $\mathcal{F}_s$, with probability at least $1 - n^{-20}$,

$$\|\mathcal{T}_{s+t}\| \le [(1-\widetilde{\eta})^t + b_{\mathrm{rk}}]\tau_s, \qquad 0 \le t \le n-s, \tag{C.49}$$

where

$$b_{\mathrm{rk}} := C\big[C_\psi^2\kappa_*\sqrt{r\widetilde{\eta}\log n} + C_{\psi,n}^2\widetilde{\eta}\kappa_* r\log n + C_\psi^4\widetilde{\eta}\kappa_*^2 r^{3/2}\big].$$

We next bound b_{rk}. Since $\widetilde{\eta} = c_\eta(\log n)/n$, we have

$$\kappa_*\sqrt{r\widetilde{\eta}\log n} = \frac{\Phi_{\mathrm{rk}}}{\sqrt{r}\,\log n},$$

$$\widetilde{\eta}\kappa_* r(\log n)^2 = \frac{\Phi_{\rm rk}^2}{\kappa_* r\log n},$$
$$\widetilde{\eta}\kappa_*^2 r^{3/2} = \frac{\Phi_{\rm rk}^2}{\sqrt{r}\,(\log n)^3}.$$

Since $\kappa_*, r \geq 1$ and $C_{\psi,n}^2 \lesssim C_\psi^2 \log n$, it follows from (C.45) that

$$b_{\rm rk} \lesssim \frac{\Phi_{\rm rk}}{\log n} + \frac{\Phi_{\rm rk}^2}{\log n} \lesssim \frac{1}{\sqrt{C_{\rm rk}}\, n^\varepsilon \log n}. \tag{C.50}$$

Step 2: Block chaining. Set $m_\eta := \lceil (\log 4)/\widetilde{\eta}\rceil$. Then

$$m_\eta \widetilde{\eta} \geq \log 4, \tag{C.51}$$

which combined with $1-\widetilde{\eta} \leq e^{-\widetilde{\eta}}$ gives

$$(1-\widetilde{\eta})^{m_\eta} \leq e^{-m_\eta\widetilde{\eta}} \leq \frac{1}{4}.$$

Define

$$\rho_{\rm rk} := (1-\widetilde{\eta})^{m_\eta} + b_{\rm rk}.$$

By (C.50), taking $C_{\rm rk}$ sufficiently large also ensures $b_{\rm rk} \leq 1/4$, and hence

$$\rho_{\rm rk} \leq \frac{1}{2}.$$

Applying the one-block bound (C.49) successively at the deterministic block endpoints $s_j = jm_\eta$, and using the tower property and a union bound over at most n blocks, we obtain, conditionally on $\mathcal{F}_0$, with probability at least $1-n^{-19}$,

$$\|\mathcal{T}_{s_j}\| \leq \rho_{\rm rk}^j \|\mathcal{T}_0\|, \qquad \forall s_j \leq n.$$

For arbitrary $k \leq n$, let $j = \lfloor k/m_\eta \rfloor$. By monotonicity,

$$\|\mathcal{T}_k\| \leq \|\mathcal{T}_{jm_\eta}\| \leq \rho_{\rm rk}^j \|\mathcal{T}_0\|.$$

Since $j \geq k/m_\eta - 1$ and $\rho_{\rm rk} \in (0,1)$,

$$\rho_{\rm rk}^j \leq \rho_{\rm rk}^{-1} \exp\left\{\frac{k}{m_\eta}\log\rho_{\rm rk}\right\} = \rho_{\rm rk}^{-1}\exp\{-c_{\rm rk}k\widetilde{\eta}\},$$

where

$$c_{\rm rk} := -\frac{\log \rho_{\rm rk}}{m_\eta \widetilde{\eta}}.$$

To control the prefactor $\rho_{\rm rk}^{-1}$, note that $\widetilde{\eta} \leq 1/2$ and $\log(1-\widetilde{\eta}) \geq -\frac{\widetilde{\eta}}{1-\widetilde{\eta}} \geq -2\widetilde{\eta}$. Moreover, $m_\eta\widetilde{\eta} \leq \log 4 + \widetilde{\eta} \leq \log 4 + \frac{1}{2}$. Hence

$$\rho_{\rm rk} \geq (1-\widetilde{\eta})^{m_\eta} \geq e^{-2m_\eta\widetilde{\eta}} \geq \frac{1}{16e}. \tag{C.52}$$

Thus $\rho_{\rm rk}^{-1} \leq 16e$, and therefore

$$\|\mathcal{T}_k\| \leq 16e\exp\{-c_{\rm rk}k\widetilde{\eta}\}\|\mathcal{T}_0\|, \qquad k \leq n. \tag{C.53}$$

Step 3: Lower bounding $c_{\rm rk}$. By definition,

$$c_{\rm rk} = -\frac{\log[(1-\widetilde{\eta})^{m_\eta} + b_{\rm rk}]}{m_\eta\widetilde{\eta}} = \frac{-\log(1-\widetilde{\eta})}{\widetilde{\eta}} - \frac{1}{m_\eta\widetilde{\eta}}\log\left(1 + \frac{b_{\rm rk}}{(1-\widetilde{\eta})^{m_\eta}}\right)$$
$$\geq 1 - \frac{b_{\rm rk}}{m_\eta\widetilde{\eta}\,(1-\widetilde{\eta})^{m_\eta}},$$

where the last line follows from $-\log(1-x) \geq x$ for $x \in (0,1)$ and $\log(1+x) \leq x$ for $x \geq 0$.

Using (C.51) and (C.52), we obtain

$$c_{\rm rk} \geq 1 - \frac{16e}{\log 4} b_{\rm rk}.$$

By (C.50), choosing $C_{\rm rk}$ sufficiently large yields

$$c_{\rm rk} \geq 1 - \frac{1}{n^\varepsilon \log n}.$$

Combining this with (C.53) proves (C.20).

Additional conclusion under Assumption 5. If, in addition, Assumption 5 and (C.21) hold, then

$$\|\mathcal{T}_0\| \le C_{\mathrm{init}}\delta_p^{-1}\sqrt{rd\log(1/\delta_p)} \lesssim \frac{\sqrt{d}\,n^{1/2-\epsilon}}{\sqrt{c_\eta}\kappa_*\sqrt{r}(\log n)^2}.$$

Combining this with (C.20) gives

$$\|\mathcal{T}_k\| \lesssim \frac{\sqrt{d}\,n^{1/2-\epsilon}}{\sqrt{c_\eta}\kappa_*\sqrt{r}(\log n)^2}\exp\left\{-\left(1-\frac{1}{n^\epsilon\log n}\right)k\widetilde{\eta}\right\}.$$

Since $n\widetilde{\eta} = c_\eta \log n$, the stated bound for $\mathcal{T}_n$ follows.

C.10 Proof of Lemma 10

Denote $N_* = \lfloor nc_\eta^*/c_\eta \rfloor$. Lemma 8 implies that

$$\mathbb{P}(\mathscr{G}^c) \le n^{-19}, \qquad \text{where } \mathscr{G} := \{\|\mathcal{T}_k\| \le 2\tau_* \text{ for all } k \in [N_*, n]\}.$$

Decomposing $\mathcal{T}_n$ from $\mathcal{T}_{N_*}$ gives

$$\mathcal{T}_n = (\mathcal{I}-\eta\mathcal{L})^{n-N_*}\mathcal{T}_{N_*} + \mathcal{D}^{(1,0)}_{N_*,n} + \mathcal{D}^{(1,1)}_{N_*,n} + \mathcal{D}^{(2)}_{N_*,n} + \mathcal{R}_{N_*,n}$$

where

$$\mathcal{D}^{(1,0)}_{N_*,n} := \eta\sum_{i=N_*+1}^{n}(\mathcal{I}-\eta\mathcal{L})^{n-i}(Z_iY_i^\top)$$

$$\mathcal{D}^{(1,1)}_{N_*,n} := \eta\sum_{i=N_*+1}^{n}(\mathcal{I}-\eta\mathcal{L})^{n-i}\left[-\mathcal{T}_{i-1}(Y_iY_i^\top-\Lambda_*) + (Z_iZ_i^\top - \Lambda_{*,\perp})\mathcal{T}_{i-1} - \mathcal{T}_{i-1}Y_iZ_i^\top\mathcal{T}_{i-1}\right]$$

$$\mathcal{D}^{(2)}_{N_*,n} := \sum_{i=N_*+1}^{n}(\mathcal{I}-\eta\mathcal{L})^{n-i}D_i^{(2)}$$

$$\mathcal{R}_{N_*,n} := \sum_{i=N_*+1}^{n}(\mathcal{I}-\eta\mathcal{L})^{n-i}R_i$$

Then

$$\left\|\mathcal{T}_n - \eta\sum_{i=1}^{n}(\mathcal{I}-\eta\mathcal{L})^{n-i}(Z_iY_i^\top)\right\| \le \underbrace{\left\|(\mathcal{I}-\eta\mathcal{L})^{n-N_*}\mathcal{T}_{N_*}\right\|}_{=:\alpha_1} + \underbrace{\left\|\eta\sum_{i=1}^{N_*}(\mathcal{I}-\eta\mathcal{L})^{n-i}(Z_iY_i^\top)\right\|}_{\alpha_2}$$
$$+ \underbrace{\left\|\mathcal{D}^{(1,1)}_{N_*,n}\right\|}_{=:\alpha_3} + \underbrace{\left\|\mathcal{D}^{(2)}_{N_*,n}\right\|}_{=:\alpha_4} + \underbrace{\|\mathcal{R}_{N_*,n}\|}_{=:\alpha_5}$$

We consider each term separately.

(0) Preliminaries. We first record some preliminary results that will be used in the subsequent analysis.

(i) Recall definition of τ_* (C.8)

$$\tau_* = 3C_5C_\psi^2\kappa_*\max\left(\sqrt{\widetilde{\eta}r\log n}\bar{\nu}, \widetilde{\eta}r(\log n)\nu_2\right) \asymp C_5C_\psi^2\kappa_*\sqrt{\widetilde{\eta}r\log n}\;\bar{\nu}$$

Here the last $\asymp$ follows from Lemma 7(ii).

Since $\kappa_* \ge 1$ and $\delta_p \in (0, 1/e)$, the sample size condition (C.17) implies that

$$n^{-\varepsilon} \ge C_8^{1/2}\kappa_*\sqrt{\widetilde{\eta}r^3}(\log n)^{3/2}, \quad \text{for } C_8 \text{ large enough.}$$

Thus,

$$\begin{aligned}\tau_* \asymp C_5 C_\psi^2 \kappa_* \sqrt{\widetilde{\eta} r \log n} &\leq \frac{C_5 C_\psi^2}{C_8^{1/2}} \frac{1}{r n^\varepsilon \log n} \\ &\leq \frac{1}{n^\varepsilon \log n} \qquad \text{for } C_8 \text{ large enough.}\end{aligned} \tag{C.54}$$

(ii) Recall (C.6) and (C.7)

$$\begin{aligned}\varphi_\sigma(\tau_*) &= C_5 C_\psi^2 \kappa_* \sqrt{\widetilde{\eta} r \log n}(1+\tau_* \nu_\infty)(\tau_* + \bar{\nu}) \\ \varphi_B(\tau_*) &= C_5 C_\psi^2 \kappa_* \widetilde{\eta} r (\log n)^2 (1+\tau_* \nu_2)(\tau_* + \nu_2) \\ \varphi(\tau_*) &= (\varphi_\sigma \vee_{\mathrm{cw}} \varphi_B)(\tau_*)\end{aligned}$$

and we have shown in Lemma 4(iii) that $\varphi(\tau_*) < \tau_*$. Thus

$$\max(\varphi_\sigma(\tau_*), \varphi_B(\tau_*)) \leq \varphi(\tau_*) < \tau_*$$

(1) α_1**.** Under $\mathscr{G}$, we have

$$\alpha_1 \leq 2(1-\widetilde{\eta})^{n-N_*} \tau_*$$

(2) α_2**.** Using $n - i = (n - N_*) + (N_* - i)$, we have

$$\alpha_2 \leq (1-\widetilde{\eta})^{n-N_*} \left\| \eta \sum_{i=1}^{N_*} (\mathcal{I} - \eta \mathcal{L})^{N_* - i} (Z_i Y_i^\top) \right\|$$

Denote $\Xi_{2,i} = \eta(\mathcal{I} - \eta\mathcal{L})^{N_*-i}(Z_i Y_i^\top)$. To apply Freedman's inequality, similar to the proof of Lemma 3, we have

$$\begin{aligned}\max_{i \in [N_*]} \|\Xi_{2,i}\| &\leq \eta \|Z_i\| \|Y_i\| \leq \eta C_{\psi,n}^2 \sqrt{\lambda_{1\sim r}\lambda_{r+1\sim d}} \\ &\leq \eta C_{\psi,n}^2 r \lambda_1 \nu_2 =: B_2\end{aligned}$$

and

$$\begin{aligned}&\max\left\{ \left\| \sum_{i=1}^{N_*} \mathbb{E}_{i-1} \Xi_{2,i} \Xi_{2,i}^\top \right\|^{1/2}, \left\| \sum_{i=1}^{N_*} \mathbb{E}_{i-1} \Xi_{2,i}^\top \Xi_{2,i} \right\|^{1/2} \right\} \\ &\leq \eta \max\left\{ \left\| \mathbb{E} Z Y^\top Y Z^\top \right\|^{1/2}, \left\| \mathbb{E} Y Z^\top Z Y^\top \right\|^{1/2} \right\} \left[\sum_{i=1}^{N_*} (1-\widetilde{\eta})^{2(N_*-i)} \right]^{1/2} \\ &\leq C_\psi^2 \eta \max\left\{ \sqrt{\lambda_{1\sim r}\lambda_{r+1}}, \sqrt{\lambda_1 \lambda_{r+1\sim d}} \right\} \widetilde{\eta}^{-1/2} \\ &\leq C_\psi^2 \sqrt{\widetilde{\eta} r} \kappa_* \bar{\nu} =: \sigma_2\end{aligned}$$

Then Freedman's inequality implies that with probability exceeding $1 - n^{-20}$,

$$\begin{aligned}\left\| \eta \sum_{i=1}^{N_*} (\mathcal{I} - \eta\mathcal{L})^{N_*-i} (Z_i Y_i^\top) \right\| &\overset{(i)}{\leq} C_5 \left[\sigma_2 \sqrt{\log n} + B_2 \log n \right] \\ &= C_5 C_\psi^2 \sqrt{\widetilde{\eta} r}\, \kappa_* \bar{\nu} \sqrt{\log n} + C_5 C_\psi^2 \widetilde{\eta} r \kappa_* \nu_2 (\log n)^2 \\ &= \sqrt{\widetilde{\eta}\rho_\sigma \bar{\nu}} + \widetilde{\eta} \rho_B \nu_2 \asymp \tau_*\end{aligned}$$

Here the constant in (i) is the same as that in Lemma 5 since both invoke Freedman's inequality and the matrix has the same dimension $(d - r) \times r$. Therefore, we have

$$\alpha_2 \lesssim (1-\widetilde{\eta})^{n-N_*} \tau_*$$

(3) α_3**.** Denote

$$S_1 = \eta \sum_{i=N_*+1}^{n} (\mathcal{I} - \eta\mathcal{L})^{n-i} \left[-\mathcal{T}_{i-1}(Y_iY_i^\top - \Lambda_*)\right],$$

$$S_2 = \eta \sum_{i=N_*+1}^{n} (\mathcal{I} - \eta\mathcal{L})^{n-i} \left[(Z_iZ_i^\top - \Lambda_{*,\perp})\mathcal{T}_{i-1}\right],$$

$$S_3 = \eta \sum_{i=N_*+1}^{n} (\mathcal{I} - \eta\mathcal{L})^{n-i} \left[-\mathcal{T}_{i-1}Y_iZ_i^\top\mathcal{T}_{i-1}\right].$$

We consider S_1, S_2, S_3 separately.

Term S_1: Denote

$$S_1^* = \sum_{i=N_*+1}^{n} \Xi_{1,i}, \quad \text{where } \Xi_{1,i} := \eta(\mathcal{I} - \eta\mathcal{L})^{n-i} \left[\mathcal{T}_{i-1}(Y_iY_i^\top - \Lambda_*)\right] \mathbb{1}\left\{\|\mathcal{T}_{i-1}\| \le 2\tau_*\right\}$$

To apply Freedman's inequality on S_1^*, we can calculate

$$\begin{aligned}\max_i \|\Xi_{1,i}\| \le 2\eta\tau_* \left\|Y_iY_i^\top - \Lambda_*\right\| \le 2\eta\tau_*(C_{\psi,n}^2\lambda_{1\sim r} + \lambda_1) \le 4C_{\psi,n}^2\widetilde{\eta}\tau_* r\kappa_* \\ \lesssim C_\psi^2\widetilde{\eta} r\kappa_*\tau_* \log n =: B_{3,1}\end{aligned}$$

and

$$\begin{aligned}&\max\Big\{\Big\|\sum_{i=N_*+1}^{n} \mathbb{E}_{i-1}\Xi_{1,i}\Xi_{1,i}^\top\Big\|^{1/2}, \Big\|\sum_{i=N_*+1}^{n} \mathbb{E}_{i-1}\Xi_{1,i}^\top\Xi_{1,i}\Big\|^{1/2}\Big\} \\ &\overset{(a)}{\le} 2\eta\tau_*\Big[\sum_{i=N_*+1}^{n} (1-\widetilde{\eta})^{2(n-i)} \left\|\mathbb{E}_{i-1}(Y_iY_i^\top - \Lambda_*)^2\right\|\Big]^{1/2} \overset{(b)}{\le} 2\eta\tau_* \cdot \widetilde{\eta}^{-1/2} C_\psi^2 \sqrt{\lambda_1\lambda_{1\sim r}} \\ &\lesssim C_\psi^2\sqrt{\widetilde{\eta}} r\tau_*\kappa_* =: \sigma_{3,1}\end{aligned}$$

where (a) follows from

$$\left\|\mathbb{1}\left\{\|\mathcal{T}_{i-1}\| \le 2\tau_*\right\}\mathbb{E}_{i-1}\mathcal{T}_{i-1}(Y_iY_i^\top - \Lambda_*)(Y_iY_i^\top - \Lambda_*)^\top\mathcal{T}_{i-1}^\top\right\| \le 4\tau_*^2 \left\|\mathbb{E}_{i-1}(Y_iY_i^\top - \Lambda_*)^2\right\|$$
$$\left\|\mathbb{1}\left\{\|\mathcal{T}_{i-1}\| \le 2\tau_*\right\}\mathbb{E}_{i-1}(Y_iY_i^\top - \Lambda_*)^\top\mathcal{T}_{i-1}^\top\mathcal{T}_{i-1}(Y_iY_i^\top - \Lambda_*)\right\| \le 4\tau_*^2 \left\|\mathbb{E}_{i-1}(Y_iY_i^\top - \Lambda_*)^2\right\|$$

and (b) is due to Lemma 1(i). Freedman inequality (Theorem 12) then implies that with probability at least $1 - n^{-20}$,

$$\|S_1^*\| \le C_5\left[\sigma_{3,1}\sqrt{\log n} + B_{3,1}\log n\right].$$

On good event $\mathscr{G}$, we have $S_1\mathbb{1}_{\mathscr{G}} = S_1^*\mathbb{1}_{\mathscr{G}}$. Thus, with probability at least $1 - n^{-20}$,

$$\begin{aligned}\|S_1\| &\le C_5\left[\sigma_{3,1}\sqrt{\log n} + B_{3,1}\log n\right]. \\ &\le C_5C_\psi^2\kappa_*\sqrt{r}\left[\sqrt{\widetilde{\eta}\log n} + \sqrt{r}\widetilde{\eta}\log^2 n\right]\tau_* \\ &\lesssim C_5C_\psi^2\kappa_*\sqrt{r\widetilde{\eta}\log n}\,\tau_*\end{aligned}$$

Here the last line follows from $\sqrt{\widetilde{\eta}\log n} \gg \sqrt{r}\widetilde{\eta}\log^2 n$, which is a consequence of the sample size condition (C.17). To see this, note that $\kappa_* \ge 1$ and $\delta_p \in (0, 1/e)$, thus (C.17) implies

$$n^{-2\varepsilon} \ge C_8\widetilde{\eta}\kappa_*^2 r^3(\log n)^3$$

Applying $n^{-2\varepsilon} \ge C_8\widetilde{\eta}\kappa_*^2 r^3(\log n)^3$ once more gives

$$C_5C_\psi^2\kappa_*\sqrt{r\widetilde{\eta}\log n}\,\tau_* \le \frac{1}{n^\varepsilon \log n}\tau_*$$

Thus, we have

$$\|S_1\| \lesssim C_5 C_\psi^2 \kappa_* \sqrt{r\widetilde{\eta}\log n}\;\tau_* \lesssim \frac{\tau_*}{n^{\varepsilon}\log n}$$

Term S_2: Similar to S_1, by exploiting $\lambda_{r+1\sim d} = \lambda_{1\sim r}\nu_2^2$ and $\lambda_{r+1} = \lambda_1 \nu_\infty^2$, we can obtain that with probability at least $1 - n^{-20}$,

$$\begin{aligned}
\|S_2\| &\lesssim C_5 C_\psi^2 \kappa_* \sqrt{r}\left[\sqrt{\widetilde{\eta}\log n}\,\nu_2\nu_\infty + \sqrt{r}\widetilde{\eta}\nu_2^2\log^2 n\right]\tau_* \\
&\le C_5 C_\psi^2 \kappa_* \sqrt{r}\left[\sqrt{\widetilde{\eta}\log n}\,\bar{\nu} + \sqrt{r}\widetilde{\eta}\bar{\nu}^2\log^2 n\right]\tau_* \\
&\lesssim \left[\tau_* + \tau_*^2\right]\tau_* \lesssim \tau_*^2
\end{aligned}$$

where the last step follows since $\tau_* \lesssim n^{-\varepsilon}\log^{-1} n$ by (C.17).

Term S_3: define

$$S_3^* = \sum_{i=N_*+1}^{n} \Xi_{3,i}, \quad \text{where } \Xi_{3,i} = \eta(\mathcal{I} - \eta\mathcal{L})^{n-i}\left[\mathcal{T}_{i-1}Y_i Z_i^\top \mathcal{T}_{i-1}\right]\mathbb{1}\left\{\|\mathcal{T}_{i-1}\| \le 2\tau_*\right\}$$

Note that by Lemma 36, we can obtain that $\max_i \|\Xi_{3,i}\| \le \eta(2\tau_*)^2 \left\|Y_i Z_i^\top\right\|$ and

$$\begin{aligned}
\left\|\mathbb{E}_{i-1}\Xi_{3,i}\Xi_{3,i}^\top\right\| &\le (1-\widetilde{\eta})^{2(n-i)}\eta^2 \left\|\mathbb{E}_{i-1}\left[\mathcal{T}_{i-1}Y_i Z_i^\top \mathcal{T}_{i-1}\right]\left[\mathcal{T}_{i-1}Y_i Z_i^\top \mathcal{T}_{i-1}\right]^\top \mathbb{1}\left\{\|\mathcal{T}_{i-1}\| \le 2\tau_*\right\}\right\| \\
&\le (1-\widetilde{\eta})^{2(n-i)}\eta^2 (2\tau_*)^4 \left\|\mathbb{E}_{i-1}\left[Y_i Z_i^\top Z_i Y_i^\top\right]\right\| \\
\left\|\mathbb{E}_{i-1}\Xi_{3,i}^\top\Xi_{3,i}\right\| &\le (1-\widetilde{\eta})^{2(n-i)}\eta^2 \left\|\mathbb{E}_{i-1}\left[\mathcal{T}_{i-1}Y_i Z_i^\top \mathcal{T}_{i-1}\right]^\top\left[\mathcal{T}_{i-1}Y_i Z_i^\top \mathcal{T}_{i-1}\right]\mathbb{1}\left\{\|\mathcal{T}_{i-1}\| \le 2\tau_*\right\}\right\| \\
&\le (1-\widetilde{\eta})^{2(n-i)}\eta^2 (2\tau_*)^4 \left\|\mathbb{E}_{i-1}\left[Z_i Y_i^\top Y_i Z_i^\top\right]\right\|
\end{aligned}$$

Using results from (2) α_2, we have

$$\|S_3^*\| \lesssim \tau_*^3$$

Combining S_1, S_2, S_3: We have

$$S_1 + S_2 + S_3 \lesssim \tau_*\left[C_5 C_\psi^2 \kappa_* \sqrt{r\widetilde{\eta}\log n} + \tau_* + \tau_*^2\right]$$

up to absolute constant.

(4) α_4. Define

$$\mathcal{D}_{N_*,n}^{(2)*} := \sum_{i=N_*+1}^{n} \Xi_{4,i}, \quad \text{where } \Xi_{4,i} = (\mathcal{I} - \eta\mathcal{L})^{n-i} D_i^{(2)}\mathbb{1}\left\{\|\mathcal{T}_{i-1}\| \le 2\tau_*\right\}$$

To apply Freedman's inequality, by Lemma 3 and 36, we can calculate

$$\max_i \|\Xi_{4,i}\| \le \left\|D_i^{(2)}\mathbb{1}\left\{\|\mathcal{T}_{i-1}\| \le 2\tau_*\right\}\right\| \lesssim C_\psi^4 \widetilde{\eta}^2 r^2 \kappa_*^2 (1+\tau_*\nu_2)^2(\tau_* + \nu_2)\log^2 n =: B_4$$

and

$$\begin{aligned}
&\max\Big\{\Big\|\sum_{i=N_*+1}^{n}\mathbb{E}_{i-1}\Xi_{4,i}\Xi_{4,i}^\top\Big\|^{1/2}, \Big\|\sum_{i=N_*+1}^{n}\mathbb{E}_{i-1}\Xi_{4,i}^\top\Xi_{4,i}\Big\|^{1/2}\Big\} \\
&\le \frac{1}{\sqrt{\widetilde{\eta}}}\max\Big\{\left\|\mathbb{E}_{i-1}D_i^{(2)}D_i^{(2)\top}\mathbb{1}\left\{\|\mathcal{T}_{i-1}\| \le 2\tau_*\right\}\right\|^{1/2}, \left\|\mathbb{E}_{i-1}D_i^{(2)\top}D_i^{(2)}\mathbb{1}\left\{\|\mathcal{T}_{i-1}\| \le 2\tau_*\right\}\right\|^{1/2}\Big\} \\
&\le C_\psi^4 \widetilde{\eta}^{3/2} r^{3/2}\kappa_*^2 (1+\tau_*\nu_\infty)^2(\tau_* + \bar{\nu}) =: \sigma_4
\end{aligned}$$

Then Freedman's inequality implies that with probability at least $1 - n^{-20}$,

$$\left\|\mathcal{D}_{N_*,n}^{(2)*}\right\| \le C_5\left[\sigma_4\sqrt{\log n} + B_4\log n\right]$$

On good event $\mathscr{G}$, we have $\mathcal{D}^{(2)}_{N_*,n}\mathbb{1}_{\mathscr{G}} = \mathcal{D}^{(2)*}_{N_*,n}\mathbb{1}_{\mathscr{G}}$. Thus, with probability at least $1-n^{-20}$,

$$\begin{aligned}
\left\|\mathcal{D}^{(2)}_{N_*,n}\right\| &\le \varphi^{(2)}_{\sigma,*} + \varphi^{(2)}_{B,*} \\
\text{where} \qquad \varphi^{(2)}_{\sigma,*} &:= C_5 C_\psi^4 \widetilde{\eta}^{3/2}\kappa_*^2 r^{3/2}(1+\tau_*\nu_\infty)^2(\tau_*+\bar{\nu})\sqrt{\log n} \\
\varphi^{(2)}_{B,*} &:= C_5 C_\psi^4 \widetilde{\eta}^{2}\kappa_*^2 r^{2}(1+\tau_*\nu_2)^2(\tau_*+\nu_2)\log^3 n
\end{aligned}$$

Comparing $\varphi^{(2)}_{\sigma,*}$ and $\varphi^{(2)}_{B,*}$ with $\varphi_\sigma(\tau_*)$ and $\varphi_B(\tau_*)$ gives

$$\begin{aligned}
\frac{\varphi^{(2)}_{\sigma,*}}{\varphi_\sigma(\tau_*)} &= C_\psi^2\widetilde{\eta}\kappa_* r(1+\tau_*\nu_\infty) \le 2C_\psi^2\widetilde{\eta} r\kappa_* \\
&= 2\sqrt{\widetilde{\eta}}\cdot C_\psi^2\sqrt{\widetilde{\eta}}\,\kappa_* r
\end{aligned}$$

and

$$\begin{aligned}
\frac{\varphi^{(2)}_{B,*}}{\varphi_B(\tau_*)} &= C_\psi^2\widetilde{\eta}\kappa_* r(1+\tau_*\nu_2)\log n \\
&\le \sqrt{\widetilde{\eta}}\cdot C_\psi^2\sqrt{\widetilde{\eta}}\kappa_* r\left(1+\tau_*\sqrt{d\nu_\infty}\right)\log n
\end{aligned}$$

Combining the above two bounds gives

$$\begin{aligned}
\alpha_4 &\lesssim \tau_*\sqrt{\widetilde{\eta}}\cdot C_\psi^2\sqrt{\widetilde{\eta}}\kappa_* r\max(1,\tau_*\sqrt{d\nu_\infty})\log n \\
&\lesssim \tau_*\sqrt{\widetilde{\eta}}\cdot\frac{1}{n^\varepsilon\log n}
\end{aligned}$$

where the last step follows from $\tau_*\le 1$ and $C_\psi^2\sqrt{\widetilde{\eta}}\,\kappa_* r\max(1,\sqrt{d\nu_\infty})\lesssim n^{-\varepsilon}\log^{-3/2}n$, which is a consequence of the sample size condition (C.17).

(5) α_5**.** Similar to the proof of Claim 1, we have

$$\begin{aligned}
\alpha_5 \lesssim \tau_*\cdot C_\psi^2\sqrt{\widetilde{\eta}}\kappa_* r(1+\tau_*\nu_\infty) &\overset{(i)}{\lesssim} \tau_*\cdot C_\psi^2\sqrt{\widetilde{\eta}}\kappa_* r \\
&\overset{(ii)}{\lesssim} \tau_*\cdot\frac{1}{n^\varepsilon\log n}
\end{aligned}$$

where (i) is due to $\tau_*\le 1$ and $\nu_\infty\le 1$; (ii) is a consequence of the sample size condition (C.17).

(6) Collecting all terms, we have with probability at least $1-n^{-18}$,

$$\sum_{i=1}^5\alpha_i \lesssim \tau_*\left[e^{-\widetilde{\eta}(n-N_*)} + C_5C_\psi^2\kappa_*\sqrt{r\widetilde{\eta}\log n} + \tau_* + \sqrt{\widetilde{\eta}}\cdot\frac{1}{n^\varepsilon\log n}\right]$$

Note that $(1-\widetilde{\eta})^{n-N_*}\le e^{-\widetilde{\eta}(n-N_*)}$. Since $N_*\le\frac{c_\eta^*}{c_\eta}n$, we have $n-N_*\ge n(1-\frac{c_\eta^*}{c_\eta})$ and

$$\widetilde{\eta}(n-N_*)\ge c_\eta\frac{\log n}{n}\cdot n(1-c_\eta^*/c_\eta) = (c_\eta-c_\eta^*)\log n \ge \frac{1}{2}\log n.$$

Thus, $(1-\widetilde{\eta})^{n-N_*}\le e^{-\frac{1}{2}\log n} = n^{-1/2}\lesssim\widetilde{\eta}^{1/2}$. Since $C_5\ge 1$, $C_\psi\ge 1$, $\kappa_*\ge 1$, we finally have

$$\sum_{i=1}^5\alpha_i \lesssim \tau_*\left[C_5C_\psi^2\kappa_*\sqrt{r\widetilde{\eta}\log n}+\tau_*\right]\lesssim\tau_*\cdot\frac{1}{n^\epsilon\log n}.$$

D Analysis of Oja's algorithm: unbounded sub-Gaussian observations

This appendix extends the bounded-observation analysis of Appendix C to the original sub-Gaussian model in Assumption 1. The argument proceeds by jointly truncating the signal and tail coordinates and controlling the resulting covariance perturbation. The proof has three steps.

- First, Lemma 11 constructs the truncated observations and shows that they satisfy the required almost-sure signal and tail bounds, up to a small covariance perturbation H.
- Second, we show that this perturbation can be absorbed into the drift remainder, yielding the one-step bounds in Lemma 12. These bounds allow the three main arguments of Appendix C to be transferred to the truncated process: convergence to the noise floor (Lemma 13), exact-rank geometric contraction (Lemma 14), and linearization (Lemma 15).
- Finally, Lemma 16 proves a lower bound for the original linearized term directly under Assumption 1 and $Y \perp Z$.

Appendix E combines these results with the truncation coupling and Haar initialization to prove Theorems 1 and 2, together with Propositions 3–4.

Recall Y, Z, $\widetilde{Y}$ and $\widetilde{Z}$ from (6), we have

$$X = U_* Y + U_{*,\perp} Z = U_* \Lambda_*^{1/2} \widetilde{Y} + U_{*,\perp} \Lambda_{*,\perp}^{1/2} \widetilde{Z}. \tag{D.1}$$

Lemma 11 (Truncation)**.** *Under Assumption 1, let $W = \begin{pmatrix} \widetilde{Y} \\ \widetilde{Z} \end{pmatrix}$ where $\widetilde{Y} = U_*^\top \widetilde{X}$ and $\widetilde{Z} = U_{*,\perp}^\top \widetilde{X}$. Then*

$$\mathbb{E}W = 0, \qquad \mathbb{E}WW^\top = I_d, \qquad \|W\|_{\psi_2} \le C_\psi.$$

Moreover, there exist universal constants $\widetilde{C}_1, \widetilde{C}_2, \widetilde{C}_3 > 0$ such that, for every $\delta_{\mathrm{tr}} \in (0, e^{-1})$, one can construct a random vector

$$\bar{W} = \begin{pmatrix} \bar{\widetilde{Y}} \\ \bar{\widetilde{Z}} \end{pmatrix}$$

with the following properties.

(i) $\mathbb{E}\overline{W} = 0$ *and* $\mathbb{P}(\overline{W} \neq W) \le \delta_{\mathrm{tr}}$.

(ii) *Writing* $\mathbb{E}\overline{W}\,\overline{W}^\top = I_d + H$, *then* $\|H\| \le \widetilde{C}_1 C_\psi^2 \delta_{\mathrm{tr}} \log(e/\delta_{\mathrm{tr}})$.

(iii) $\|\overline{W}\|_{\psi_2} \le \widetilde{C}_2 C_\psi$.

(iv) *Defining* $\overline{Y} := \Lambda_*^{1/2} \overline{\widetilde{Y}}$ *and* $\overline{Z} := \Lambda_{*,\perp}^{1/2} \overline{\widetilde{Z}}$, *then* $\mathbb{E}\overline{Y} = 0$, $\mathbb{E}\overline{Z} = 0$ *and*

$$\|\overline{Y}\|_2 \le \widetilde{C}_3 C_\psi \sqrt{\lambda_{1\sim r} \log(e/\delta_{\mathrm{tr}})} \qquad \textit{almost surely},$$
$$\|\overline{Z}\|_2 \le \widetilde{C}_3 C_\psi \sqrt{\lambda_{r+1\sim d} \log(e/\delta_{\mathrm{tr}})} \qquad \textit{almost surely}.$$

Proof. See Appendix D.1. □

Fix a constant $\delta_H \in (0, e^{-1})$ such that $\widetilde{C}_1 C_\psi^2 \delta_H \log(e/\delta_H) \le \frac{1}{2}$ where $\widetilde{C}_1$ is the universal constant in Lemma 11(ii), and set

$$\delta_{\mathrm{tr}} := n^{-20} \wedge \delta_H. \tag{D.2}$$

Such a constant δ_H exists because $x \mapsto x \log(e/x)$ is increasing on $(0, 1)$, and $x \log(e/x) \to 0$ as $x \downarrow 0$. Apply Lemma 11 with δ_{tr}, and let H be the covariance perturbation in Lemma 11(ii). Then we have

$$\begin{aligned} \|H\| &\le \frac{1}{2}, \\ \|H\| &\lesssim C_\psi^2 n^{-20} \log(en^{20}) \lesssim \widetilde{\eta}, \\ \|H\| &\lesssim C_\psi^4 \widetilde{\eta} \kappa_* r^{3/2}. \end{aligned} \tag{D.3}$$

Here the second line uses $\widetilde{\eta} = c_\eta (\log n)/n$, $c_\eta \ge 1/2$, and $n \ge 4$, while the third uses $C_\psi \ge 1$, $\kappa_* \ge 1$, and $r \ge 1$.

Let $\overline{Y}$, $\overline{Z}$, $\overline{\widetilde{Y}}$, $\overline{\widetilde{Z}}$ be the truncated random vectors from Lemma 11 with δ_{tr} as given in (D.2). Motivated by (D.1), define the truncated observation

$$\overline{X} := U_* \overline{Y} + U_{*,\perp} \overline{Z} = U_* \Lambda_*^{1/2} \overline{\widetilde{Y}} + U_{*,\perp} \Lambda_{*,\perp}^{1/2} \overline{\widetilde{Z}} \tag{D.4}$$

and

$$\overline{U}_0 = U_0; \qquad \overline{U}_k = \mathsf{QR}(\overline{U}_{k-1} + \eta \overline{X}_k \overline{X}_k^\top \overline{U}_{k-1}), \quad k \in [n]. \tag{D.5}$$

Then Lemma 11 implies that

- simultaneously for all $k \in [n]$,
$$\|\overline{Y}_k\|_2 \le C_{\psi,n}\sqrt{\lambda_{1\sim r}}, \qquad \|\overline{Z}_k\|_2 \le C_{\psi,n}\sqrt{\lambda_{r+1\sim d}}, \tag{D.6}$$
where $C_{\psi,n} = C_{\mathsf{noise}} C_\psi \sqrt{\log n}$ for some universal constant $C_{\mathsf{noise}} > 0$.
- (D.2) implies $n\delta_{\mathrm{tr}} \le n^{-19}$. Hence
$$\overline{X}_k = X_k, \qquad \overline{U}_k = U_k, \qquad \forall k \in [n], \qquad \text{with probability at least } 1 - n^{-19}. \tag{D.7}$$

Therefore, it suffices to show $\overline{U}_k$ is close to U_*. To this end, define

$$\overline{\mathcal{C}}_k := U_*^\top \overline{U}_k, \qquad \overline{S}_k := U_{*,\perp}^\top \overline{U}_k, \qquad \overline{\mathcal{T}}_k := \overline{S}_k \overline{\mathcal{C}}_k^{-1},$$

we show that $\overline{\mathcal{T}}_k$ satisfies analogues of Lemma 2–10 only up to constant factors.

Tangent recursion for the truncated process. Applying Sherman–Morrison formula as in Lemma 2 to the update (D.5) gives

$$\overline{\mathcal{T}}_+ - \overline{\mathcal{T}} = \eta \frac{\overline{F}_+}{1 + \eta \overline{\zeta}_+} = \eta \overline{F}_+ - \eta^2 \frac{\overline{\zeta}_+ \overline{F}_+}{1 + \eta \overline{\zeta}_+}, \tag{D.8}$$

where

$$\overline{F}_+ := (-\overline{\mathcal{T}}\overline{Y}_+ + \overline{Z}_+)(\overline{Y}_+^\top + \overline{Z}_+^\top \overline{\mathcal{T}}) \in \mathbb{R}^{(d-r)\times r}, \qquad \overline{\zeta}_+ = (\overline{Y}_+^\top + \overline{Z}_+^\top \overline{\mathcal{T}})\overline{Y}_+ \in \mathbb{R}.$$

Similar to (C.3)–(C.5), we can obtain

$$\overline{\mathcal{T}}_+ - \overline{\mathcal{T}} = (\mathcal{I} - \eta \overline{\mathcal{L}})\overline{\mathcal{T}} + \overline{D}_+ + \overline{R}_+, \tag{D.9}$$

where

$$\begin{aligned}
&\overline{\mathcal{L}}\overline{\mathcal{T}} := -\mathbb{E}' \overline{F}_+, \\
&\overline{D}_+ := (\overline{\mathcal{T}}_+ - \overline{\mathcal{T}}) - \mathbb{E}'(\overline{\mathcal{T}}_+ - \overline{\mathcal{T}}) = \underbrace{\eta(\overline{F}_+ - \mathbb{E}'\overline{F}_+)}_{=:\overline{D}_+^{(1)}} + \underbrace{\eta^2 \left(-\frac{\overline{\zeta}_+ \overline{F}_+}{1 + \eta \overline{\zeta}_+} + \mathbb{E}' \frac{\overline{\zeta}_+ \overline{F}_+}{1 + \eta \overline{\zeta}_+} \right)}_{=:\overline{D}_+^{(2)}}, \\
&\overline{R}_+ := -\eta^2 \mathbb{E}' \left[\frac{\overline{\zeta}_+ \overline{F}_+}{1 + \eta \overline{\zeta}_+} \right].
\end{aligned}$$

Note that due to covariance perturbation H in Lemma 11 (ii), $\overline{\mathcal{L}}$ differs from $\mathcal{L}$. Denote $\overline{R}_+^H := \overline{R}_+ + \eta(\mathcal{L} - \overline{\mathcal{L}})\overline{\mathcal{T}}$, we have

$$\overline{\mathcal{T}}_+ = (\mathcal{I} - \eta \mathcal{L})\overline{\mathcal{T}} + \overline{D}_+ + \overline{R}_+^H. \tag{D.10}$$

Effect of the covariance perturbation. Unlike the original coordinates, the truncated coordinates need not satisfy $\mathbb{E}[\overline{Z}\overline{Y}^\top] = 0$. Consequently, the conditional drift operator $\overline{\mathcal{L}}$ differs from $\mathcal{L}$. We now quantify this difference and show that it can be included in the drift remainder. Let $H = \left(\begin{smallmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{smallmatrix}\right)$ and

$$\begin{aligned}
\overline{\Lambda}_* &:= \mathbb{E}\overline{Y}\overline{Y}^\top = \Lambda_* + A_H, & A_H &:= \Lambda_*^{1/2} H_{11} \Lambda_*^{1/2}, \\
\overline{\Lambda}_{*,\perp} &:= \mathbb{E}\overline{Z}\overline{Z}^\top = \Lambda_{*,\perp} + B_H, & B_H &:= \Lambda_{*,\perp}^{1/2} H_{22} \Lambda_{*,\perp}^{1/2}, \\
\overline{\Lambda}_{*,\times} &:= \mathbb{E}\overline{Z}\overline{Y}^\top = C_H, & C_H &:= \Lambda_{*,\perp}^{1/2} H_{21} \Lambda_*^{1/2}.
\end{aligned} \tag{D.11}$$

Then

$$\overline{\mathcal{L}\mathcal{T}} = \overline{\mathcal{T}}\Lambda_* + \overline{\mathcal{T}}\overline{\Lambda}_{*,\times}^\top\overline{\mathcal{T}} - \overline{\Lambda}_{*,\times} - \overline{\Lambda}_{*,\perp}\overline{\mathcal{T}} = \mathcal{L}\overline{\mathcal{T}} + \overline{\mathcal{T}}A_H + \overline{\mathcal{T}}C_H^\top\overline{\mathcal{T}} - C_H - B_H\overline{\mathcal{T}}.$$

Since every block of H has operator norm at most $\|H\|$, we have

$$\|A_H\| \le \lambda_1\|H\|, \quad \|B_H\| \le \lambda_{r+1}\|H\| = \lambda_1\nu_\infty^2\|H\|, \quad \|C_H\| \le \sqrt{\lambda_1\lambda_{r+1}}\|H\| = \lambda_1\nu_\infty\|H\|, \tag{D.12}$$

which implies

$$\begin{aligned}\left\|(\overline{\mathcal{L}} - \mathcal{L})\overline{\mathcal{T}}\right\| &\le \|H\|\,\lambda_1\left(\|\overline{\mathcal{T}}\| + \nu_\infty\|\overline{\mathcal{T}}\|^2 + \nu_\infty + \nu_\infty^2\|\overline{\mathcal{T}}\|\right) \\ &= \|H\|\,\lambda_1(1 + \|\overline{\mathcal{T}}\|\nu_\infty)(\|\overline{\mathcal{T}}\| + \nu_\infty).\end{aligned} \tag{D.13}$$

One-step bounds for the truncated process. We next verify that the centered noise $\overline{D}_+$, $\overline{D}_+^{(2)}$ and modified drift remainder $\overline{R}_+$ in (D.10) satisfy the same bounds as their counterparts D_+, $D_+^{(2)}$ and R_+ in Lemma 3.

Lemma 12 (One step bound)**.** *Assume Assumption 1 holds, and let the truncated process be constructed using* (D.2)*. Assume $\overline{\mathcal{T}}$ is well-defined with $\|\overline{\mathcal{T}}\| = \tau$. If $C_{\psi,n}^2\widetilde{\eta}\kappa_* r(1+\tau\nu_2) \le \frac{1}{2}$, then the following hold.*

(I) *Noise $\overline{D}_+$: $\|\overline{D}_+\| \le 4C_{\psi,n}^2\widetilde{\eta}\kappa_* r(1+\tau\nu_2)(\tau+\nu_2)$ almost surely and*

$$\max\Big\{\left\|\mathbb{E}'\overline{D}_+\overline{D}_+^\top\right\|^{1/2}, \left\|\mathbb{E}'\overline{D}_+^\top\overline{D}_+\right\|^{1/2}\Big\} \lesssim C_\psi^2\widetilde{\eta}\kappa_*\sqrt{r}\left(1+\tau\nu_\infty\right)(\tau+\bar{\nu}).$$

(II) *Higher-order noise $\overline{D}_+^{(2)}$: $\left\|\overline{D}_+^{(2)}\right\| \le 4C_{\psi,n}^4\widetilde{\eta}^2\kappa_*^2 r^2(1+\tau\nu_2)^2(\tau+\nu_2)$ almost surely and*

$$\max\Big\{\left\|\mathbb{E}'\overline{D}_+^{(2)}\overline{D}_+^{(2)\top}\right\|^{1/2}, \left\|\mathbb{E}'\overline{D}_+^{(2)\top}\overline{D}_+^{(2)}\right\|^{1/2}\Big\} \lesssim C_\psi^4\widetilde{\eta}^2\kappa_*^2 r^{3/2}\left(1+\tau\nu_\infty\right)^2(\tau+\bar{\nu}).$$

(III) *Drift remainder $\overline{R}_+^H$: $\left\|\overline{R}_+^H\right\| \lesssim C_\psi^4\widetilde{\eta}^2\kappa_*^2 r^{3/2}\left(1+\tau\nu_\infty\right)^2(\tau+\nu_\infty)$.*

Proof. See Appendix D.2. □

Lemma 12 provides precisely the inputs used in the convergence analysis of Appendix C: the recursion has the same contractive linear part $\mathcal{I} - \eta\mathcal{L}$, the centered increments satisfy the same almost-sure and conditional-variance bounds, and the modified drift remainder has the same order as the remainder in Lemma 3. Therefore, after enlarging the absolute constants, the stopped-process argument, matrix-Freedman bound, one-epoch estimate, epoch chaining, and iteration-budget argument in Lemmas 5–8 apply verbatim to $\{\overline{\mathcal{T}}_k\}_{k\le n}$. This yields the following counterpart of Lemma 8.

Lemma 13 (Convergence)**.** *Suppose Assumptions 1,4, 5 hold. Assume*

$$\nu_\infty \ge c_\nu n^{-C_\nu} \quad \text{for constants } C_\nu \ge 0, c_\nu > 0.$$

Let $\varepsilon \in [0, 1/2)$ and set

$$\eta = c_\eta\frac{\log n}{n\Delta_{\min}}, \quad \text{where} \quad c_\eta \ge c_\eta^* := \Big(1 - \varepsilon + \frac{3C_\nu}{2}\Big)\Big(1 + \frac{\log^{-3} n + \log 2}{\varepsilon\log n + \log\log n}\Big).$$

If

$$n^{1-2\varepsilon} \ge C_{13}c_\eta\kappa_*^2 r^3(\log n)^4\frac{\log(1/\delta_p)}{\delta_p^2}\max(1, \nu_\infty d)$$

for some constant $C_{13} > 0$ large enough (depending only on C_ψ, C_ν, c_ν and C_d). Then with probability at least $1 - n^{-19}$,

$$\left\|\overline{\mathcal{T}}_k\right\| \le 2\overline{\tau}_* \asymp \kappa_*\bar{\nu}\sqrt{\widetilde{\eta}r\log n} = \frac{\lambda_1}{\Delta_{\min}}\sqrt{c_\eta r}\,\bar{\nu}\frac{\log n}{\sqrt{n}}, \qquad \text{for all } \left\lfloor\frac{c_\eta^*}{c_\eta}n\right\rfloor \le k \le n,$$

where

$$\overline{\tau}_* := \overline{C}\max\big(C_\psi^2\kappa_*\bar{\nu}\sqrt{\widetilde{\eta}r\log n}, C_\psi^2\kappa_*\nu_2\widetilde{\eta}r(\log n)^2\big)$$

for some constant $\overline{C} > 0$ (depending only on C_d).

Exact rank-r contraction. We next transfer Lemma 9 to the truncated process. When $\operatorname{rank}(\Sigma) = r$, $\Lambda_{*,\perp} = 0$, so the truncated tail coordinates $\overline{Z}_i$ vanish identically. Hence the exact multiplicative structure and the associated column-space reduction from Appendix C.9 are preserved. Lemma 12 controls the additional covariance-perturbation remainder.

Lemma 14 (Exact-rank contraction)**.** *Suppose Assumption 1 holds,* $\operatorname{rank}(\Sigma) = r$ *and* $\overline{\mathcal{T}}_0 = \mathcal{T}_0$ *is well-defined. Fix* $\varepsilon \in [0, 1/2)$ *and set*

$$\eta = c_\eta \frac{\log n}{n\lambda_r}, \qquad c_\eta \geq \frac{1}{2}.$$

If

$$n^{1-2\varepsilon} \geq C_{\mathrm{rk,tr}} c_\eta \kappa_*^2 r^2 (\log n)^4$$

for some constant $C_{\mathrm{rk,tr}} > 0$ *sufficiently large depending only on* C_ψ*, then, conditionally on* $\mathcal{F}_0 = \sigma(U_0)$*, with probability at least* $1 - n^{-19}$*,*

$$\|\overline{\mathcal{T}}_k\| \leq 16e \exp\left\{-\left(1 - \frac{1}{n^\varepsilon \log n}\right) k\widetilde{\eta}\right\} \|\mathcal{T}_0\|, \qquad k = 0, 1, \ldots, n.$$

In particular,

$$\|\overline{\mathcal{T}}_n\| \leq 16e\, n^{-c_\eta(1-(n^\varepsilon \log n)^{-1})} \|\mathcal{T}_0\|.$$

If, in addition, Assumption 5 holds and

$$n^{1-2\varepsilon} \geq C_{\mathrm{rk,tr}} c_\eta \kappa_*^2 r^2 (\log n)^4 \frac{\log(1/\delta_p)}{\delta_p^2}, \tag{D.14}$$

then on the same event,

$$\|\mathcal{T}_k\| \lesssim \frac{\sqrt{d}\, n^{1/2-\varepsilon}}{\sqrt{c_\eta}\kappa_* \sqrt{r} (\log n)^2} \exp\left\{-\left(1 - \frac{1}{n^\varepsilon \log n}\right) k\widetilde{\eta}\right\}, \qquad k = 0, 1, \ldots, n. \tag{D.15}$$

In particular,

$$\|\mathcal{T}_n\| \lesssim \frac{\sqrt{d}}{\sqrt{c_\eta}\kappa_* \sqrt{r} (\log n)^2} n^{1/2-\varepsilon-c_\eta(1-(n^\varepsilon \log n)^{-1})}.$$

Proof. See Appendix D.3. ☐

Linearization. We next transfer the linearization argument of Lemma 10. The only new term arises because $\overline{\Lambda}_{*,\times} := \mathbb{E}[\overline{ZY}^\top]$ need not vanish. After centering $\overline{Z}_i \overline{Y}_i^\top$, its post-burn-in mean contribution ($i \in (N_*, n]$) is absorbed into the modified drift remainder, while its pre-burn-in contribution ($i \in [1, N_*]$) is exponentially damped by the contraction between N_* and n. All other terms are controlled exactly as in the proof of Lemma 10 using Lemma 12.

Lemma 15 (Linearization)**.** *Under the assumptions of Lemma 13 with* $c_\eta \geq c_\eta^* + 1/2$*, let* $N_* := \lfloor n c_\eta^* / c_\eta \rfloor$*. Then, with probability at least* $1 - n^{-18}$*,*

$$\Big\|\overline{\mathcal{T}}_n - \eta \sum_{i=1}^n (\mathcal{I} - \eta\mathcal{L})^{n-i} (\overline{Z}_i \overline{Y}_i^\top)\Big\| \lesssim \tau_* \big[C_\psi^2 \kappa_* \sqrt{r\widetilde{\eta}\log n} + \overline{\tau}_*\big] + e^{-\widetilde{\eta}(n-N_*)} \kappa_* \nu_\infty \|H\|. \tag{D.16}$$

In particular, (D.3) *gives* $\|H\| \lesssim \widetilde{\eta}$*, and therefore*

$$\Big\|\overline{\mathcal{T}}_n - \eta \sum_{i=1}^n (\mathcal{I} - \eta\mathcal{L})^{n-i} (\overline{Z}_i \overline{Y}_i^\top)\Big\| \lesssim \overline{\tau}_* \big[C_\psi^2 \kappa_* \sqrt{r\widetilde{\eta}\log n} + \overline{\tau}_*\big]. \tag{D.17}$$

Proof. See Appendix D.4. ☐

Lower bound for the original linearized term. The truncation argument above is needed only for the convergence and linearization upper bounds. For the lower bound, we return to the original observations and analyze the linearized term directly. This argument does not require bounded observations, Assumption 4, or Assumption 5.

Lemma 16 (Lower bound for the linearized term)**.** *Assume Assumption 1 holds with Y independent of Z. If $\widetilde{\eta}\kappa_*^2 \to 0$ and $n\widetilde{\eta} \geq \frac{1}{2}\log n$, then*

$$\mathbb{P}\left\{\left\|\eta\sum_{i=1}^{n}(\mathcal{I}-\eta\mathcal{L})^{n-i}(Z_iY_i^\top)\right\|^2 \geq \left(1-\frac{1}{n}\right)\frac{\widetilde{\eta}\lambda_1\lambda_{1\sim r}}{4\Delta_{\max}\Delta_{\min}}\bar{\nu}^2\right\} \geq \frac{1}{12}\left[1+o(1)\right].$$

Furthermore, if $\kappa_* \lesssim 1$, then

$$\widetilde{\eta}\frac{\lambda_1\lambda_{1\sim r}}{\Delta_{\max}\Delta_{\min}}\bar{\nu}^2 \asymp \frac{\tau_*^2}{\log n}.$$

Proof. See Appendix D.5. □

Remark 9 (Weaker nondegeneracy condition). The independence condition $Y \perp Z$ in Lemma 16 can be relaxed to $\lambda_{\min}(\widetilde{\mathfrak{C}}) \gtrsim 1$. Indeed, this condition implies

$$\mathbb{E}[\widetilde{Z}_\ell^2\widetilde{Y}_j^2] \gtrsim 1, \qquad \ell \in [d-r],\ j \in [r],$$

which yields the same row- and column-wise second-moment lower bounds used in the proof, up to constant factors. The fourth-moment upper bounds require only Assumption 1. Hence the conclusion of Lemma 16 continues to hold under this weaker nondegeneracy condition.

D.1 Proof of Lemma 11

$\mathbb{E}W = 0$, $\mathbb{E}WW^\top = I_d$, and $\|W\|_{\psi_2} \leq C_\psi$ follow directly from Assumption 1 and the properties of orthogonal transformations. The rest of the proof is devoted to constructing $\overline{W}$ and verifying its properties.

Preliminary moment estimates. We first show the following rare-event estimates. If a real random variable V satisfies $\|V\|_{\psi_2} \leq \sigma$ and $\mathcal{E}$ is an arbitrary event with $\mathbb{P}(\mathcal{E}) \leq \delta_{\mathrm{tr}} \leq e^{-1}$, then

$$\mathbb{E}\left[|V|\mathbb{1}_{\mathcal{E}}\right] \lesssim \sigma\delta_{\mathrm{tr}}\sqrt{\log(e/\delta_{\mathrm{tr}})} \tag{D.18}$$

$$\mathbb{E}\left[V^2\mathbb{1}_{\mathcal{E}}\right] \lesssim \sigma^2\delta_{\mathrm{tr}}\log(e/\delta_{\mathrm{tr}}). \tag{D.19}$$

Indeed, the sub-Gaussian tail estimate gives

$$\mathbb{P}\left(|V|>t\right) \leq 2\exp\left(-\frac{ct^2}{\sigma^2}\right).$$

Taking $t_{\mathrm{tr}} = \widetilde{C}\sigma\sqrt{\log(e/\delta_{\mathrm{tr}})}$ for $\widetilde{C}$ sufficiently large, we have

$$\begin{aligned}\mathbb{E}[|V|\mathbb{1}_{\mathcal{E}}] &\leq t_{\mathrm{tr}}\delta_{\mathrm{tr}} + \mathbb{E}\left[|V|\mathbb{1}_{\{|V|>t_{\mathrm{tr}}\}}\right]\\ &= t_{\mathrm{tr}}\delta_{\mathrm{tr}} + t_{\mathrm{tr}}\mathbb{P}\left(|V|>t_{\mathrm{tr}}\right) + \int_{t_{\mathrm{tr}}}^{\infty}\mathbb{P}\left(|V|>s\right)ds\\ &\lesssim \sigma\delta_{\mathrm{tr}}\sqrt{\log(e/\delta_{\mathrm{tr}})}.\end{aligned}$$

The proof of (D.19) is analogous:

$$\begin{aligned}\mathbb{E}[V^2\mathbb{1}_{\mathcal{E}}] &\leq t_{\mathrm{tr}}^2\delta_{\mathrm{tr}} + \mathbb{E}[V^2\mathbb{1}_{\{|V|>t_{\mathrm{tr}}\}}]\\ &= t_{\mathrm{tr}}^2\delta_{\mathrm{tr}} + t_{\mathrm{tr}}^2\mathbb{P}(|V|>t_{\mathrm{tr}}) + \int_{t_{\mathrm{tr}}}^{\infty}2s\mathbb{P}(|V|>s)ds\\ &\lesssim \sigma^2\delta_{\mathrm{tr}}\log(e/\delta_{\mathrm{tr}}).\end{aligned}$$

Construction of $\overline{W}$. Define the two deterministic linear maps

$$A_Y := \begin{pmatrix}\Lambda_*^{1/2} & 0\end{pmatrix} \in \mathbb{R}^{r\times d}, \qquad A_Z := \begin{pmatrix}0 & \Lambda_{*,\perp}^{1/2}\end{pmatrix} \in \mathbb{R}^{(d-r)\times d}.$$

Then we have $Y = A_Y W$, $Z = A_Z W$ and $\|A_Y\|_{\mathrm{F}}^2 = \lambda_{1\sim r}$, $\|A_Z\|_{\mathrm{F}}^2 = \lambda_{r+1\sim d}$. Let

$$q := \log(e/\delta_{\mathrm{tr}}).$$

Since $\delta_{\mathrm{tr}} \le e^{-1}$, we have $q \ge 2$. By Lemma 37, we have

$$\mathbb{E}\|Y\|_2^{2q} \le \left(CC_\psi\sqrt{q\lambda_{1\sim r}}\right)^{2q}.$$

Define

$$L_Y := C_1 C_\psi \sqrt{\lambda_{1\sim r}\log\left(e/\delta_{\mathrm{tr}}\right)}, \quad L_Z := C_1 C_\psi \sqrt{\lambda_{r+1\sim d}\log\left(e/\delta_{\mathrm{tr}}\right)}, \quad \text{where} \quad C_1 = Ce.$$

Then Markov's inequality gives

$$\mathbb{P}(\|Y\|_2 > L_Y) \le \frac{\mathbb{E}\|Y\|_2^{2q}}{L_Y^{2q}} \le \left(\frac{C}{C_1}\right)^{2q} \le e^{-2q} = \left(\frac{\delta_{\mathrm{tr}}}{e}\right)^2 \le \frac{\delta_{\mathrm{tr}}}{4},$$

where the last inequality holds because $\delta_{\mathrm{tr}} \le e^{-1} \le e^2/4$. Similarly, we have

$$\mathbb{P}(\|Z\|_2 > L_Z) \le \frac{\mathbb{E}\|Z\|_2^{2q}}{L_Z^{2q}} \le \left(\frac{C}{C_1}\right)^{2q} \le e^{-2q} \le \frac{\delta_{\mathrm{tr}}}{4}.$$

Hence the event $\mathcal{A} := \{\|Y\|_2 > L_Y\} \cup \{\|Z\|_2 > L_Z\}$ satisfies

$$\mathbb{P}(\mathcal{A}) \le \frac{\delta_{\mathrm{tr}}}{2}. \tag{D.20}$$

Next, we define an event $\mathcal{B} \supset \mathcal{A}$ so that $\mathbb{P}(\mathcal{B}) = \delta_{\mathrm{tr}}$. Specifically, we define

$$\mathcal{B} := \mathcal{A} \cup \left(\mathcal{A}^c \cap \{U \le \theta_{\mathcal{A}}\}\right),$$

where $U \sim \mathrm{Unif}(0,1)$ is independent of W and $\theta_{\mathcal{A}} := \frac{\delta_{\mathrm{tr}} - \mathbb{P}(\mathcal{A})}{1-\mathbb{P}(\mathcal{A})}$. Under this definition, $\mathcal{A} \subseteq \mathcal{B}$ and (D.20) implies that $\theta_{\mathcal{A}} \in [0,1]$. Since U is independent of W, we have

$$\mathbb{P}(\mathcal{A}^c \cap \{U \le \theta_{\mathcal{A}}\}) = \mathbb{P}(\mathcal{A}^c)\mathbb{P}(U \le \theta_{\mathcal{A}}) = (1-\mathbb{P}(\mathcal{A}))\theta_{\mathcal{A}} = \delta_{\mathrm{tr}} - \mathbb{P}(\mathcal{A}),$$

which further implies that

$$\mathbb{P}(\mathcal{B}) = \mathbb{P}(\mathcal{A}) + \mathbb{P}(\mathcal{A}^c \cap \{U \le \theta_{\mathcal{A}}\}) = \delta_{\mathrm{tr}}. \tag{D.21}$$

Then, we define the random vector $\overline{W}$ as

$$\overline{W} := W\mathbb{1}_{\mathcal{B}^c} + b\mathbb{1}_{\mathcal{B}}, \quad \text{where} \quad b := \frac{\mathbb{E}\left[W\mathbb{1}_{\mathcal{B}}\right]}{\delta_{\mathrm{tr}}}. \tag{D.22}$$

Next, we verify that $\overline{W}$ satisfies Properties (i)–(iv).

Property (i). By (D.21), $\mathbb{P}(\overline{W} \ne W) \le \mathbb{P}(\mathcal{B}) = \delta_{\mathrm{tr}}$. Moreover,

$$\mathbb{E}\overline{W} = \mathbb{E}\left[W\mathbb{1}_{\mathcal{B}^c}\right] + b\delta_{\mathrm{tr}} = \mathbb{E}W - \mathbb{E}\left[W\mathbb{1}_{\mathcal{B}}\right] + \mathbb{E}\left[W\mathbb{1}_{\mathcal{B}}\right] = 0.$$

Property (ii): Since $\mathbb{E}\overline{W} = 0$, the covariance of $\overline{W}$ equals its second-moment matrix. (D.22) implies that

$$\begin{aligned}\mathbb{E}\overline{W}\overline{W}^\top &= \mathbb{E}\left[WW^\top\mathbb{1}_{\mathcal{B}^c}\right] + \delta_{\mathrm{tr}}bb^\top \\ &= I_d - \mathbb{E}\left[WW^\top\mathbb{1}_{\mathcal{B}}\right] + \delta_{\mathrm{tr}}bb^\top.\end{aligned}$$

Hence

$$H = -\mathbb{E}\left[WW^\top\mathbb{1}_{\mathcal{B}}\right] + \delta_{\mathrm{tr}}bb^\top.$$

For any unit vector $u \in \mathbb{R}^d$, by the definition of b in (D.22), $\mathbb{P}(\mathcal{B}) = \delta_{\mathrm{tr}}$, (D.18) and (D.19), we have

$$|u^\top b| = \frac{1}{\delta_{\mathrm{tr}}} \left|\mathbb{E}[u^\top W \mathbb{1}_{\mathcal{B}}]\right| \le \frac{1}{\delta_{\mathrm{tr}}} \mathbb{E}[|u^\top W| \mathbb{1}_{\mathcal{B}}] \lesssim C_\psi \sqrt{\log(e/\delta_{\mathrm{tr}})}. \tag{D.23}$$

and

$$\mathbb{E}[(u^\top W)^2 \mathbb{1}_{\mathcal{B}}] \lesssim C_\psi^2 \delta_{\mathrm{tr}} \log(e/\delta_{\mathrm{tr}}). \tag{D.24}$$

The constants in (D.23) and (D.24) are independent of the unit vector u. Then,

$$|u^\top H u| \le \mathbb{E}[(u^\top W)^2 \mathbb{1}_{\mathcal{B}}] + \delta_{\mathrm{tr}} (u^\top b)^2 \lesssim C_\psi^2 \delta_{\mathrm{tr}} \log(e/\delta_{\mathrm{tr}}).$$

Since H is symmetric, taking supremum over all unit vectors $u \in \mathbb{R}^d$ gives

$$\|H\| \lesssim C_\psi^2 \delta_{\mathrm{tr}} \log(e/\delta_{\mathrm{tr}}).$$

Property (iii) For any unit vector $u \in \mathbb{R}^d$, we have

$$\|u^\top \overline{W}\|_{\psi_2} \le \|u^\top W \mathbb{1}_{\mathcal{B}^c}\|_{\psi_2} + \|u^\top b \mathbb{1}_{\mathcal{B}}\|_{\psi_2} \le \|u^\top W\|_{\psi_2} + \|u^\top b \mathbb{1}_{\mathcal{B}}\|_{\psi_2} \le C_\psi + \|u^\top b \mathbb{1}_{\mathcal{B}}\|_{\psi_2}.$$

Then it suffices to bound $\|u^\top b \mathbb{1}_{\mathcal{B}}\|_{\psi_2}$.

By (D.23), we can choose a universal constant $C_2 > 0$ sufficiently large so that

$$\frac{|u^\top b|^2}{C_2^2 C_\psi^2} \le \frac{1}{2} \log(e/\delta_{\mathrm{tr}}).$$

Then

$$\begin{aligned}
\mathbb{E} \exp\left(\frac{|u^\top b|^2 \mathbb{1}_{\mathcal{B}}}{C_2^2 C_\psi^2}\right) &= 1 - \delta_{\mathrm{tr}} + \delta_{\mathrm{tr}} \exp\left(\frac{|u^\top b|^2}{C_2^2 C_\psi^2}\right) \\
&\le 1 - \delta_{\mathrm{tr}} + \delta_{\mathrm{tr}} (e/\delta_{\mathrm{tr}})^{1/2} \\
&= 1 - \delta_{\mathrm{tr}} + \sqrt{e \delta_{\mathrm{tr}}} \le 2,
\end{aligned}$$

where the last inequality follows from $\delta_{\mathrm{tr}} \le e^{-1}$. Therefore, we have

$$\|u^\top b \mathbb{1}_{\mathcal{B}}\|_{\psi_2} \le C_2 C_\psi.$$

which then implies

$$\|u^\top \overline{W}\|_{\psi_2} \le (1 + C_2) C_\psi.$$

Taking the supremum over all unit vectors $u \in \mathbb{R}^d$ gives

$$\|\overline{W}\|_{\psi_2} \le (1 + C_2) C_\psi,$$

which proves Property (iii).

Property (iv): we consider two cases:

- On $\mathcal{B}^c$, one has $\mathcal{B}^c \subseteq \mathcal{A}^c$, and therefore
$$\left\|\overline{Y}\right\|_2 = \|Y\|_2 \le L_Y, \qquad \left\|\overline{Z}\right\|_2 = \|Z\|_2 \le L_Z.$$
- On $\mathcal{B}$, partition $b = \binom{b_Y}{b_Z}$ where
$$b_Y = \frac{\mathbb{E}[\widetilde{Y} \mathbb{1}_{\mathcal{B}}]}{\delta_{\mathrm{tr}}}, \qquad b_Z = \frac{\mathbb{E}[\widetilde{Z} \mathbb{1}_{\mathcal{B}}]}{\delta_{\mathrm{tr}}}$$
according to the signal and tail blocks. Then it suffices to consider $\|\Lambda_*^{1/2} b_Y\|_2$ and $\|\Lambda_{*,\perp}^{1/2} b_Z\|_2$. Note that for any unit vector $u \in \mathbb{R}^r$, we have
$$\begin{aligned}
\left|u^\top \Lambda_*^{1/2} b_Y\right| &= \delta_{\mathrm{tr}}^{-1} \left|\mathbb{E} u^\top \Lambda_*^{1/2} \widetilde{Y} \mathbb{1}_{\mathcal{B}}\right| \\
&\lesssim C_\psi \sqrt{\lambda_1 \log(e/\delta_{\mathrm{tr}})} \le C_\psi \sqrt{\lambda_{1\sim r} \log(e/\delta_{\mathrm{tr}})}.
\end{aligned}$$
Here the second line applies (D.18) with $\|u^\top \Lambda_*^{1/2} \widetilde{Y}\|_{\psi_2} \le \sqrt{\lambda_1} C_\psi$. Taking the supremum over unit u gives
$$\|\Lambda_*^{1/2} b_Y\|_2 \lesssim C_\psi \sqrt{\lambda_1 \log(e/\delta_{\mathrm{tr}})} \le C_\psi \sqrt{\lambda_{1\sim r} \log(e/\delta_{\mathrm{tr}})}. \tag{7}$$
Similarly, we can obtain
$$\|\Lambda_{*,\perp}^{1/2} b_Z\|_2 \lesssim C_\psi \sqrt{\lambda_{r+1} \log(e/\delta_{\mathrm{tr}})} \le C_\psi \sqrt{\lambda_{r+1\sim d} \log(e/\delta_{\mathrm{tr}})}. \tag{8}$$

Combining the two cases above, we conclude that

$$\left\|\overline{Y}\right\|_2 \lesssim C_\psi \sqrt{\lambda_{1\sim r}\log(e/\delta_{\rm tr})}, \quad \left\|\overline{Z}\right\|_2 \lesssim C_\psi \sqrt{\lambda_{r+1\sim d}\log(e/\delta_{\rm tr})} \quad \text{almost surely.}$$

The proof is then complete.

D.2 Proof of Lemma 12

Proof. By (D.2), Lemma 11 and $\log(e/\delta_{\rm tr}) \lesssim \log n$, choosing the constant $C_{\sf noise} > 0$ in $C_{\psi,n} := C_{\sf noise} C_\psi \sqrt{\log n}$ sufficiently large gives

$$\|\overline{Y}\|_2 \le C_{\psi,n}\sqrt{\lambda_{1\sim r}}, \qquad \|\overline{Z}\|_2 \le C_{\psi,n}\sqrt{\lambda_{r+1\sim d}}, \qquad \text{almost surely.}$$

For brevity, define

$$\overline{V}_1 := -\overline{\mathcal{T}}\overline{Y} + \overline{Z}, \qquad \overline{V}_2 := \overline{Y} + \overline{\mathcal{T}}^\top \overline{Z}.$$

Then

$$\overline{F}_+ = \overline{V}_1 \overline{V}_2^\top, \qquad \overline{\zeta}_+ = \overline{V}_2^\top \overline{Y}.$$

Since $\|\overline{\mathcal{T}}\| = \tau$, the preceding block bounds give

$$\|\overline{V}_1\|_2 \le C_{\psi,n}\sqrt{\lambda_{1\sim r}}(\tau + \nu_2), \qquad \|\overline{V}_2\|_2 \le C_{\psi,n}\sqrt{\lambda_{1\sim r}}(1 + \tau\nu_2).$$

Consequently,

$$|\eta\overline{\zeta}_+| \le \eta \|\overline{V}_2\|_2 \|\overline{Y}\|_2 \le C_{\psi,n}^2 \eta\lambda_{1\sim r}(1+\tau\nu_2) \le C_{\psi,n}^2 \widetilde{\eta}\kappa_* r(1+\tau\nu_2) \le \frac{1}{2}, \quad \text{almost surely,}$$

where we used $\lambda_{1\sim r} \le r\lambda_1$ and $\eta\lambda_1 = \widetilde{\eta}\kappa_*$. Hence

$$|1 + \eta\overline{\zeta}_+| \ge \frac{1}{2}.$$

Almost-sure bounds in (I) and (II). Recall that

$$\overline{\mathcal{T}}_+ - \overline{\mathcal{T}} = \eta \frac{\overline{F}_+}{1 + \eta\overline{\zeta}_+}, \qquad \overline{D}_+ = (\overline{\mathcal{T}}_+ - \overline{\mathcal{T}}) - \mathbb{E}'(\overline{\mathcal{T}}_+ - \overline{\mathcal{T}}).$$

It follows that

$$\|\overline{\mathcal{T}}_+ - \overline{\mathcal{T}}\| \le 2\eta \|\overline{V}_1\|_2 \|\overline{V}_2\|_2 \le 2C_{\psi,n}^2 \widetilde{\eta}\kappa_* r(1+\tau\nu_2)(\tau+\nu_2).$$

Centering increases the almost-sure bound by at most a factor of two, and therefore

$$\|\overline{D}_+\| \le 4C_{\psi,n}^2 \widetilde{\eta}\kappa_* r(1+\tau\nu_2)(\tau+\nu_2).$$

proving the almost-sure bound in (I). Similarly, writing

$$\overline{Q}_+ := -\eta^2 \frac{\overline{\zeta}_+ \overline{F}_+}{1 + \eta\overline{\zeta}_+}, \qquad \overline{D}_+^{(2)} = \overline{Q}_+ - \mathbb{E}'\overline{Q}_+,$$

we have

$$\|\overline{Q}_+\| \le 2\eta^2 |\overline{\zeta}_+| \|\overline{V}_1\|_2 \|\overline{V}_2\|_2 \le 2C_{\psi,n}^4 \widetilde{\eta}^2 \kappa_*^2 r^2 (1+\tau\nu_2)^2 (\tau+\nu_2).$$

Centering again costs at most a factor of two, proving the almost-sure bound in (II).

Quadratic variation bounds in (I) and (II). We first verify the moment bound (D.25), which is the key input for the quadratic-variation bounds in the proof of Lemma 3. By Lemma 11,

$$\mathbb{E}\overline{W}\overline{W}^\top = I_d + H, \qquad \|\overline{W}\|_{\psi_2} \lesssim C_\psi, \qquad \|H\| \lesssim C_\psi^2 \delta_{\rm tr}\log n.$$

By (D.3), $\|H\| \le 1/2$. Define

$$W^\circ := (I_d + H)^{-1/2}\overline{W}.$$

Then

$$\mathbb{E}W^\circ W^{\circ\top} = I_d, \qquad \|W^\circ\|_{\psi_2} \lesssim C_\psi.$$

Therefore, Lemma 37 implies that, for every deterministic matrix A and every $q \geq 1$,

$$\begin{aligned}\left[\mathbb{E}\|A\overline{W}\|_2^{2q}\right]^{1/(2q)} &= \left[\mathbb{E}\big\|A(I_d+H)^{1/2}W^\circ\big\|_2^{2q}\right]^{1/(2q)} \\ &\lesssim C_\psi\sqrt{q}\|A(I_d+H)^{1/2}\|_{\mathrm{F}} \lesssim C_\psi\sqrt{q}\|A\|_{\mathrm{F}}. \end{aligned} \tag{D.25}$$

Applying (D.25) to the linear maps defining $\overline{V}_1$ and $\overline{V}_2$ gives, for $q \in \{2,4\}$,

$$\begin{aligned}\sup_{\|u\|_2=1}\left[\mathbb{E}'|u^\top\overline{V}_1|^{2q}\right]^{1/(2q)} &\lesssim C_\psi\sqrt{\lambda_1}(\tau+\nu_\infty), \\ \sup_{\|v\|_2=1}\left[\mathbb{E}'|v^\top\overline{V}_2|^{2q}\right]^{1/(2q)} &\lesssim C_\psi\sqrt{\lambda_1}(1+\tau\nu_\infty), \\ \left[\mathbb{E}'\|\overline{V}_1\|_2^{2q}\right]^{1/(2q)} &\lesssim C_\psi\sqrt{\lambda_{1\sim r}}(\tau+\nu_2), \\ \left[\mathbb{E}'\|\overline{V}_2\|_2^{2q}\right]^{1/(2q)} &\lesssim C_\psi\sqrt{r\lambda_1}(1+\tau\nu_\infty).\end{aligned}$$

Moreover, since

$$\overline{\zeta}_+ = \|\overline{Y}\|_2^2 + \overline{Z}^\top\overline{\mathcal{T}Y},$$

the same moment bound and Cauchy–Schwarz imply

$$\left[\mathbb{E}'|\overline{\zeta}_+|^4\right]^{1/4} \lesssim C_\psi^2 r\lambda_1(1+\tau\nu_\infty).$$

These are precisely the moment estimates used in the proof of Lemma 3. Indeed,

$$\overline{F}_+\overline{F}_+^\top = \|\overline{V}_2\|_2^2\overline{V}_1\overline{V}_1^\top, \qquad \overline{F}_+^\top\overline{F}_+ = \|\overline{V}_1\|_2^2\overline{V}_2\overline{V}_2^\top.$$

Thus, Cauchy–Schwarz gives

$$\begin{aligned}\left\|\mathbb{E}'\overline{F}_+\overline{F}_+^\top\right\|^{1/2} &\lesssim C_\psi^2\sqrt{r}\,\lambda_1(1+\tau\nu_\infty)(\tau+\nu_\infty), \\ \left\|\mathbb{E}'\overline{F}_+^\top\overline{F}_+\right\|^{1/2} &\lesssim C_\psi^2\sqrt{r}\,\lambda_1(1+\tau\nu_\infty)(\tau+\nu_2).\end{aligned}$$

Similarly, Holder's inequality gives

$$\begin{aligned}\left\|\mathbb{E}'\left[\overline{\zeta}_+^2\,\overline{F}_+\overline{F}_+^\top\right]\right\|^{1/2} &\lesssim C_\psi^4 r^{3/2}\lambda_1^2(1+\tau\nu_\infty)^2(\tau+\nu_\infty), \\ \left\|\mathbb{E}'\left[\overline{\zeta}_+^2\,\overline{F}_+^\top\overline{F}_+\right]\right\|^{1/2} &\lesssim C_\psi^4 r^{3/2}\lambda_1^2(1+\tau\nu_\infty)^2(\tau+\nu_2).\end{aligned}$$

Since centering reduces the second moment and $|1+\eta\overline{\zeta}_+| \geq 1/2$,

$$\begin{aligned}\mathbb{E}'\overline{D}_+\overline{D}_+^\top &\preceq 4\eta^2\mathbb{E}'\overline{F}_+\overline{F}_+^\top, \\ \mathbb{E}'\overline{D}_+^\top\overline{D}_+ &\preceq 4\eta^2\mathbb{E}'\overline{F}_+^\top\overline{F}_+.\end{aligned}$$

The preceding two moment bounds therefore prove the quadratic-variation bound in (I). Likewise,

$$\begin{aligned}\mathbb{E}'\overline{D}_+^{(2)}(\overline{D}_+^{(2)})^\top &\preceq 4\eta^4\mathbb{E}'\left[\overline{\zeta}_+^2\,\overline{F}_+\overline{F}_+^\top\right], \\ \mathbb{E}'(\overline{D}_+^{(2)})^\top\overline{D}_+^{(2)} &\preceq 4\eta^4\mathbb{E}'\left[\overline{\zeta}_+^2\,\overline{F}_+^\top\overline{F}_+\right],\end{aligned}$$

which proves the quadratic-variation bound in (II).

Remainder bound in (III). Recall that

$$\overline{R}_+ = -\eta^2\mathbb{E}'\left[\frac{\overline{\zeta}_+\overline{F}_+}{1+\eta\overline{\zeta}_+}\right].$$

By Lemma 36, the denominator bound, and the preceding row-moment estimate,

$$\|\overline{R}_{+}\| \le 2\eta^2 \left\|\mathbb{E}'\left[\overline{\zeta}_{+}^{-2}\overline{F}_{+}\overline{F}_{+}^{\top}\right]\right\|^{1/2} \lesssim C_{\psi}^4 \widetilde{\eta}^2 \kappa_*^2 r^{3/2}(1+\tau\nu_\infty)^2(\tau+\nu_\infty).$$

Meanwhile, (D.3) and (D.13) give

$$\eta\|(\mathcal{L}-\overline{\mathcal{L}})\overline{\mathcal{T}}\| \le \widetilde{\eta}\kappa_*\|H\|(1+\tau\nu_\infty)(\tau+\nu_\infty) \lesssim C_{\psi}^4\widetilde{\eta}^2\kappa_*^2 r^{3/2}(1+\tau\nu_\infty)^2(\tau+\nu_\infty).$$

Recall from (D.10) that $\overline{R}_{+}^{H} = \overline{R}_{+} + \eta(\mathcal{L}-\overline{\mathcal{L}})\overline{\mathcal{T}}$, combining the last bound with the bound for $\overline{R}_{+}$ proves (III). ☐

D.3 Proof of Lemma 14

Proof. Since $\operatorname{rank}(\Sigma) = r$, we have $\Lambda_{*,\perp} = 0$, and hence $\overline{Z}_i = 0$ almost surely for every $i \in [n]$. Therefore, exactly as in (C.43),

$$\overline{\mathcal{T}}_{k+1} = \overline{\mathcal{T}}_k\big(I_r + \eta\overline{Y}_{k+1}\overline{Y}_{k+1}^{\top}\big)^{-1}.$$

Consequently, as in (C.44),

$$\operatorname{col}(\overline{\mathcal{T}}_{k+1}) = \operatorname{col}(\overline{\mathcal{T}}_k), \qquad \|\overline{\mathcal{T}}_{k+1}\| \le \|\overline{\mathcal{T}}_k\|, \qquad k \ge 0.$$

The sample-size condition implies the analogue of (C.46), so Lemma 12 applies at every step. Specializing Lemma 12 to $\nu_2 = \nu_\infty = \bar{\nu} = 0$ gives, up to absolute constants, the same one-step almost-sure, conditional-variance, and drift-remainder bounds as (C.48), with $\overline{R}_i^{H}$ in place of R_i.

It remains to note that the dimension-reduction argument in the proof of Lemma 9 is preserved by truncation. Indeed, writing

$$\overline{K}_i := \frac{\overline{Y}_i\overline{Y}_i^{\top}}{1+\eta\|\overline{Y}_i\|^2},$$

we have

$$\overline{\mathcal{T}}_i - \overline{\mathcal{T}}_{i-1} = -\eta\overline{\mathcal{T}}_{i-1}\overline{K}_i, \qquad \overline{D}_i = -\eta\overline{\mathcal{T}}_{i-1}\big(\overline{K}_i - \mathbb{E}_{i-1}\overline{K}_i\big).$$

Hence, for any block starting at time s,

$$\operatorname{col}(\overline{D}_i) \subseteq \operatorname{col}(\overline{\mathcal{T}}_s), \qquad i > s.$$

Moreover, since $\mathcal{L}M = M\Lambda_*$ in the exact rank-r case, right multiplication by $(I_r - \eta\Lambda_*)^{s+t-i}$ preserves this column-space inclusion. Thus the weighted martingale in the analogue of(C.47) can again be compressed to a $q_s \times r$ martingale, where $q_s = \operatorname{rank}(\overline{\mathcal{T}}_s) \le r$. Applying matrix Freedman's inequality for each fixed terminal time and then taking a union bound over the at most n terminal times therefore gives the analogue of (C.49), with the same bound (C.50) on the block error, up to absolute constants.

The block-chaining argument in Steps 2–3 of the proof of Lemma 9 then applies unchanged, after enlarging the absolute constants, and gives the first conclusion.

Finally, if Assumption 5 and (D.14) also hold, the same calculation as in the last paragraph of the proof of Lemma 9 gives

$$\|\mathcal{T}_0\| \lesssim \frac{\sqrt{d}\,n^{1/2-\epsilon}}{\sqrt{c_\eta}\kappa_*\sqrt{r}(\log n)^2}.$$

Combining this with the first conclusion yields (D.15), and the stated bound for $\|\overline{\mathcal{T}}_n\|$ follows from $n\widetilde{\eta} = c_\eta \log n$. ☐

D.4 Proof of Lemma 15

Proof. Starting from (D.10), the same decomposition as in Lemma 10 gives

$$\overline{\mathcal{T}}_n = (\mathcal{I}-\eta\mathcal{L})^{n-N_*}\overline{\mathcal{T}}_{N_*} + \overline{\mathcal{D}}_{N_*,n}^{(1,0)} + \overline{\mathcal{D}}_{N_*,n}^{(1,1)} + \overline{\mathcal{D}}_{N_*,n}^{(2)} + \overline{\mathcal{R}}_{N_*,n}^{H},$$

where

$$\overline{\mathcal{D}}_{N_*,n}^{(1,0)} := \eta \sum_{i=N_*+1}^{n} (\mathcal{I} - \eta\mathcal{L})^{n-i}(\overline{Z}_i\overline{Y}_i^\top - \overline{\Lambda}_{*,\times});$$

$$\overline{\mathcal{D}}_{N_*,n}^{(1,1)} := \overline{S}_1 + \overline{S}_2 + \overline{S}_3,$$

$$\text{where} \quad \overline{S}_1 := \eta \sum_{i=N_*+1}^{n} (\mathcal{I} - \eta\mathcal{L})^{n-i}[-\overline{\mathcal{T}}_{i-1}(\overline{Y}_i\overline{Y}_i^\top - \overline{\Lambda}_*)],$$

$$\overline{S}_2 := \eta \sum_{i=N_*+1}^{n} (\mathcal{I} - \eta\mathcal{L})^{n-i}[(\overline{Z}_i\overline{Z}_i^\top - \overline{\Lambda}_{*,\perp})\overline{\mathcal{T}}_{i-1}],$$

$$\overline{S}_3 := \eta \sum_{i=N_*+1}^{n} (\mathcal{I} - \eta\mathcal{L})^{n-i}[-\overline{\mathcal{T}}_{i-1}(\overline{Y}_i\overline{Z}_i^\top - \overline{\Lambda}_{*,\times}^\top)\overline{\mathcal{T}}_{i-1}];$$

$$\overline{\mathcal{D}}_{N_*,n}^{(2)} := \sum_{i=N_*+1}^{n} (\mathcal{I} - \eta\mathcal{L})^{n-i}\overline{D}_i^{(2)};$$

$$\overline{\mathcal{R}}_{N_*,n}^{H} := \sum_{i=N_*+1}^{n} (\mathcal{I} - \eta\mathcal{L})^{n-i}\overline{R}_i^{H}.$$

Then

$$\begin{aligned}
&\left\|\overline{\mathcal{T}}_n - \eta\sum_{i=1}^{n}(\mathcal{I} - \eta\mathcal{L})^{n-i}(\overline{Z}_i\overline{Y}_i^\top)\right\| \\
&\le \underbrace{\left\|(\mathcal{I} - \eta\mathcal{L})^{n-N_*}\overline{\mathcal{T}}_{N_*}\right\|}_{=:\overline{\alpha}_1} + \underbrace{\left\|\eta\sum_{i=1}^{N_*}(\mathcal{I} - \eta\mathcal{L})^{n-i}(\overline{Z}_i\overline{Y}_i^\top - \overline{\Lambda}_{*,\times})\right\|}_{=:\overline{\alpha}_2} \\
&+ \underbrace{\left\|\overline{\mathcal{D}}_{N_*,n}^{(1,1)}\right\|}_{=:\overline{\alpha}_3} + \underbrace{\left\|\overline{\mathcal{D}}_{N_*,n}^{(2)}\right\|}_{=:\overline{\alpha}_4} + \underbrace{\left\|\overline{\mathcal{R}}_{N_*,n}^{H} - \eta\sum_{i=N_*+1}^{n}(\mathcal{I} - \eta\mathcal{L})^{n-i}\overline{\Lambda}_{*,\times}\right\|}_{=:\overline{\alpha}_5} + \underbrace{\left\|\eta\sum_{i=1}^{N_*}(\mathcal{I} - \eta\mathcal{L})^{n-i}\overline{\Lambda}_{*,\times}\right\|}_{=:\overline{\alpha}_6}.
\end{aligned}$$

Note that by Lemma 36 and a geometric series bound, we have

$$\overline{\alpha}_6 \le \eta\sum_{i=1}^{N_*}(1-\widetilde{\eta})^{n-i}\|\Lambda_{*,\times}\| \le (1-\widetilde{\eta})^{n-N_*}\frac{\|\Lambda_{*,\times}\|}{\Delta_{\min}} \lesssim e^{-\widetilde{\eta}(n-N_*)}\kappa_*\nu_\infty\|H\|.$$

Moreover, (D.12) and (D.3) gives

$$\eta\|\Lambda_{*,\times}\| \le \widetilde{\eta}\kappa_*\nu_\infty\|H\| \lesssim C_\psi^4\widetilde{\eta}^2\kappa_*^2 r^{3/2}\nu_\infty \le C_\psi^4\widetilde{\eta}^2\kappa_*^2 r^{3/2}(1+\tau\nu_\infty)^2(\tau+\nu_\infty).$$

Consequently, $\overline{R}_i^H - \eta\overline{\Lambda}_{*,\times}$ satisfies the same bound (up to an absolute constant) as $\overline{R}_i^H$ in Lemma 12. Thus, the same argument as in the proof of Lemma 10, with Lemma 12 in place of Lemma 3, shows that with probability at least $1 - n^{-18}$,

$$\overline{\alpha}_1 + \overline{\alpha}_2 + \overline{\alpha}_3 + \overline{\alpha}_4 + \overline{\alpha}_5 \lesssim \overline{\tau}_*\left[C_\psi^2\kappa_*\sqrt{r\widetilde{\eta}\log n} + \overline{\tau}_*\right].$$

Combining the above bounds for $\overline{\alpha}_1$, $\overline{\alpha}_2$, $\overline{\alpha}_3$, $\overline{\alpha}_4$, $\overline{\alpha}_5$ and $\overline{\alpha}_6$ gives (D.16).

Finally, since $c_\eta \ge c_\eta^* + 1/2$ and $N_* = \lfloor (c_\eta^*/c_\eta)n \rfloor$, we have

$$e^{-\widetilde{\eta}(n-N_*)} \le e^{-(c_\eta - c_\eta^*)\log n} \le n^{-1/2}.$$

By (D.3), $\|H\| \lesssim \widetilde{\eta}$. Hence,

$$e^{-\widetilde{\eta}(n-N_*)}\kappa_*\nu_\infty\|H\| \lesssim n^{-1/2}\kappa_*\nu_\infty\widetilde{\eta} \lesssim \overline{\tau}_*\, C_\psi^2\kappa_*\sqrt{r\widetilde{\eta}\log n}.$$

Thus, the last term in (D.16) is absorbed into the first term on its right-hand side, which proves (D.17). □

D.5 Proof of Lemma 16

We only prove the column bound, as the row bound follows by symmetry. Recall

$$\mathcal{D}_n^{(1,0)} = \eta \sum_{i=1}^{n} (\mathcal{I} - \eta \mathcal{L})^{n-i} (Z_i Y_i^\top)$$

Denote the first column of $\mathcal{D}_n^{(1,0)}$ by

$$W := \left[\mathcal{D}_n^{(1,0)}\right]_{\cdot,1} = \sum_{i=1}^{n} \gamma_i, \qquad \text{where } \gamma_i := \eta \operatorname{diag}\left\{(1-\eta\Delta_{l1})^{n-i}\right\}_{l=1}^{n-r} Z_i \left[Y_i\right]_1 \in \mathbb{R}^{d-r}.$$

Note that we have $\mathbb{E}\gamma_i = 0$ since $Y = U_*^\top X$ and $Z = U_{*,\perp}^\top X$.

Step 1: Column second moment By independence $\gamma_i \perp \gamma_{i'}$ for $i \neq i'$ and $\mathbb{E}\gamma_i = 0$, we have

$$\mathbb{E}\left\|W\right\|^2 = \mathbb{E}\sum_{i=1}^{n} \left\|\gamma_i\right\|^2 = \sum_{l=1}^{d-r} \underbrace{\mathbb{E}\sum_{i=1}^{n} \left[\gamma_i\right]_l^2}_{=:\sigma_l^2}$$

Under $\eta\Delta_{\max} \in (0,1)$, we have $\eta\Delta_{l1} \in (0,1)$, hence $2 - \eta\Delta_{l1} \in (1,2)$ and

$$\begin{aligned}
\sigma_l^2 &= \sum_{i=1}^{n} \eta^2 (1-\eta\Delta_{l1})^{2(n-i)} \mathbb{E}\left[Z_i\right]_l^2 \left[Y_i\right]_1^2 = \sum_{i=1}^{n} \eta^2 (1-\eta\Delta_{l1})^{2(n-i)} \lambda_{r+l}\lambda_1 \\
&\geq \sum_{i=1}^{n} \eta^2 (1-\eta\Delta_{\max})^{2(n-i)} \lambda_{r+l}\lambda_1 = \eta\lambda_{r+l}\lambda_1 \frac{\left[1-(1-\eta\Delta_{\max})^{2n}\right]}{(2-\eta\Delta_{\max})\Delta_{\max}} \\
&\geq \eta\lambda_{r+l}\lambda_1 \frac{1-(1-\eta\Delta_{\max})^{2n}}{2\Delta_{\max}}
\end{aligned}$$

where the last line follows from $\eta\Delta_{\max} \in (0,1)$.
Moreover, since $n\widetilde{\eta} \geq \frac{1}{2}\log n$. We have

$$(1-\eta\Delta_{\max})^{2n} \leq (1-\eta\Delta_{\min})^{2n} \leq e^{-2n\eta\Delta_{\min}} = e^{-2n\widetilde{\eta}} \leq e^{-\log n} = \frac{1}{n}.$$

Thus, we have

$$\mathbb{E}\left\|W\right\|^2 = \mathbb{E}\sum_{i=1}^{n} \left\|\gamma_i\right\|^2 \geq \widetilde{\eta} \frac{\lambda_1 \lambda_{r+1\sim d}}{2\Delta_{\max}\Delta_{\min}} \left[1 - \tfrac{1}{n}\right] \tag{D.26}$$

Step 2: Column fourth moment

- First, we show that

$$\mathbb{E}\left\|W\right\|^4 \leq \sum_i \mathbb{E}\left\|\gamma_i\right\|^4 + 3\Big(\sum_i \mathbb{E}\left\|\gamma_i\right\|^2\Big)^2$$

  To see this, note that

$$\mathbb{E}\left\|W\right\|^4 = \mathbb{E}\Big\|\sum_{i=1}^{n}\gamma_i\Big\|^4 = \sum_{i,j,a,b=1}^{n} \mathbb{E}\big\langle\gamma_i,\, \gamma_j\big\rangle \left\langle\gamma_a,\, \gamma_b\right\rangle.$$

  By independence and $\mathbb{E}\gamma_i = 0$, any term with a singleton index (an index appearing exactly once among $\{i,j,a,b\}$) vanishes: conditioning on all other variables isolates a factor $\mathbb{E}[\gamma_i] = 0$. The surviving pairings are:

  (i) $(i=j,\ a=b)$: contributes $\sum_{i,a} \mathbb{E}\left[\left\|\gamma_i\right\|^2 \left\|\gamma_a\right\|^2\right]$.

(ii) $(i=a,\, j=b,\, i\neq j)$: contributes $\sum_{i\neq j}\mathbb{E}[\langle\gamma_i,\,\gamma_j\rangle^2]$.

(iii) $(i=b,\, j=a,\, i\neq j)$: same value as (ii).

Hence

$$\mathbb{E}\,\|W\|^4=\sum_{i,a}\mathbb{E}\big[\|\gamma_i\|^2\,\|\gamma_a\|^2\big]+2\sum_{i\neq j}\mathbb{E}[\langle\gamma_i,\,\gamma_j\rangle^2].$$

First sum. Splitting the diagonal and off-diagonal:

$$\begin{aligned}\sum_{i,a}\mathbb{E}\big[\|\gamma_i\|^2\,\|\gamma_a\|^2\big]&=\sum_i\mathbb{E}\big\|\gamma_i\big\|^4+\sum_{i\neq a}\mathbb{E}\,\|\gamma_i\|^2\cdot\mathbb{E}\,\|\gamma_a\|^2\\&\le\sum_i\mathbb{E}\,\|\gamma_i\|^4+\Big(\sum_i\mathbb{E}\,\|\gamma_i\|^2\Big)^2.\end{aligned}$$

Second sum. For $i\neq j$, by independence,

$$\mathbb{E}\big[\langle\gamma_i,\gamma_j\rangle^2\big]=\operatorname{tr}(\Sigma_i\Sigma_j),\quad\text{ where }\Sigma_m=\mathbb{E}[\gamma_m\gamma_m^\top].$$

Since $\operatorname{tr}(AB)\le\operatorname{tr}(A)\operatorname{tr}(B)$ for PSD matrices A,B, we have

$$2\sum_{i\neq j}\operatorname{tr}(\Sigma_i\Sigma_j)\le 2\sum_{i\neq j}\operatorname{tr}(\Sigma_i)\operatorname{tr}(\Sigma_j)\le 2\Big(\sum_i\operatorname{tr}\Sigma_i\Big)^2=2\Big(\sum_i\mathbb{E}\|\gamma_i\|^2\Big)^2.$$

Adding gives

$$\mathbb{E}\,\|W\|^4\le\sum_{i=1}^n\mathbb{E}\,\|\gamma_i\|^4+3\Big(\sum_{i=1}^n\mathbb{E}\,\|\gamma_i\|^2\Big)^2$$

- Next, we compute $\sum_{i=1}^n\mathbb{E}\,\|\gamma_i\|^4$. To this end, note that

$$\|\gamma_i\|^2=[Y_i]_1^2\sum_{l=1}^{d-r}\eta^2(1-\eta\Delta_{l1})^{2(n-i)}\,[Z_i]_l^2\,.$$

By independence $Y\perp Z$ and $\mathbb{E}\,[Y_i]_1^4\lesssim\lambda_1^2$ (sub-Gaussian), we have

$$\begin{aligned}\mathbb{E}\,\|\gamma_i\|^4&=\mathbb{E}\Big[[Y_i]_1^2\sum_{l=1}^{d-r}\eta^2(1-\eta\Delta_{lj})^{2(n-i)}\,[Z_i]_l^2\Big]^2\\&\lesssim\lambda_1^2\mathbb{E}\Big[\sum_{l=1}^{d-r}\eta^2(1-\eta\Delta_{l1})^{2(n-i)}\,[Z_i]_l^2\Big]^2\\&\overset{(i)}{=}\lambda_1^2\mathbb{E}\Big[\sum_{l=1}^{d-r}\eta^2(1-\eta\Delta_{l1})^{2(n-i)}\lambda_{r+l}\big[\widetilde{Z}_i\big]_l^2\Big]^2\\&\le\lambda_1^2\eta^4(1-\eta\Delta_{\min})^{4(n-i)}\mathbb{E}\Big[\sum_{l=1}^{d-r}\lambda_{r+l}\big[\widetilde{Z}_i\big]_l^2\Big]^2\\&\overset{(ii)}{=}\lambda_1^2\eta^4(1-\eta\Delta_{\min})^{4(n-i)}\mathbb{E}\big\|B\widetilde{Z}_i\big\|^4\end{aligned}$$

where (i) follows from $[Z_i]_j=\lambda_j^{1/2}\big[\widetilde{Z}_i\big]_j$ and (ii) holds with $B=\operatorname{diag}\Big(\lambda_{l+r}^{1/2}\Big)_{l=1}^{d-r}\in\mathbb{R}^{(d-r)\times(d-r)}$. Applying Lemma 37 gives

$$\big[\mathbb{E}\big\|B\widetilde{Z}_i\big\|^4\big]^{1/4}\lesssim\|B\|_{\mathrm{F}}\,C_\psi=C_\psi\lambda_{r+1\sim d}^{1/2}$$

Thus,

$$\mathbb{E}\,\|\gamma_i\|^4\lesssim C_\psi^4\eta^4(1-\eta\Delta_{\min})^{4(n-i)}\lambda_1^2\lambda_{r+1\sim d}^2\le C_\psi^4\widetilde{\eta}^4(1-\widetilde{\eta})^{4(n-i)}\frac{\lambda_1^2\lambda_{r+1\sim d}^2}{\Delta_{\min}^4}$$

which implies

$$\sum_{i=1}^n\mathbb{E}\,\|\gamma_i\|^4\lesssim C_\psi^4\widetilde{\eta}^4\frac{\lambda_1^2\lambda_{1\sim r}^2}{\Delta_{\min}^4}\nu_2^4\sum_{i=1}^n(1-\widetilde{\eta})^{4(n-i)}\le C_\psi^4\widetilde{\eta}^3\frac{\lambda_1^2\lambda_{r+1\sim d}^2}{\Delta_{\min}^4}$$

- With the above two bounds, we have

$$\frac{\sum_{i=1}^n \mathbb{E}\|\gamma_i\|^4}{\left(\sum_{i=1}^n \mathbb{E}\|\gamma_i\|^2\right)^2} \lesssim \frac{C_\psi^4 \widetilde{\eta}^3 \frac{\lambda_1^2 \lambda_{r+1\sim d}^2}{\Delta_{\min}^4}}{\widetilde{\eta}^2 \frac{\lambda_1^2 \lambda_{r+1\sim d}^2}{4\Delta_{\max}^2 \Delta_{\min}^2} \left[1-\frac{1}{n}\right]^2} \lesssim 4C_\psi^4 \widetilde{\eta} \frac{\Delta_{\max}^2}{\Delta_{\min}^2} = o(1)$$

which implies that

$$\begin{aligned}\mathbb{E}\|W\|^4 &\le \sum_{i=1}^n \mathbb{E}\|\gamma_i\|^4 + 3\Big(\sum_{i=1}^n \mathbb{E}\|\gamma_i\|^2\Big)^2 \\ &= (3+o(1))\Big(\sum_{i=1}^n \mathbb{E}\|\gamma_i\|^2\Big)^2 = (3+o(1))(\mathbb{E}\|W\|^2)^2\end{aligned}$$

Finally, applying the Paley-Zygmund inequality gives

$$\mathbb{P}\left\{\|W\|^2 \ge \frac{1}{2}\mathbb{E}\|W\|^2\right\} \ge \frac{1}{4}\frac{\left(\mathbb{E}\|W\|^2\right)^2}{\mathbb{E}\|W\|^4} \ge \frac{1}{12}(1+o(1))$$

which combined with (D.26) gives

$$\mathbb{P}\Big\{\|W\|^2 \ge \left[1-\tfrac{1}{n}\right]\widetilde{\eta}\frac{\lambda_1 \lambda_{r+1\sim d}}{4\Delta_{\max}\Delta_{\min}}\Big\} \ge \frac{1}{4}\frac{\left(\mathbb{E}\|W\|^2\right)^2}{\mathbb{E}\|W\|^4} \ge \frac{1}{12}(1+o(1))$$

E Proofs for Section 3.1

E.1 Proof of Theorem 1

We apply Lemma 13 to the truncated process and then transfer its conclusion to the original process through the truncation coupling (D.7). We first record two consequences of the sample-size condition (11):

$$n^{1-2\varepsilon} \ge C_1 c_\eta \kappa_*^2 r^3 (1+\nu_\infty d)(\log n)^4 \frac{\log(1/\delta_p)}{\delta_p^2}. \tag{11}$$

- $n^{1-2\varepsilon} \ge C_1 c_\eta \kappa_*^2 r^3 (\log n)^4 \frac{\log(1/\delta_p)}{\delta_p^2}$ implies that

  $$\delta_p \ge \delta_p^2 \ge \frac{\log(1/\delta_p)}{n^{1-2\varepsilon}} \ge \frac{1}{n}, \tag{E.1}$$

  where the first and third inequalities follow from $\delta_p \le 1/e$, while the second follows from $c_\eta \ge 1/2$, $\kappa_*, r \ge 1$, and C_1 being sufficiently large.
- $n^{1-2\varepsilon} \ge C_1 c_\eta \kappa_*^2 d \nu_\infty r^3 (\log n)^4 \frac{\log(1/\delta_p)}{\delta_p^2}$ similarly implies that

  $$d \le n^{1-2\varepsilon}\nu_\infty^{-1}\frac{\delta_p^2}{\log(1/\delta_p)} \le n^{1-2\varepsilon+C_\nu} c_\nu^{-1}$$

  where the second inequality follows from $\delta_p \le 1/e$ and (9). Hence Assumption 4 holds.

Moreover, by Lemma 41,

$$\mathbb{P}\Big\{\|\mathcal{T}_0\| \le C_{\mathsf{init}}\delta_p^{-1}\sqrt{rd\log(1/\delta_p)}\Big\} \ge 1-\frac{\delta_p}{2}.$$

Denote this event by $\mathcal{E}_0$. On $\mathcal{E}_0$, Assumption 5 holds, so Lemma 13 applies to the truncated process. Hence, with probability at least $1-n^{-19}$,

$$\|\overline{\mathcal{T}}_k\| \lesssim \tau_* \qquad \text{for all } \left\lfloor \tfrac{c_\eta^*}{c_\eta} n \right\rfloor \le k \le n.$$

By the truncation coupling (D.7),

$$\mathbb{P}\{\overline{U}_k = U_k \text{ for all } k \in [n]\} \ge 1-n^{-19}.$$

Therefore, with probability at least

$$\mathbb{P}\left\{\|\sin\Theta(U_k, U_*)\| \lesssim \tau_*,\ \text{for all } \left\lfloor \tfrac{c_\eta^*}{c_\eta} n \right\rfloor \le k \le n\right\} \ge (1-\delta_p/2)\left(1-2n^{-19}\right) \ge 1-\delta_p$$

where the last inequality follows from (E.1).

E.2 Proof of Theorem 2

We apply Lemma 14 to the truncated process and then transfer its conclusion to the original process through the truncation coupling. Let $\{\overline{X}_i\}_{i=1}^n$, $\{\overline{U}_k\}_{k=0}^n$, and $\{\overline{\mathcal{T}}_k\}_{k=0}^n$ denote the truncated observations, iterates, and tangent matrices constructed in Appendix D.

As in the proof of Theorem 1, the sample-size condition (13) implies, for C_2 sufficiently large,

$$\delta_p \geq \delta_p^2 \geq \frac{\log(1/\delta_p)}{n^{1-2\varepsilon}} \geq \frac{1}{n}. \tag{E.2}$$

By Lemma 41, Haar initialization gives

$$\mathbb{P}\left\{\|\mathcal{T}_0\| \leq C_{\rm init}\delta_p^{-1}\sqrt{rd\log(1/\delta_p)}\right\} \geq 1 - \frac{\delta_p}{2},$$

after enlarging the absolute constant if necessary. Denote this initialization event by $\mathcal{E}_0$.

On $\mathcal{E}_0$, Assumption 5 holds. Since $\operatorname{rank}(\Sigma) = r$, Lemma 14 therefore gives, conditionally on $\mathcal{E}_0$, with probability at least $1 - n^{-19}$,

$$\|\overline{\mathcal{T}}_k\| \lesssim \frac{\sqrt{d}\, n^{1/2-\epsilon}}{\sqrt{c_\eta}\,\kappa_*\sqrt{r}\,(\log n)^2}\exp\left\{-\left(1 - \frac{1}{n^\epsilon \log n}\right)k\widetilde{\eta}\right\}, \qquad k = 0, 1, \ldots, n.$$

Here the sample-size condition in the theorem is precisely (D.14), up to enlarging $C_{\rm rk,tr}$.

By the truncation coupling (D.7) in Appendix D,

$$\mathbb{P}\left\{\overline{U}_k = U_k \text{ for all } k \in [n]\right\} \geq 1 - n^{-19}.$$

Therefore, with probability at least

$$\left(1 - \frac{\delta_p}{2}\right)\left(1 - 2n^{-19}\right) \geq 1 - \delta_p,$$

where the last inequality follows from (E.2), the preceding bound holds with $\mathcal{T}_k$ in place of $\overline{\mathcal{T}}_k$ simultaneously for all $k = 0, \ldots, n$.

Finally, Lemma 39 gives

$$\|\sin\Theta(U_k, U_*)\| = \left\|\mathcal{T}_k(\mathcal{C}_k\mathcal{C}_k^\top)^{1/2}\right\| \leq \|\mathcal{T}_k\|,$$

since $\|(\mathcal{C}_k\mathcal{C}_k^\top)^{1/2}\| \leq 1$. This proves the uniform bound.

For $k = n$, using $n\widetilde{\eta} = c_\eta \log n$ gives

$$\exp\left\{-\left(1 - \frac{1}{n^\epsilon \log n}\right)n\widetilde{\eta}\right\} = n^{-c_\eta\left(1-(n^\epsilon \log n)^{-1}\right)},$$

which yields the stated bound for U_n and completes the proof.

E.3 Proof of Proposition 3

We apply Lemma 15 to the truncated process and then transfer its conclusion to the original process through the truncation coupling.

Let $\mathcal{E}_0$ be the initialization event defined in the proof of Theorem 1. On $\mathcal{E}_0$, Assumption 5 holds, so Lemma 15 gives, with probability at least $1 - n^{-18}$,

$$\left\|\overline{\mathcal{T}}_n - \eta\sum_{i=1}^n(\mathcal{I} - \eta\mathcal{L})^{n-i}(\overline{Z}_i\overline{Y}_i^\top)\right\| \lesssim \tau_*\left[\kappa_*\sqrt{r\widetilde{\eta}\log n}\,(1 + \bar{\nu})\right] \lesssim \frac{\tau_*}{n^\epsilon \log n},$$

where we used Lemma 15 together with $\overline{\tau}_* \asymp \tau_*$.

On the truncation coupling event (D.7), we have

$$\overline{\mathcal{T}}_n = \mathcal{T}_n, \qquad \overline{Z}_i\overline{Y}_i^\top = Z_iY_i^\top, \quad i \in [n].$$

Hence the same bound holds for the original process. The total failure probability is at most

$$\frac{\delta_p}{2} + n^{-18} + n^{-19} \leq \delta_p$$

by (E.1). This proves the proposition.

E.4 Proof of Proposition 4

The proof combines Proposition 3 with Lemma 16.

Lower bound for the linearized term. Let

$$L_n := \eta \sum_{i=1}^{n} (\mathcal{I} - \eta\mathcal{L})^{n-i}(Z_i Y_i^\top).$$

By Lemma 16, under $Y \perp Z$,

$$\mathbb{P}\left\{ \|L_n\|^2 \geq \left(1 - \frac{1}{n}\right) \frac{\widetilde{\eta}\lambda_1\lambda_{1\sim r}}{4\Delta_{\max}\Delta_{\min}} \bar{\nu}^2 \right\} \geq \frac{1}{12}[1 + o(1)].$$

Moreover, when $\kappa_* \lesssim 1$,

$$\frac{\widetilde{\eta}\lambda_1\lambda_{1\sim r}}{\Delta_{\max}\Delta_{\min}} \bar{\nu}^2 \asymp \frac{\tau_*^2}{\log n}.$$

Therefore,

$$\mathbb{P}\left\{ \|L_n\| \gtrsim \frac{\tau_*}{\sqrt{\log n}} \right\} \geq \frac{1}{12}[1 + o(1)].$$

Transfer to the tangent matrix. By Proposition 3, with probability at least $1 - \delta_p$,

$$\|\mathcal{T}_n - L_n\| \lesssim \frac{\tau_*}{n^\epsilon \log n} = o\left(\frac{\tau_*}{\sqrt{\log n}}\right).$$

Intersecting this event with the event above and using the triangle inequality,

$$\|\mathcal{T}_n\| \geq \|L_n\| - \|\mathcal{T}_n - L_n\| \gtrsim \frac{\tau_*}{\sqrt{\log n}}$$

with probability at least

$$\frac{1}{12}[1 + o(1)] - \delta_p.$$

Transfer to the sine distance. By Lemma 39,

$$\|\sin\Theta(U_n, U_*)\| = \|\mathcal{T}_n(\mathcal{C}_n\mathcal{C}_n^\top)^{1/2}\|.$$

Moreover, on the same high-probability event used in the proof of Proposition 3,

$$\|\mathcal{T}_n\| \lesssim \tau_* = o(1).$$

Hence

$$\sigma_{\min}\left((\mathcal{C}_n\mathcal{C}_n^\top)^{1/2}\right) = 1 - o(1),$$

and therefore

$$\|\sin\Theta(U_n, U_*)\| \geq (1 - o(1))\|\mathcal{T}_n\| \gtrsim \frac{\tau_*}{\sqrt{\log n}}.$$

This proves the claim.

F Setup for Gaussian approximation and Bootstrap

F.1 Notation and definitions.

This section introduces the operator-theoretic framework underlying Theorems 5–8. All the results rest on a common set of objects, which we now define.

Hilbert space $\mathcal{H}$. Let $\mathcal{H} = \mathbb{R}^{(d-r)\times r}$ equipped with the Frobenius inner product $\langle\cdot, \cdot\rangle_{\mathrm{F}}$. Every random element ξ of $\mathcal{H}$ with $\mathbb{E}\xi = 0$ and $\mathbb{E}\,\|\xi\|_{\mathrm{F}}^2 < \infty$ induces a *covariance operator* $\mathrm{cov}(\xi)$, which is the unique self-adjoint positive-semidefinite linear operator on $\mathcal{H}$ satisfying

$$\mathrm{cov}(\xi)(A) = \mathbb{E}\left[\langle A, \xi\rangle_{\mathrm{F}}\, \xi\right], \qquad \forall\, A \in \mathcal{H}.$$

Equivalently, for any $A, B \in \mathcal{H}$, $\langle B, \mathrm{cov}(\xi)(A)\rangle_{\mathrm{F}} = \mathbb{E}[\langle A, \xi\rangle_{\mathrm{F}} \langle B, \xi\rangle_{\mathrm{F}}]$. In particular, $\mathrm{tr}\left[\mathrm{cov}(\xi)\right] = \mathbb{E}\,\|\xi\|_{\mathrm{F}}^2$ and $\|\mathrm{cov}(\xi)\|_{\mathrm{op}} = \sup_{\|A\|_{\mathrm{F}}=1} \mathbb{E}\,\langle A, \xi\rangle_{\mathrm{F}}^2$.

Gap operator. Define the linear operator $\mathcal{L} : \mathcal{H} \to \mathcal{H}$ by

$$[\mathcal{L}A]_{lj} = \Delta_{lj} A_{lj}, \qquad \Delta_{lj} := \lambda_j - \lambda_{r+l}, \quad 1 \le l \le d-r,\ 1 \le j \le r.$$

The operator $\mathcal{L}$ is diagonal in the standard basis $\{E_{lj} := e_l e_j^\top\}_{l,j}$ of $\mathcal{H}$ with eigenvalues $\{\Delta_{lj}\}$. Write $\Delta_{\min} := \min_{l,j} \Delta_{lj} > 0$ and $\Delta_{\max} := \max_{l,j} \Delta_{lj}$. Since $\mathcal{L}$ is strictly positive, the semigroup $\{e^{-t\mathcal{L}}\}_{t\ge 0}$ is exponentially stable: $\|e^{-t\mathcal{L}}\|_{\mathrm{op}} \le e^{-t\Delta_{\min}}$ for all $t \ge 0$.

Covariance operator of $ZY^\top$ and $\widetilde{Z}\widetilde{Y}^\top$. Define the operator $\mathfrak{C} : \mathcal{H} \to \mathcal{H}$ as the covariance of the random element $ZY^\top \in \mathcal{H}$:

$$\mathfrak{C}(B) := \mathbb{E}\left[\langle ZY^\top, B\rangle_{\mathcal{H}}\, ZY^\top\right], \qquad B \in \mathcal{H}. \tag{F.1}$$

In coordinates, $\langle E_{l'j'}, \mathfrak{C}(E_{lj})\rangle_{\mathrm{F}} = \mathbb{E}\big([Z]_l\, [Y]_j\, [Z]_{l'}\, [Y]_{j'}\big)$. The operator $\mathfrak{C}$ is self-adjoint and positive-semidefinite on $\mathcal{H}$, with $\mathrm{tr}(\mathfrak{C}) = \mathbb{E}\,\|ZY^\top\|_{\mathrm{F}}^2$ and $\|\mathfrak{C}\|_{\mathrm{HS}} = \big[\sum_{lj,l'j'} \langle E_{l'j'},\, \mathfrak{C}(E_{lj})\rangle_{\mathrm{F}}^2\big]^{1/2}$.

Similarly, define the *whitened covariance operator* $\widetilde{\mathfrak{C}} : \mathcal{H} \to \mathcal{H}$ using the isotropic coordinates $\widetilde{X} = (\widetilde{Y}, \widetilde{Z}) \in \mathbb{R}^r \times \mathbb{R}^{d-r}$ (where $Y = \Lambda_*^{1/2}\widetilde{Y}$, $Z = \Lambda_{*,\perp}^{1/2}\widetilde{Z}$):

$$\widetilde{\mathfrak{C}}(B) := \mathbb{E}\big[\langle \widetilde{Z}\widetilde{Y}^\top,\, B\rangle_{\mathrm{F}}\, \widetilde{Z}\widetilde{Y}^\top\big], \qquad B \in \mathcal{H}.$$

Summands. All the Gaussian approximation and bootstrap theorems rest on the decomposition

$$\frac{1}{\sqrt{\eta}}\left[U_n \operatorname{sgn}(U_n^\top U_*) - U_*\right] \approx \frac{1}{\sqrt{\eta}} U_{*,\perp} T_n \approx U_{*,\perp}\Big[\sum_{i=1}^n \sqrt{\eta}(I - \eta\mathcal{L})^{n-i}\, Z_i Y_i^\top\Big]$$

Thus, we define the *summands* $\xi_i \in \mathcal{H}$ and their covariances Ξ_i by

$$\xi_i := \sqrt{\eta}(\mathcal{I}-\eta\mathcal{L})^{n-i}\,(Z_i Y_i^\top) \in \mathcal{H}, \qquad \Xi_i := \mathrm{cov}(\xi_i) = \eta(\mathcal{I}-\eta\mathcal{L})^{n-i}\mathfrak{C}(\mathcal{I}-\eta\mathcal{L})^{n-i}. \tag{F.2}$$

Note that each Ξ_i is a self-adjoint positive-semidefinite linear operator on $\mathcal{H}$, obtained by conjugating $\mathfrak{C}$ with the contraction $(\mathcal{I} - \eta\mathcal{L})^{n-i}$.

Cumulative covariances. Define the finite-sample and limiting cumulative covariances Γ_n and Γ by

$$\Gamma_n := \sum_{i=1}^n \Xi_i = \eta \sum_{k=0}^{n-1} (\mathcal{I} - \eta\mathcal{L})^k \mathfrak{C} (\mathcal{I} - \eta\mathcal{L})^k, \qquad \Gamma := \int_0^\infty e^{-t\mathcal{L}}\mathfrak{C}e^{-t\mathcal{L}}\, dt.$$

The sum Γ_n is an approximation to Γ. The operator Γ is the unique solution to the *Lyapunov equation* $\mathcal{L}\Gamma + \Gamma\mathcal{L} = \mathfrak{C}$ (Lemma 17).

The whitened analogues $\widetilde{\Gamma}_n$ and $\widetilde{\Gamma}$ are defined identically with $\widetilde{\mathfrak{C}}$ replacing $\mathfrak{C}$.

Projection to $\mathbb{R}^r$. For $u \in \mathbb{R}^{d-r}$, define the bounded linear map $T_u : \mathcal{H} \to \mathbb{R}^r$ by $T_u(B) := B^\top u$. Its adjoint is $T_u^* : \mathbb{R}^r \to \mathcal{H}$, $T_u^*(v) = uv^\top$. Given $u \in \mathbb{R}^{d-r}$, define the *projected summands* $\xi_{u,i}$ and covariances $\Xi_{u,i}$ by

$$\xi_{u,i} := T_u(\xi_i) = \sqrt{\eta}[(\mathcal{I} - \eta\mathcal{L})^{n-i}(Z_i Y_i^\top)]^\top u \in \mathbb{R}^r, \qquad \Xi_{u,i} := \mathrm{cov}(\xi_{u,i}) = T_u \Xi_i T_u^* \in \mathbb{R}^{r\times r},$$

and the cumulative covariances $\Gamma_{u,n}$ and Γ_u by

$$\Gamma_{u,n} := \sum_{i=1}^n \Xi_{u,i} = T_u \Gamma_n T_u^* \in \mathbb{R}^{r\times r}, \qquad \Gamma_u := T_u \Gamma T_u^* \in \mathbb{R}^{r\times r}. \tag{F.3}$$

The whitened analogues $\widetilde{\xi}_{u,i}$, $\widetilde{\Xi}_{u,i}$, $\widetilde{\Gamma}_{u,n}$, and $\widetilde{\Gamma}_u$ are defined identically with $\widetilde{Z}_i\widetilde{Y}_i^\top$ replacing $Z_i Y_i^\top$ and $\widetilde{\mathfrak{C}}$ replacing $\mathfrak{C}$, i.e.

$$\begin{aligned} &\widetilde{\xi}_{u,i} := T_u(\widetilde{\xi}_i) = \sqrt{\eta}[(\mathcal{I} - \eta\mathcal{L})^{n-i}(\widetilde{Z}_i\widetilde{Y}_i^\top)]^\top u \in \mathbb{R}^r, && \widetilde{\Xi}_{u,i} := \mathrm{cov}(\widetilde{\xi}_{u,i}) = T_u\widetilde{\Xi}_i T_u^* \in \mathbb{R}^{r\times r}, \\ &\widetilde{\Gamma}_{u,n} := \sum_{i=1}^n \widetilde{\Xi}_{u,i} = T_u\widetilde{\Gamma}_n T_u^* \in \mathbb{R}^{r\times r}, && \widetilde{\Gamma}_u := T_u\widetilde{\Gamma} T_u^* \in \mathbb{R}^{r\times r}. \end{aligned} \tag{F.4}$$

Summary of notation. The table below collects the key objects given $u \in \mathbb{R}^{d-r}$. The first column concerns the $\mathcal{H}$-level analysis (Theorem 5), while the last three columns concern the projected analysis (Theorem 6).

Table 3: Notation for summands and covariances.

	$\mathcal{H}$-level		Projected to $\mathbb{R}^r$	
	unwhitened	whitened	unwhitened	whitened
summand	ξ_i	$\widetilde{\xi}_i$	$\xi_{u,i} = T_u(\xi_i)$	$\widetilde{\xi}_{u,i} = T_u(\widetilde{\xi}_i)$
covariance	Ξ_i	$\widetilde{\Xi}_i$	$T_u \Xi_i T_u^*$	$T_u \widetilde{\Xi}_i T_u^*$
cum. cov. (finite)	Γ_n	$\widetilde{\Gamma}_n$	$\Gamma_{u,n} = T_u \Gamma_n T_u^*$	$\widetilde{\Gamma}_{u,n} = T_u \widetilde{\Gamma}_n T_u^*$
cum. cov. (limit)	Γ	$\widetilde{\Gamma}$	$\Gamma_u = T_u \Gamma T_u^*$	$\widetilde{\Gamma}_u = T_u \widetilde{\Gamma} T_u^*$

F.2 Technical lemmas

In this section, we collect a few useful results used in Appendix G and H.

Properties of Γ, Γ_n, and ξ_i.

Lemma 17. *Assume $0 < \eta\Delta_{\max} < 1$ and let $\lvert\lvert\lvert\cdot\rvert\rvert\rvert$ denote any unitarily invariant norm on $\mathcal{H}$.*

(i) (Coordinate representation) *For $l, l' \in [d-r]$ and $j, j' \in [r]$,*

$$\langle E_{l'j'}, \Gamma E_{lj} \rangle_{\mathrm{F}} = \frac{\langle E_{l'j'}, \mathfrak{C} E_{lj} \rangle_{\mathrm{F}}}{\Delta_{lj} + \Delta_{l'j'}},$$

$$\langle E_{l'j'}, \Gamma_n E_{lj} \rangle_{\mathrm{F}} = \eta \sum_{i=1}^{n} (1 - \eta\Delta_{lj})^{n-i} (1 - \eta\Delta_{l'j'})^{n-i} \langle E_{l'j'}, \mathfrak{C} E_{lj} \rangle_{\mathrm{F}}.$$

(ii) (Bounds from the Lyapunov equation) *Γ is the unique solution to $\mathcal{L}\Gamma + \Gamma\mathcal{L} = \mathfrak{C}$, with*

$$\frac{\lvert\lvert\lvert\mathfrak{C}\rvert\rvert\rvert}{2\Delta_{\max}} \le \lvert\lvert\lvert\Gamma\rvert\rvert\rvert \le \frac{\lvert\lvert\lvert\mathfrak{C}\rvert\rvert\rvert}{2\Delta_{\min}}, \qquad \text{and} \qquad \Gamma \succeq \frac{\lambda_{\min}(\mathfrak{C})}{2\Delta_{\max}} \mathcal{I}.$$

(iii) (Lower bound for Γ_n) *If $\eta\Delta_{\max} \in (0,1)$, then*

$$\Gamma_n \succeq \frac{\lambda_{\min}(\mathfrak{C})}{2\Delta_{\max}} \left[1 - (1 - \eta\Delta_{\max})^{2n}\right] \mathcal{I}.$$

If $\eta\Delta_{\max} \in (0,1)$ and $n\eta\Delta_{\max} \ge 1$, then

$$\Gamma_n \succeq \frac{\lambda_{\min}(\mathfrak{C})}{3\Delta_{\max}} \mathcal{I}.$$

(iv) (Approximation) *Let $\epsilon_n := (1 - \widetilde{\eta})^{2n} + \widetilde{\eta}\kappa_\Delta^2$ where $\widetilde{\eta} = \eta\Delta_{\min}$. Then*

$$\lvert\lvert\lvert\Gamma_n - \Gamma\rvert\rvert\rvert \le \epsilon_n \lvert\lvert\lvert\Gamma\rvert\rvert\rvert$$

As a consequence, $|\mathrm{tr}(\Gamma_n - \Gamma)| \le \epsilon_n \,\mathrm{tr}\,\Gamma$.

(v) (Moment bound for ξ_i) *For any $q \ge 1$,*

$$\sum_{i=1}^{n} \mathbb{E}\,\|\xi_i\|_{\mathrm{F}}^q \;\le\; \frac{\widetilde{\eta}^{\,q/2-1}}{\Delta_{\min}^{q/2}}\, \mathbb{E}\left\|ZY^\top\right\|_{\mathrm{F}}^q.$$

(vi) (Matrix moment bound) *Under Assumption 1, for any $q \ge 1$,*

$$\left[\mathbb{E}\left\|ZY^\top\right\|_{\mathrm{F}}^{2q}\right]^{1/(2q)} \le Cq[\mathrm{tr}\,\Lambda_*]^{1/2}[\mathrm{tr}\,\Lambda_{*,\perp}]^{1/2}\left\|\widetilde{X}\right\|_{\psi_2}^2,$$

and the whitened analogue holds with Λ_* and $\Lambda_{*,\perp}$ replaced by I_r and I_{d-r}, respectively:

$$\left[\mathbb{E}\left\|\widetilde{Z}\widetilde{Y}^\top\right\|_{\mathrm{F}}^{2q}\right]^{1/(2q)} \le Cqr^{1/2}(d-r)^{1/2}\left\|\widetilde{X}\right\|_{\psi_2}^2.$$

Here $C > 0$ is an absolute constant.

(vii) (Norm of $\widetilde{\mathfrak{C}}$) $\left\|\widetilde{\mathfrak{C}}\right\|_{\mathrm{op}} \lesssim \min\{r, d-r\}$.

Proof. See Appendix F.3.1. $\square$

Remark 10 (Whitened analogue). Lemma 17 holds verbatim with $(\widetilde{\mathfrak{C}}, \widetilde{\Gamma}_n, \widetilde{\Gamma}, \widetilde{\xi}_i)$ replacing $(\mathfrak{C}, \Gamma_n, \Gamma, \xi_i)$, since the proof uses only that $\mathfrak{C}$ is positive-semidefinite and $\mathcal{L}$ is strictly positive.

Lemma 18. *Let $u \in \mathbb{R}^{d-r}$ and $\widetilde{u} = \Lambda_{*,\perp}^{1/2} u$.*

(i) (Coordinate representation) *For any $u \in \mathbb{R}^{d-r}$,*

$$\langle e_{j'}, \Gamma_u e_j \rangle_2 = \sum_{l,l'=1}^{d-r} \frac{u_l\, u_{l'}\, \langle E_{l'j'}, \mathfrak{C}(E_{lj}) \rangle_{\mathrm{F}}}{\Delta_{lj} + \Delta_{l'j'}}, \qquad \forall j, j' \in [r],$$

and in whitened form,

$$\langle e_{j'}, \widetilde{\Gamma}_u e_j \rangle_2 = \sum_{l,l'=1}^{d-r} \frac{\widetilde{u}_l\, \widetilde{u}_{l'} \langle E_{l'j'}, \widetilde{\mathfrak{C}}(E_{lj}) \rangle_{\mathrm{F}}}{\Delta_{lj} + \Delta_{l'j'}}, \qquad \forall j, j' \in [r].$$

(ii) (Whitened–unwhitened connection) *For any $u \in \mathbb{R}^{d-r}$, let $\widetilde{u} = \Lambda_{*,\perp}^{1/2} u$. Then*

$$\Gamma_u = \Lambda_*^{1/2} \widetilde{\Gamma}_{\widetilde{u}} \Lambda_*^{1/2}, \qquad \textit{and} \qquad \Gamma_{u,n} = \Lambda_*^{1/2} \widetilde{\Gamma}_{\widetilde{u},n} \Lambda_*^{1/2}.$$

(iii) (Loewner order lower bound)

$$\widetilde{\Gamma}_{\widetilde{u}} \succeq \frac{\|\widetilde{u}\|^2}{2\Delta_{\max}} \lambda_{\min}(\widetilde{\mathfrak{C}})\, I_r, \qquad \textit{and} \qquad \Gamma_u \succeq \frac{\lambda_r \|\widetilde{u}\|^2}{2\Delta_{\max}} \lambda_{\min}(\widetilde{\mathfrak{C}})\, I_r.$$

If $\eta\Delta_{\max} \in (0,1)$ and $n\eta\Delta_{\max} \geq 1$, then

$$\widetilde{\Gamma}_{\widetilde{u},n} \succeq \frac{\|\widetilde{u}\|^2}{3\Delta_{\max}} \lambda_{\min}(\widetilde{\mathfrak{C}})\, I_r, \qquad \textit{and} \qquad \Gamma_{u,n} \succeq \frac{\lambda_r \|\widetilde{u}\|^2}{3\Delta_{\max}} \lambda_{\min}(\widetilde{\mathfrak{C}})\, I_r.$$

(iv) (Approximation) *With ϵ_n as in Lemma 17(iv). If $\lambda_{\min}(\widetilde{\mathfrak{C}}) > 0$, then*

$$\left|\!\left|\!\left| \widetilde{\Gamma}_{\widetilde{u},n} - \widetilde{\Gamma}_{\widetilde{u}} \right|\!\right|\!\right| \leq \epsilon_n \kappa_\Delta \frac{\|\widetilde{\mathfrak{C}}\|_{\mathrm{op}}}{\lambda_{\min}(\widetilde{\mathfrak{C}})} \left|\!\left|\!\left| \widetilde{\Gamma}_{\widetilde{u}} \right|\!\right|\!\right|.$$

(v) (Independence case) *If $\widetilde{Y} \perp \widetilde{Z}$, then $\widetilde{C} = \mathcal{I}$ and*

$$[\widetilde{\Gamma}_{\widetilde{u}}]_{jj'} = \delta_{jj'} \sum_{l=1}^{d-r} \frac{\widetilde{u}_l^2}{2\Delta_{lj}}, \qquad [\Gamma_u]_{jj'} = \delta_{jj'} \sum_{l=1}^{d-r} \frac{\lambda_j \widetilde{u}_l^2}{2\Delta_{lj}}.$$

Proof. See Appendix F.3.2. $\square$

Lemma 19. *For independent $X_1, \ldots, X_n \geq 0$ and $p \geq 1$,*

$$\mathbb{E}\Big(\sum_{i=1}^n X_i\Big)^p \leq C_p \left[\Big(\sum_{i=1}^n \mathbb{E} X_i\Big)^p + \sum_{i=1}^n \mathbb{E} X_i^p\right]$$

Let $\mathfrak{g} \sim \mathcal{N}(0, I_d)$ be a Gaussian vector in $\mathbb{R}^d$. If $G \succeq 0$, then for $p \geq 1$,

$$\mathbb{E}\left(\mathfrak{g}^\top G \mathfrak{g}\right)^p \lesssim_p [\operatorname{tr} G]^p$$

Proof. See Appendix F.3.3. $\square$

Lemma 20 (weighted χ^2 anti-concentration)**.** *Let $a_1 \geq \cdots \geq a_p \geq 0$ such that $\sum_{i=1}^p a_i^2 = 1$ and $\omega_1, \ldots, \omega_p$ be i.i.d. χ_1^2 random variables. Then*

$$\sup_{t \in \mathbb{R}} \mathbb{P}\Big(t \leq \sum_{i=1}^p a_i \omega_i \leq t + h\Big) \leq \sqrt{\frac{4h}{\pi}}$$

Proof. See [26, Proposition B.7]. □

Lemma 21 (χ^2 concentration)**.** *Let $\mathfrak{g} \sim \mathcal{N}(0, \mathcal{I})$ be a Gaussian random element in $\mathcal{H}$ with identity covariance. Then for every $t \geq 0$, we have*

$$\mathbb{P}\left\{\left|\mathfrak{g}^\top \Gamma \mathfrak{g} - \operatorname{tr}\Gamma\right| \geq t\right\} \leq 2\exp\Big\{-c\Big(\frac{t^2}{\|\Gamma\|_{\mathrm{HS}}^2} \bigwedge \frac{t}{\|\Gamma\|_{\mathrm{op}}}\Big)\Big\}$$

As a consequence, with probability exceeding $1 - n^{-20}$,

$$\left|\mathfrak{g}^\top \Gamma \mathfrak{g} - \operatorname{tr}\Gamma\right| \lesssim \|\Gamma\|_{\mathrm{HS}} \sqrt{\log n} + \|\Gamma\|_{\mathrm{op}} \log n$$

Proof. See [Hanson-Wright inequality from 29, Theorem 6.2.2.] □

Lemma 22 (χ^2 mixture comparison)**.** *Suppose $\mathfrak{g}$ and $\widetilde{\mathfrak{g}}$ are centered Gaussian random elements in $\mathcal{H}$ with $\mathfrak{g} \sim \mathcal{N}(0, \Gamma)$, $\widetilde{\mathfrak{g}} \sim \mathcal{N}(0, \widetilde{\Gamma})$. Let $\Delta_\Gamma = \Gamma - \widetilde{\Gamma}$. If $\|\Gamma\|_{\mathrm{HS}} > 0$, then for any $t > 0$,*

$$d_K\Big(\|\mathfrak{g}\|_{\mathrm{F}}^2, \|\widetilde{\mathfrak{g}}\|_{\mathrm{F}}^2\Big) \lesssim \sqrt{\frac{|\operatorname{tr}\Delta_\Gamma| + t}{\|\Gamma\|_{\mathrm{HS}}}} + \exp\Big\{-c\Big(\frac{t^2}{\|\Delta_\Gamma\|_{\mathrm{HS}}^2} \bigwedge \frac{t}{\|\Delta_\Gamma\|_{\mathrm{op}}}\Big)\Big\}$$

Proof. See [26, Lemma C.5]. □

Lemma 23 (High-dimensional CLT)**.** *Let $\{\xi_i\}_{i=1}^n$ be independent random variables in $\mathcal{H}$ such that $\mathbb{E}\xi_i = 0$ and $\operatorname{cov}\xi_i = \Xi_i$. Let $\{\mathfrak{g}_i\}_{i=1}^n$ be independent Gaussian analogues of $\{\xi_i\}_{i=1}^n$ with $\mathfrak{g}_i \sim \mathcal{N}(0, \Xi_i)$. Furthermore, let $\Gamma_n = \sum_{i=1}^n \Xi_i$. For $0 < \delta \leq 1$, $q = 2 + \delta$ and $\beta \geq 2$ define the following quantities:*

$$\begin{aligned}
L_q^\xi &= \sum_{i=1}^n \frac{\mathbb{E}\left\langle \xi_i,\, \Gamma_n \xi_i\right\rangle_{\mathrm{F}}^{q/2}}{\|\Gamma_n\|_{\mathrm{HS}}^q} + \frac{n^2}{\binom{n}{2}} \sum_{1 \leq i < j \leq n} \frac{\mathbb{E}\left|\langle \xi_i,\, \xi_j\rangle_{\mathrm{F}}\right|^q}{\|\Gamma_n\|_{\mathrm{HS}}^q} \\
L_q^{\mathfrak{g}} &= \sum_{i=1}^n \frac{\mathbb{E}\left\langle \mathfrak{g}_i,\, \Gamma_n \mathfrak{g}_i\right\rangle_{\mathrm{F}}^{q/2}}{\|\Gamma_n\|_{\mathrm{HS}}^q} \\
K_\beta^\beta &= \frac{n^\beta}{n} \sum_{i=1}^n \mathbb{E}\left|\frac{\langle \xi_i,\, \xi_i\rangle_{\mathrm{F}} - \mathbb{E}\langle \xi_i,\, \xi_i\rangle_{\mathrm{F}}}{\|\Gamma_n\|_{\mathrm{HS}}}\right|^\beta \\
J_n &= \frac{\sum_{i=1}^n \mathfrak{v}_i}{\|\Gamma_n\|_{\mathrm{HS}}^2}, \qquad \textit{where} \quad \mathfrak{v}_i = \operatorname{Var}(\|\xi_i\|_{\mathrm{F}}^2)
\end{aligned}$$

Suppose $L_q^\xi \to 0$, $L_q^{\mathfrak{g}} \to 0$, $n^{1-\beta} K_\beta^\beta \to 0$, and $J_n \to 0$. Then

$$d_K\Big(\Big\|\sum_{i=1}^n \xi_i\Big\|_{\mathrm{F}}^2, \Big\|\sum_{i=1}^n \mathfrak{g}_i\Big\|_{\mathrm{F}}^2\Big) \to 0$$

Proof. See [26, Proposition B.5]. □

Lemma 24 (G-approx additive remainder)**.** *Let $\mathsf{A}_n, \mathsf{E}_n \in \mathbb{R}^{d_n}$ be random vectors, and define $\mathsf{B}_n := \mathsf{A}_n + \mathsf{E}_n$. Let $\mathsf{G}_n \sim N(0, \Gamma_n)$ be a centered Gaussian vector in $\mathbb{R}^{d_n}$. Define*

$$\begin{aligned}
\delta_n &:= \sup_{t \in \mathbb{R}} \Big|\mathbb{P}\big\{\|\mathsf{A}_n\|^2 \leq t\big\} - \mathbb{P}\big\{\|\mathsf{G}_n\|^2 \leq t\big\}\Big|, \\
w_n(h) &:= \sup_{t \in \mathbb{R}} \mathbb{P}\big\{t < \|\mathsf{G}_n\|^2 \leq t + h\big\}, \quad \forall h \geq 0.
\end{aligned}$$

For any deterministic sequences $e_n, L_n > 0$, define $h_n := 2L_n e_n + e_n^2$. Then

$$\sup_{t \in \mathbb{R}} \Big|\mathbb{P}\big\{\|\mathsf{B}_n\|^2 \leq t\big\} - \mathbb{P}\big\{\|\mathsf{G}_n\|^2 \leq t\big\}\Big| \leq 2\delta_n + w_n(h_n) + \mathbb{P}\big\{\|\mathsf{E}_n\| > e_n\big\} + \mathbb{P}\big\{\|\mathsf{G}_n\| > L_n\big\}.$$

Proof. See Appendix F.3.4. □

Lemma 25 (G-approx multiplicative remainder)**.** *Let random $\mathsf{A}_n, \mathsf{B}_n \in \mathbb{R}^{d_n}$, and suppose $\mathsf{B}_n = (I_{d_n} + \mathsf{E}_n)\mathsf{A}_n$ for some random $\mathsf{E}_n \in \mathbb{R}^{d_n \times d_n}$. Let $\mathsf{G}_n \sim N(0, \Gamma_n) \in \mathbb{R}^{d_n}$. Define*

$$
\begin{aligned}
\delta_n &:= \sup_{t\in\mathbb{R}} \Big| \mathbb{P}\big\{ \|\mathsf{A}_n\|^2 \le t \big\} - \mathbb{P}\big\{ \|\mathsf{G}_n\|^2 \le t \big\} \Big|, \\
w_n(h) &:= \sup_{t\in\mathbb{R}} \mathbb{P}\big\{ t < \|\mathsf{G}_n\|^2 \le t + h \big\}, \quad \forall h \ge 0.
\end{aligned}
$$

Let $e_n \in [0, 1/2]$ and $L_n > 0$ be deterministic sequences, and define $h_n := \big[\frac{1}{(1-e_n)^2} - 1\big] L_n^2$. Then

$$
\sup_{t\in\mathbb{R}} \Big| \mathbb{P}\big\{ \|\mathsf{B}_n\|^2 \le t \big\} - \mathbb{P}\big\{ \|\mathsf{G}_n\|^2 \le t \big\} \Big| \le \delta_n + w_n(h_n) + \mathbb{P}(\|\mathsf{E}_n\| > e_n) + \mathbb{P}(\|\mathsf{G}_n\| > L_n).
$$

Proof. See Appendix F.3.5. □

F.3 Proof of technical lemmas

F.3.1 Proof of Lemma 17

(i) Coordinate representation. Note that $\mathcal{L}$ is diagonal under the basis $\{E_{lj}\}_{l,j}$. For any $l, l' \in [d-r]$ and $j, j' \in [r]$, we have $\mathcal{L}E_{lj} = \Delta_{lj} E_{lj}$ with $\Delta_{lj} = \lambda_j - \lambda_{r+l} > 0$. By the definition of Γ, we have

$$
\begin{aligned}
\langle E_{l'j'}, \Gamma E_{lj} \rangle_{\mathrm{F}} &= \int_0^\infty \big\langle e^{-t\mathcal{L}} E_{l'j'}, \mathfrak{C} e^{-t\mathcal{L}} E_{lj} \big\rangle_{\mathrm{F}} \, dt \\
&= \int_0^\infty e^{-t(\Delta_{lj} + \Delta_{l'j'})} dt \cdot \langle E_{l'j'}, \mathfrak{C} E_{lj} \rangle_{\mathrm{F}} = \frac{\langle E_{l'j'}, \mathfrak{C} E_{lj} \rangle_{\mathrm{F}}}{\Delta_{lj} + \Delta_{l'j'}}
\end{aligned}
$$

Similarly for Γ_n, we have

$$
\begin{aligned}
\langle E_{l'j'}, \Gamma_n E_{lj} \rangle_{\mathrm{F}} &= \eta \sum_{i=1}^n \big\langle (\mathcal{I} - \eta\mathcal{L})^{n-i} E_{l'j'}, \mathfrak{C} (\mathcal{I} - \eta\mathcal{L})^{n-i} E_{lj} \big\rangle_{\mathrm{F}} \\
&= \eta \sum_{i=1}^n (1 - \eta\Delta_{lj})^{n-i} (1 - \eta\Delta_{l'j'})^{n-i} \cdot \langle E_{l'j'}, \mathfrak{C} E_{lj} \rangle_{\mathrm{F}}
\end{aligned}
$$

(ii) Bounds from the Lyapunov equation. We first show that Γ is the unique solution $\mathcal{L}\Gamma + \Gamma\mathcal{L} = \mathfrak{C}$. Note that

$$
\begin{aligned}
\mathcal{L}\Gamma_* + \Gamma_*\mathcal{L} = \mathfrak{C} &\iff \langle E_{l'j'}, (\mathcal{L}\Gamma_* + \Gamma_*\mathcal{L}) E_{lj} \rangle_{\mathrm{F}} = \langle E_{l'j'}, \mathfrak{C} E_{lj} \rangle_{\mathrm{F}}, \quad \forall l, l' \in [d-r]; j, j' \in [r] \\
&\iff \langle \mathcal{L} E_{l'j'}, \Gamma_* E_{lj} \rangle_{\mathrm{F}} + \langle E_{l'j'}, \Gamma_* \mathcal{L} E_{lj} \rangle_{\mathrm{F}} = \langle E_{l'j'}, \mathfrak{C} E_{lj} \rangle_{\mathrm{F}} \\
&\iff \langle E_{l'j'}, \Gamma_* E_{lj} \rangle_{\mathrm{F}} = \frac{\langle E_{l'j'}, \mathfrak{C} E_{lj} \rangle_{\mathrm{F}}}{\Delta_{lj} + \Delta_{l'j'}} \\
&\overset{(i)}{\iff} \langle E_{l'j'}, \Gamma_* E_{lj} \rangle_{\mathrm{F}} = \langle E_{l'j'}, \Gamma E_{lj} \rangle_{\mathrm{F}}.
\end{aligned} \tag{F.5}
$$

Here (i) follows from the coordinate representation of Γ. Hence Γ is the unique solution to the Lyapunov equation.

Next, we prove the two-sided bound for unitarily invariant norm $|||\cdot|||$. For the upper bound, taking $|||\cdot|||$ on the integral definition $\Gamma = \int_0^\infty e^{-t\mathcal{L}} \mathfrak{C} e^{-t\mathcal{L}} dt$ gives

$$
|||\Gamma||| \le \int_0^\infty \big|\big|\big| e^{-t\mathcal{L}} \mathfrak{C} e^{-t\mathcal{L}} \big|\big|\big| \, dt \overset{(i)}{\le} \int_0^\infty \big\| e^{-t\mathcal{L}} \big\|_{\mathrm{op}}^2 |||\mathfrak{C}||| \, dt \le \int_0^\infty e^{-2t\Delta_{\min}} |||\mathfrak{C}||| \, dt = \frac{|||\mathfrak{C}|||}{2\Delta_{\min}}
$$

For the lower bound, taking $|||\cdot|||$ on both sides of the Lyapunov equation we have

$$
|||\mathfrak{C}||| = |||\mathcal{L}\Gamma + \Gamma\mathcal{L}||| \overset{(i)}{\le} 2 \|\mathcal{L}\|_{\mathrm{op}} |||\Gamma||| \le 2\Delta_{\max} |||\Gamma|||
$$

For both bounds, (i) follows from the ideal property of unitarily invariant norms.

Then, we establish the Loewner order lower bound. Let $\mathcal{A} = \frac{\lambda_{\min}(\mathfrak{C})}{2}\mathcal{L}^{-1}$. Then we have

$$\mathcal{L}\mathcal{A} + \mathcal{A}\mathcal{L} = \lambda_{\min}(\mathfrak{C})\mathcal{I} \preceq \mathfrak{C} = \mathcal{L}\Gamma + \Gamma\mathcal{L}$$

which implies that $\mathcal{B} = \Gamma - \mathcal{A}$ solves the Lyapunov equation $\mathcal{L}\mathcal{B} + \mathcal{B}\mathcal{L} = \mathfrak{C} - \lambda_{\min}(\mathfrak{C})\mathcal{I} \succeq 0$. The same argument as (F.5) shows that the unique solution is

$$\Gamma - \mathcal{A} = \int_0^\infty e^{-t\mathcal{L}}(\mathfrak{C} - \lambda_{\min}(\mathfrak{C})\mathcal{I})e^{-t\mathcal{L}}dt \succeq 0$$

Hence $\Gamma \succeq \mathcal{A}$. Finally, since $\mathcal{L}^{-1}$ is diagonal with eigenvalues $\Delta_{lj}^{-1} \geq \Delta_{\max}^{-1}$, we have $\mathcal{A} \succeq \frac{\lambda_{\min}(\mathfrak{C})}{2\Delta_{\max}}\mathcal{I}$ and consequently $\Gamma \succeq \frac{\lambda_{\min}(\mathfrak{C})}{2\Delta_{\max}}\mathcal{I}$.

(iii) Lower bound for Γ_n. Since $\mathcal{L}$ is diagonal with eigenvalues in $[\Delta_{\min}, \Delta_{\max}]$, we have

$$(\mathcal{I} - \eta\mathcal{L})^{n-i} \succeq (1 - \eta\Delta_{\max})^{n-i}\mathcal{I}, \qquad \forall i = 1, 2, \ldots, n$$

which combined with $\mathfrak{C} \succeq \lambda_{\min}(\mathfrak{C})\mathcal{I}$ gives

$$(\mathcal{I} - \eta\mathcal{L})^{n-i}\mathfrak{C}(\mathcal{I} - \eta\mathcal{L})^{n-i} \succeq \lambda_{\min}(\mathfrak{C})(\mathcal{I} - \eta\mathcal{L})^{2(n-i)} \succeq \lambda_{\min}(\mathfrak{C})(1 - \eta\Delta_{\max})^{2(n-i)}\mathcal{I}$$

Summing over i gives

$$\Gamma_n = \eta\sum_{i=1}^{n}(\mathcal{I} - \eta\mathcal{L})^{n-i}\mathfrak{C}(\mathcal{I} - \eta\mathcal{L})^{n-i} \succeq \eta\lambda_{\min}(\mathfrak{C})\sum_{i=1}^{n}(1 - \eta\Delta_{\max})^{2(n-i)}\mathcal{I}$$

Therefore, we have

$$\lambda_{\min}(\Gamma_n) \geq \lambda_{\min}(\mathfrak{C})\frac{1 - (1 - \eta\Delta_{\max})^{2n}}{\Delta_{\max}\left(2 - \eta\Delta_{\max}\right)}$$

If further $n\eta\Delta_{\max} \geq 1$, then $(1 - \eta\Delta_{\max})^{2n} \leq \exp\left(-2n\eta\Delta_{\max}\right) < e^{-2}$. Hence $1 - (1 - \eta\Delta_{\max})^{2n} \geq 1 - e^{-2}$. Since $2 - \eta\Delta_{\max} < 2$, we have

$$\lambda_{\min}(\Gamma_n) \geq \lambda_{\min}(\mathfrak{C})\frac{1 - e^{-2}}{2\Delta_{\max}} > \frac{\lambda_{\min}(\mathfrak{C})}{3\Delta_{\max}}$$

where the last inequality uses $(1 - e^{-2})/2 > 1/3$.

(iv) Approximation. By definition of Γ_n, for $n \geq 1$, we have

$$\Gamma_n = \eta\sum_{i=1}^{n}(\mathcal{I} - \eta\mathcal{L})^{n-i}\mathfrak{C}(\mathcal{I} - \eta\mathcal{L})^{n-i} \overset{(i)}{=} \eta\sum_{k=0}^{n-1}(\mathcal{I} - \eta\mathcal{L})^{k}\mathfrak{C}(\mathcal{I} - \eta\mathcal{L})^{k}$$

Here (i) follows from change of variable $k = n - i$. Let $\Gamma_0 = 0$. Then we have recursion formula

$$\Gamma_n = (\mathcal{I} - \eta\mathcal{L})\Gamma_{n-1}(\mathcal{I} - \eta\mathcal{L}) + \eta\mathfrak{C}, \qquad n = 1, 2, \ldots$$

Subtracting Γ on both sides and applying the Lyapunov equation, we obtain that for any $n \geq 1$,

$$\begin{aligned}
\Gamma_n - \Gamma &= (\mathcal{I} - \eta\mathcal{L})\left(\Gamma_{n-1} - \Gamma\right)(\mathcal{I} - \eta\mathcal{L}) - \eta\underbrace{\left[\mathcal{L}\Gamma + \Gamma\mathcal{L} - \mathfrak{C}\right]}_{=0} + \eta^2\mathcal{L}\Gamma\mathcal{L} \\
&\overset{(i)}{=} (\mathcal{I} - \eta\mathcal{L})^n\left(\Gamma_0 - \Gamma\right)(\mathcal{I} - \eta\mathcal{L})^n + \eta^2\sum_{k=0}^{n-1}(\mathcal{I} - \eta\mathcal{L})^k\mathcal{L}\Gamma\mathcal{L}(\mathcal{I} - \eta\mathcal{L})^k \\
&= -(\mathcal{I} - \eta\mathcal{L})^n\Gamma(\mathcal{I} - \eta\mathcal{L})^n + \eta^2\sum_{k=0}^{n-1}(\mathcal{I} - \eta\mathcal{L})^k\mathcal{L}\Gamma\mathcal{L}(\mathcal{I} - \eta\mathcal{L})^k
\end{aligned}$$

Here (i) follows by iterating the recursion formula.

Taking $\|\!|\cdot|\!\|$ on both sides of the above equation, by triangle inequality and the ideal property of $\|\!|\cdot|\!\|$, we have

$$\|\!|\Gamma_n - \Gamma|\!\| \leq \|\mathcal{I} - \eta\mathcal{L}\|_{\text{op}}^{2n}\|\!|\Gamma|\!\| + \eta^2\sum_{k=0}^{n-1}\|\mathcal{I} - \eta\mathcal{L}\|_{\text{op}}^{2k}\|\mathcal{L}\|_{\text{op}}^2\|\!|\Gamma|\!\|$$

$$
\begin{aligned}
&\leq (1-\eta\Delta_{\min})^{2n} \left\|\!\left|\Gamma\right|\!\right\| + \eta^2 \sum_{k=0}^{n-1} (1-\eta\Delta_{\min})^{2k} \Delta_{\max}^2 \left\|\!\left|\Gamma\right|\!\right\| \\
&\overset{(i)}{\leq} (1-\eta\Delta_{\min})^{2n} \left\|\!\left|\Gamma\right|\!\right\| + \eta \frac{\Delta_{\max}^2}{\Delta_{\min}} \left\|\!\left|\Gamma\right|\!\right\| \\
&= (1-\widetilde{\eta})^{2n} \left\|\!\left|\Gamma\right|\!\right\| + \widetilde{\eta} \frac{\Delta_{\max}^2}{\Delta_{\min}^2} \left\|\!\left|\Gamma\right|\!\right\|
\end{aligned}
$$

Here (i) follows from the fact that $\sum_{k=0}^{n-1}(1-\eta\Delta_{\min})^{2k} \leq \sum_{k=0}^{\infty}(1-\eta\Delta_{\min})^{k} \leq 1/(\eta\Delta_{\min})$.

For the trace bound, we apply the above bound to the nuclear norm $\|\cdot\|_*$, which is unitarily invariant. Since Γ_n and Γ are both PSD, we have

$$
|\operatorname{tr}(\Gamma_n - \Gamma)| \leq \|\Gamma_n - \Gamma\|_* \leq \left[(1-\widetilde{\eta})^{2n} + \widetilde{\eta}\frac{\Delta_{\max}^2}{\Delta_{\min}^2}\right] \|\Gamma\|_* = \left[(1-\widetilde{\eta})^{2n} + \widetilde{\eta}\frac{\Delta_{\max}^2}{\Delta_{\min}^2}\right] \operatorname{tr}\Gamma
$$

(v) Moment bound for ξ_i. By definition of ξ_i, for any $q \geq 1$, we have

$$
\|\xi_i\|_{\mathrm{F}}^q \leq \eta^{q/2} \|\mathcal{I} - \eta\mathcal{L}\|_{\mathrm{op}}^{(n-i)q} \left\|ZY^\top\right\|_{\mathrm{F}}^q \leq \eta^{q/2}(1-\eta\Delta_{\min})^{(n-i)q} \left\|ZY^\top\right\|_{\mathrm{F}}^q
$$

Summing over $i = 1, 2, \ldots, n$ and taking expectation on both sides gives

$$
\begin{aligned}
\sum_{i=1}^n \mathbb{E}\,\|\xi_i\|_{\mathrm{F}}^q &\leq \eta^{q/2} \sum_{i=1}^n (1-\eta\Delta_{\min})^{(n-i)q} \mathbb{E}\left\|ZY^\top\right\|_{\mathrm{F}}^q \\
&\leq \frac{\eta^{q/2-1}}{\Delta_{\min}} \mathbb{E}\left\|ZY^\top\right\|_{\mathrm{F}}^q = \frac{\widetilde{\eta}^{q/2-1}}{\Delta_{\min}^{q/2}} \mathbb{E}\left\|ZY^\top\right\|_{\mathrm{F}}^q
\end{aligned}
$$

(vi) Matrix moment bound. Applying Lemma 38 to the random matrix $ZY^\top$ gives the desired bound.

(vii) $\|\widetilde{\mathfrak{C}}\|_{\mathrm{op}}$ upper bound follows from Lemma 38.

F.3.2 Proof of Lemma 18

(i) (Coordinate representation.) By the definition of Γ_u, for any $j \in [r]$, we have

$$
\begin{aligned}
\langle e_{j'},\, \Gamma_u e_j\rangle_{\mathrm{F}} = \langle T_u^* e_{j'},\, \Gamma T_u^* e_j\rangle_{\mathrm{F}} = \left\langle u e_{j'}^\top,\, \Gamma(u e_j^\top)\right\rangle_{\mathrm{F}} &= \sum_{l,l'=1}^{d-r} u_l u_{l'} \langle E_{l'j'},\, \Gamma E_{lj}\rangle_{\mathrm{F}} \\
&\overset{(i)}{=} \sum_{l,l'=1}^{d-r} \frac{u_l u_{l'}}{\Delta_{lj} + \Delta_{l'j'}} \langle E_{l'j'},\, \mathfrak{C} E_{lj}\rangle_{\mathrm{F}}
\end{aligned}
$$

Here (i) follows from the coordinate representation of Γ in Lemma 17(i). The coordinate representation for $\widetilde{\Gamma}_{\widetilde{u}}$ follows similarly.

(ii) (Whitened–unwhitened connection) For any $j, j' \in [r]$, we have

$$
\begin{aligned}
\left\langle e_{j'},\, T_u e^{-t\mathcal{L}} \mathfrak{C} e^{-t\mathcal{L}} T_u^* e_j\right\rangle_2 &= \left\langle T_u^* e_{j'},\, e^{-t\mathcal{L}} \mathfrak{C} e^{-t\mathcal{L}} T_u^* e_j\right\rangle_{\mathrm{F}} \\
&\overset{(i)}{=} \mathbb{E}\left\langle u e_{j'}^\top,\, e^{-t\mathcal{L}} Z Y^\top\right\rangle_{\mathrm{F}} \left\langle e^{-t\mathcal{L}} Z Y^\top,\, u e_j^\top\right\rangle_{\mathrm{F}} \\
&\overset{(ii)}{=} \mathbb{E}\left\langle u e_{j'}^\top,\, \Lambda_{*,\perp}^{1/2} e^{-t\mathcal{L}} \widetilde{Z}\widetilde{Y}^\top \Lambda_*^{1/2}\right\rangle_{\mathrm{F}} \left\langle \Lambda_{*,\perp}^{1/2} e^{-t\mathcal{L}} \widetilde{Z}\widetilde{Y}^\top \Lambda_*^{1/2},\, u e_j^\top\right\rangle_{\mathrm{F}} \\
&= \mathbb{E}\left\langle \Lambda_{*,\perp}^{1/2} u e_{j'}^\top \Lambda_*^{1/2},\, e^{-t\mathcal{L}} \widetilde{Z}\widetilde{Y}^\top\right\rangle_{\mathrm{F}} \left\langle e^{-t\mathcal{L}} \widetilde{Z}\widetilde{Y}^\top,\, \Lambda_{*,\perp}^{1/2} u e_j^\top \Lambda_*^{1/2}\right\rangle_{\mathrm{F}} \\
&= \mathbb{E}\left\langle \widetilde{u}\left[\Lambda_*^{1/2} e_{j'}\right]^\top,\, e^{-t\mathcal{L}} \widetilde{Z}\widetilde{Y}^\top\right\rangle_{\mathrm{F}} \left\langle e^{-t\mathcal{L}} \widetilde{Z}\widetilde{Y}^\top,\, \widetilde{u}\left[\Lambda_*^{1/2} e_j\right]^\top\right\rangle_{\mathrm{F}} \\
&= \left\langle T_{\widetilde{u}}^*\left[\Lambda_*^{1/2} e_{j'}\right],\, e^{-t\mathcal{L}} \widetilde{\mathfrak{C}} e^{-t\mathcal{L}} T_{\widetilde{u}}^*\left[\Lambda_*^{1/2} e_j\right]\right\rangle_{\mathrm{F}}
\end{aligned}
$$

$$= \left\langle e_{j'},\, \Lambda_*^{1/2} T_{\widetilde{u}} e^{-t\mathcal{L}} \widetilde{\mathfrak{C}} e^{-t\mathcal{L}} T_{\widetilde{u}}^* \Lambda_*^{1/2} e_j \right\rangle_2$$

Here (i) follows from the definition of $\mathfrak{C}$ in (F.1), $T_u^* e_j = u e_j^\top$ and $e^{-t\mathcal{L}}$ is PSD; (ii) follows due to the commutation $\Lambda_{*,\perp}^{1/2} e^{-t\mathcal{L}}(\widetilde{Z} Y^\top) = e^{-t\mathcal{L}}(\Lambda_{*,\perp}^{1/2} \widetilde{Z} Y^\top)$.

Integrating over t from 0 to ∞ gives

$$\underbrace{T_u \Gamma T_u^*}_{=\Gamma_u} = \Lambda_*^{1/2} \underbrace{T_{\widetilde{u}} \widetilde{\Gamma} T_{\widetilde{u}}^*}_{=\widetilde{\Gamma}_{\widetilde{u}}} \Lambda_*^{1/2}$$

The proof for the connection between $\Gamma_{u,n}$ and $\widetilde{\Gamma}_{\widetilde{u},n}$ follows similarly.

(iii) (Loewner order lower bound) For any unit vector $v \in \mathbb{R}^r$, we have

$$\left\langle v,\, \widetilde{\Gamma}_{\widetilde{u}} v \right\rangle_2 = \left\langle \widetilde{u} v^\top,\, \widetilde{\Gamma} \widetilde{u} v^\top \right\rangle_{\mathrm{F}} \overset{(i)}{\geq} \frac{\lambda_{\min}(\widetilde{\mathfrak{C}}) \left\langle \widetilde{u} v^\top,\, \widetilde{u} v^\top \right\rangle_{\mathrm{F}}}{2\Delta_{\max}} = \frac{\lambda_{\min}(\widetilde{\mathfrak{C}}) \left\|\widetilde{u}\right\|^2}{2\Delta_{\max}}$$

Here (i) follows from the Loewner order lower bound for Γ in Lemma 17(iii) and Remark 10. Thus we get the lower bound for $\widetilde{\Gamma}_{\widetilde{u}}$. The lower bound for Γ_u follows from the connection between Γ_u and $\widetilde{\Gamma}_{\widetilde{u}}$ in part (ii) of this lemma. The lower bound for $\widetilde{\Gamma}_{\widetilde{u},n}$ and $\Gamma_{u,n}$ follows similarly by applying the same argument to Γ_n and using the lower bound for Γ_n in Lemma 17(iii).

(iv) (Approximation) By the definition of $\widetilde{\Gamma}_{\widetilde{u}}$ and $\widetilde{\Gamma}_{\widetilde{u},n}$, we have

$$\left|\!\left|\!\left| \widetilde{\Gamma}_{\widetilde{u},n} - \widetilde{\Gamma}_{\widetilde{u}} \right|\!\right|\!\right| = \left|\!\left|\!\left| T_{\widetilde{u}} (\widetilde{\Gamma}_n - \widetilde{\Gamma}) T_{\widetilde{u}}^* \right|\!\right|\!\right| \overset{(i)}{\leq} \left|\!\left|\!\left| \widetilde{\Gamma}_n - \widetilde{\Gamma} \right|\!\right|\!\right| \|\widetilde{u}\|^2 \leq \varepsilon_n \left|\!\left|\!\left| \widetilde{\Gamma} \right|\!\right|\!\right| \|\widetilde{u}\|^2 \leq \varepsilon_n \left\|\widetilde{\Gamma}\right\|_{\mathrm{op}} |\!|\!| \mathcal{I} |\!|\!| \, \|\widetilde{u}\|^2$$

Here (i) follows from the ideal property of unitarily invariant norms and the fact that $\|T_{\widetilde{u}}\|_{\mathrm{op}} = \|T_{\widetilde{u}}^*\|_{\mathrm{op}} = \|\widetilde{u}\|$. Hence

$$\begin{aligned}
\left|\!\left|\!\left| \widetilde{\Gamma}_{\widetilde{u},n} - \widetilde{\Gamma}_{\widetilde{u}} \right|\!\right|\!\right| &\leq \varepsilon_n \|\widetilde{\Gamma}\|_{\mathrm{op}} \, |\!|\!| \mathcal{I} |\!|\!| \, \|\widetilde{u}\|^2 = \varepsilon_n \|\widetilde{\Gamma}\|_{\mathrm{op}} \lambda_{\min}(\widetilde{\Gamma}_{\widetilde{u}}) \, |\!|\!| \mathcal{I} |\!|\!| \, \|\widetilde{u}\|^2 \cdot \frac{1}{\lambda_{\min}(\widetilde{\Gamma}_{\widetilde{u}})} \\
&\overset{(i)}{\leq} \varepsilon_n \|\widetilde{\Gamma}\|_{\mathrm{op}} \left|\!\left|\!\left| \widetilde{\Gamma}_{\widetilde{u}} \right|\!\right|\!\right| \|\widetilde{u}\|^2 \cdot \frac{2\Delta_{\max}}{\lambda_{\min}(\widetilde{\mathfrak{C}}) \|\widetilde{u}\|^2} \overset{(ii)}{\leq} \varepsilon_n \frac{\|\widetilde{\mathfrak{C}}\|_{\mathrm{op}}}{2\Delta_{\min}} \left|\!\left|\!\left| \widetilde{\Gamma}_{\widetilde{u}} \right|\!\right|\!\right| \|\widetilde{u}\|^2 \cdot \frac{2\Delta_{\max}}{\lambda_{\min}(\widetilde{\mathfrak{C}}) \|\widetilde{u}\|^2} \\
&= \varepsilon_n \frac{\Delta_{\max}}{\Delta_{\min}} \frac{\|\widetilde{\mathfrak{C}}\|_{\mathrm{op}}}{\lambda_{\min}(\widetilde{\mathfrak{C}})} \left|\!\left|\!\left| \widetilde{\Gamma}_{\widetilde{u}} \right|\!\right|\!\right|
\end{aligned}$$

Here (i) follows from $\lambda_{\min}(\widetilde{\Gamma}_{\widetilde{u}}) \, |\!|\!| \mathcal{I} |\!|\!| \leq \left|\!\left|\!\left| \widetilde{\Gamma}_{\widetilde{u}} \right|\!\right|\!\right|$ and the Loewner order lower bound for $\widetilde{\Gamma}_{\widetilde{u}}$ in part (iii) of this lemma; (ii) follows from the upper bound for $\|\widetilde{\Gamma}\|_{\mathrm{op}}$ from Lemma 17(ii) and Remark 10.

(v) (Independence case) When $\widetilde{Y} \perp \widetilde{Z}$, we have

$$\mathfrak{C} = \Lambda_* \otimes \Lambda_{*,\perp}, \qquad \widetilde{\mathfrak{C}} = I_r \otimes I_{d-r}.$$

or equivalently,

$$\left\langle E_{l'j'},\, \mathfrak{C} E_{lj} \right\rangle_{\mathrm{F}} = \lambda_j \lambda_{r+l} \delta_{ll'} \delta_{jj'}, \qquad \left\langle E_{l'j'},\, \widetilde{\mathfrak{C}} E_{lj} \right\rangle_{\mathrm{F}} = \delta_{ll'} \delta_{jj'}$$

which, combined with (i) coordinate representation, gives the desired results.

F.3.3 Proof of Lemma 19

See [19, Corollary 3]. Without loss of generality, assume G is diagonal with $G_{11} \geq \cdots \geq G_{dd} \geq 0$. Then

$$\mathbb{E}\left(\mathfrak{g}^\top G \mathfrak{g}\right)^p = \mathbb{E}\Big(\sum_{i=1}^d G_{ii} \mathfrak{g}_i^2\Big)^p \lesssim_p \left[\Big(\sum_{i=1}^d G_{ii}\Big)^p + \sum_{i=1}^d G_{ii}^p \mathbb{E} \mathfrak{g}_i^{2p}\right] \lesssim_p [\operatorname{tr} G]^p$$

Here the last inequality follows since $\sum_i G_{ii}^p \leq (\sum_i G_{ii})^p$ for $p \geq 1$ and $G \succeq 0$.

F.3.4 Proof of Lemma 24

Let

$$F_{\mathsf{A}_n}(t) := \mathbb{P}\big\{\|\mathsf{A}_n\|^2 \le t\big\}, \qquad F_{\mathsf{B}_n}(t) := \mathbb{P}\big\{\|\mathsf{B}_n\|^2 \le t\big\}, \qquad F_{\mathsf{G}_n}(t) := \mathbb{P}\big\{\|\mathsf{G}_n\|^2 \le t\big\}$$

Since $\|\mathsf{A}_n\|, \|\mathsf{B}_n\|, \|\mathsf{G}_n\| \ge 0$, it suffices to consider $t \ge 0$.

Define good event

$$\mathcal{G}_n := \{\|\mathsf{E}_n\| \le e_n, \|\mathsf{A}_n\| \le L_n\}.$$

On $\mathcal{G}_n$, we have

$$\big|\|\mathsf{B}_n\|^2 - \|\mathsf{A}_n\|^2\big| = \big|2\langle \mathsf{A}_n, \mathsf{E}_n\rangle + \|\mathsf{E}_n\|^2\big| \le 2L_n e_n + e_n^2 =: h_n$$

which implies

$$\big\{\|\mathsf{A}_n\|^2 \le t - h_n\big\} \cap \mathcal{G}_n \subseteq \big\{\|\mathsf{B}_n\|^2 \le t\big\} \subseteq \big\{\|\mathsf{A}_n\|^2 \le t + h_n\big\} \cup \mathcal{G}_n^c.$$

Thus we have

$$F_{\mathsf{A}_n}(t - h_n) - \mathbb{P}(\mathcal{G}_n^c) \le F_{\mathsf{B}_n}(t) \le F_{\mathsf{A}_n}(t + h_n) + \mathbb{P}(\mathcal{G}_n^c)$$

Subtracting $F_{\mathsf{G}_n}(t)$ and using $|F_{\mathsf{G}_n}(t) - F_{\mathsf{A}_n}(t)| \le \delta_n$ gives

$$\begin{aligned} F_{\mathsf{G}_n}(t - h_n) - F_{\mathsf{G}_n}(t) - \mathbb{P}(\mathcal{G}_n^c) - \delta_n &\le F_{\mathsf{B}_n}(t) - F_{\mathsf{G}_n}(t) \\ &\le F_{\mathsf{G}_n}(t + h_n) - F_{\mathsf{G}_n}(t) + \mathbb{P}(\mathcal{G}_n^c) + \delta_n \end{aligned}$$

which implies

$$|F_{\mathsf{B}_n}(t) - F_{\mathsf{G}_n}(t)| \le w_n(h_n) + \mathbb{P}(\mathcal{G}_n^c) + \delta_n \tag{F.6}$$

Now by the definition of good event $\mathcal{G}_n$ and δ_n, one can obtain

$$\mathbb{P}(\mathcal{G}_n^c) \le \mathbb{P}\{\|\mathsf{E}_n\| > e_n\} + \mathbb{P}\{\|\mathsf{A}_n\| > L_n\} \le \mathbb{P}\{\|\mathsf{E}_n\| > e_n\} + \mathbb{P}\{\|\mathsf{G}_n\| > L_n\} + \delta_n$$

Plugging the above bound for $\mathbb{P}(\mathcal{G}_n^c)$ into (F.6) gives

$$|F_{\mathsf{B}_n}(t) - F_{\mathsf{G}_n}(t)| \le w_n(h_n) + 2\delta_n + \mathbb{P}\{\|\mathsf{E}_n\| > e_n\} + \mathbb{P}\{\|\mathsf{G}_n\| > L_n\}$$

Taking supremum over $t \in \mathbb{R}$ gives the desired result.

F.3.5 Proof of Lemma 25

Let

$$F_{A_n}(t) := \mathbb{P}\big\{\|A_n\|^2 \le t\big\}, \qquad F_{B_n}(t) := \mathbb{P}\big\{\|B_n\|^2 \le t\big\}, \qquad F_{G_n}(t) := \mathbb{P}\big\{\|G_n\|^2 \le t\big\}.$$

Since $\|A_n\|, \|B_n\|, \|G_n\| \ge 0$, it suffices to consider $t \ge 0$.

Define events

$$\mathcal{E}_n := \{\|\mathsf{E}_n\| \le e_n\}, \qquad \mathcal{G}_n := \{\|\mathsf{G}_n\| \le L_n\}.$$

On $\mathcal{E}_n$, we have

$$(1 - e_n)\|x\| \le \|(I_{d_n} + \mathsf{E}_n)x\| \le (1 + e_n)\|x\|, \quad \forall x \in \mathbb{R}^{d_n}$$

Since $0 \le e_n \le \frac{1}{2} \le 1$, by the definition of B_n, we then have

$$(1 - e_n)^2 \|\mathsf{A}_n\|^2 \le \|\mathsf{B}_n\|^2 \le (1 + e_n)^2 \|\mathsf{A}_n\|^2$$

So for any $t \ge 0$, we have

$$\big\{\|\mathsf{B}_n\|^2 \le t\big\} \cap \mathcal{E}_n \subseteq \big\{\|\mathsf{A}_n\|^2 \le \frac{t}{(1 - e_n)^2}\big\}$$

and

$$\big\{\|\mathsf{A}_n\|^2 \le \frac{t}{(1 + e_n)^2}\big\} \cap \mathcal{E}_n \subseteq \big\{\|\mathsf{B}_n\|^2 \le t\big\}.$$

Taking probabilities yields

$$F_{\mathsf{A}_n}\left(\frac{t}{(1+e_n)^2}\right) - \mathbb{P}(\mathcal{E}_n^c) \le F_{\mathsf{B}_n}(t) \le F_{\mathsf{A}_n}\left(\frac{t}{(1-e_n)^2}\right) + \mathbb{P}(\mathcal{E}_n^c)$$

which together with the definition of δ_n gives

$$F_{\mathsf{G}_n}\left(\frac{t}{(1+e_n)^2}\right) - \mathbb{P}(\mathcal{E}_n^c) - \delta_n \le F_{\mathsf{B}_n}(t) \le F_{\mathsf{G}_n}\left(\frac{t}{(1-e_n)^2}\right) + \mathbb{P}(\mathcal{E}_n^c) + \delta_n$$

Subtracting $F_{\mathsf{G}_n}(t)$ from the above inequalities gives

$$F_{\mathsf{B}_n}(t) - F_{\mathsf{G}_n}(t) \le \underbrace{F_{\mathsf{G}_n}\left(\frac{t}{(1-e_n)^2}\right) - F_{\mathsf{G}_n}(t)}_{=:\alpha_1} + \mathbb{P}(\mathcal{E}_n^c) + \delta_n$$

$$F_{\mathsf{G}_n}(t) - F_{\mathsf{B}_n}(t) \le \underbrace{F_{\mathsf{G}_n}(t) - F_{\mathsf{G}_n}\left(\frac{t}{(1+e_n)^2}\right)}_{=:\alpha_2} + \mathbb{P}(\mathcal{E}_n^c) + \delta_n$$

So it remains to bound α_1 and α_2.

(i) (α_1)

$$\alpha_1 = \mathbb{P}\Big\{t < \|\mathsf{G}_n\|^2 \le \frac{t}{(1-e_n)^2}\Big\}$$

On the event $\big\{t < \|\mathsf{G}_n\|^2 \le t/(1-e_n)^2\big\} \cap \mathcal{G}_n$ we have $t \le \|\mathsf{G}_n\|^2 \le L_n^2$, hence

$$\frac{t}{(1-e_n)^2} - t \le \left(\frac{1}{(1-e_n)^2} - 1\right) L_n^2 = h_n$$

Therefore, we have

$$\alpha_1 = \mathbb{P}\Big\{t < \|\mathsf{G}_n\|^2 \le \frac{t}{(1-e_n)^2}\Big\} \le w_n(h_n) + \mathbb{P}(\mathcal{G}_n^c).$$

(ii) (α_2) By definition,

$$\alpha_2 = \mathbb{P}\left\{\frac{t}{(1+e_n)^2} < \|\mathsf{G}_n\|^2 \le t\right\}.$$

Set $t_\wedge := t \wedge L_n^2$. On $\mathcal{G}_n$, the event above forces $\|\mathsf{G}_n\|^2 \le t_\wedge$. Hence

$$\left\{\frac{t}{(1+e_n)^2} < \|\mathsf{G}_n\|^2 \le t\right\} \cap \mathcal{G}_n \subseteq \left\{\frac{t}{(1+e_n)^2} < \|\mathsf{G}_n\|^2 \le t_\wedge\right\}.$$

The width of the latter interval satisfies

$$t_\wedge - \frac{t}{(1+e_n)^2} \le t_\wedge - \frac{t_\wedge}{(1+e_n)^2} = t_\wedge\left(1 - \frac{1}{(1+e_n)^2}\right) \le L_n^2\left(1 - \frac{1}{(1+e_n)^2}\right) \le h_n,$$

where the first inequality uses $t_\wedge \le t$, and the last inequality follows from

$$1 - \frac{1}{(1+e_n)^2} \le \frac{1}{(1-e_n)^2} - 1.$$

Therefore,

$$\alpha_2 \le w_n(h_n) + \mathbb{P}(\mathcal{G}_n^c).$$

Combining the preceding bounds gives

$$\sup_{t\in\mathbb{R}} |F_{\mathsf{B}_n}(t) - F_{\mathsf{G}_n}(t)| \le \delta_n + \mathbb{P}\left\{\|\mathsf{E}_n\| > e_n\right\} + w_n(h_n) + \mathbb{P}\left\{\|\mathsf{G}_n\| > L_n\right\}$$

G Proofs for Section 3.2

G.1 Proof of Theorem 5

The proof proceeds in three steps, progressing right-to-left through the chain

$$\eta^{-1}\left\|\sin\Theta(U_n,U_*)\right\|_{\mathrm{F}}^2 \overset{\text{Step 3.}}{\approx} \eta^{-1}\left\|\mathcal{T}_n\right\|_{\mathrm{F}}^2 \overset{\text{Step 2.}}{\approx} \left\|\sum_{i=1}^{n}\xi_i\right\|_{\mathrm{F}}^2 \overset{\text{Step 1.}}{\approx} \|\mathsf{G}\|_{\mathrm{F}}^2 .$$

Preliminaries. For convenience, recall notation

$$d_{\mathfrak{C}} = \frac{\lambda_\times^4}{\|\mathfrak{C}\|_{\mathrm{HS}}^2} = \frac{\lambda_{1\sim r}^2\lambda_{r+1\sim d}^2}{\|\mathfrak{C}\|_{\mathrm{HS}}^2}, \quad \kappa_* = \frac{\lambda_1}{\Delta_{\min}}, \quad \kappa_\Delta = \frac{\Delta_{\max}}{\Delta_{\min}}.$$

We restate the conditions in Theorem 5 here:

$$n^{-2\varepsilon} \geq C_1\widetilde{\eta}\kappa_*^2 r^3(1+\nu_\infty d)\,(\log n)^3\frac{\log(1/\delta_p)}{\delta_p^2} \tag{11}$$

$$\widetilde{\eta}\,\kappa_*^4 r^3 d_{\mathfrak{C}}^{3/2}(\log n)^3 \to 0 \tag{15(i)}$$

$$\widetilde{\eta}\kappa_*^4 r^3 d_{\mathfrak{C}}\bar{\nu}^2(\log n)^3 \to 0, \tag{15(ii)}$$

By Lemma 38, we have

$$\|\mathfrak{C}\|_{\mathrm{HS}} \leq \operatorname{tr}\mathfrak{C} = \mathbb{E}\left\|ZY^\top\right\|_{\mathrm{F}}^2 \lesssim \lambda_{1\sim r}\lambda_{r+1\sim d}$$

Thus, $d_{\mathfrak{C}} \gtrsim 1$. Since $\lambda_1 \geq \Delta_{\max}$, we have $\kappa_* \geq \kappa_\Delta \geq 1$. Thus (11) implies

$$\eta\Delta_{\max} = \widetilde{\eta}\kappa_\Delta \leq \widetilde{\eta}\kappa_* < 1, \quad \text{and} \quad \widetilde{\eta}\kappa_\Delta^2 \to 0.$$

Note that $\widetilde{\eta} = c_\eta\frac{\log n}{n}$ with $c_\eta \geq 1$ implies that $(1-\widetilde{\eta})^{2n} \lesssim \widetilde{\eta}$, thus $\epsilon_n = (1-\widetilde{\eta})^{2n} + \widetilde{\eta}\kappa_\Delta^2 \to 0$, and hence $< \frac{1}{2}$ for n large enough.

Step 1: Gaussian approximation of $\|\sum_{i=1}^n \xi_i\|_{\mathrm{F}}^2$. By the triangle inequality for the Kolmogorov distance d_K,

$$d_K\Big(\Big\|\sum_{i=1}^n\xi_i\Big\|_{\mathrm{F}}^2, \|\mathsf{G}\|_{\mathrm{F}}^2\Big) \;\leq\; d_K\Big(\Big\|\sum_{i=1}^n\xi_i\Big\|_{\mathrm{F}}^2,\; \Big\|\sum_{i=1}^n g_i\Big\|_{\mathrm{F}}^2\Big) + d_K\Big(\Big\|\sum_{i=1}^n g_i\Big\|_{\mathrm{F}}^2,\; \|\mathsf{G}\|_{\mathrm{F}}^2\Big) \tag{G.1}$$

First term in (G.1)**.**

Lemma 26. *Assume Assumption 1 holds. Let L_q^ξ, $L_q^{\mathfrak{g}}$, K_β, and J_n be defined as in Lemma 23 for $q \in (2,3]$ and $\beta \geq 2$. If*

$$\eta\Delta_{\max} \in (0,1), \quad \epsilon_n := (1-\widetilde{\eta})^{2n} + \widetilde{\eta}\kappa_\Delta^2 \leq \frac{1}{2}, \quad \|\mathfrak{C}\|_{\mathrm{HS}} > 0,$$

then

$$\begin{aligned} L_q^\xi &\lesssim \widetilde{\eta}^{q/2-1}\kappa_\Delta^{q/2}d_C^{q/4} + \widetilde{\eta}^{q-2}\kappa_\Delta^q d_C^{q/2}, \\ L_q^{\mathfrak{g}} &\lesssim \widetilde{\eta}^{q/2-1}\kappa_\Delta^{q/2}, \\ n^{1-\beta}K_\beta^\beta &\lesssim \widetilde{\eta}^{\beta-1}\kappa_\Delta^\beta d_{\mathfrak{C}}^{\beta/2}, \\ J_n &\lesssim \widetilde{\eta}\kappa_\Delta^2 d_{\mathfrak{C}}. \end{aligned}$$

In particular, for $q = \beta = 3$, a sufficient condition for all four terms to vanish is

$$\widetilde{\eta}\kappa_\Delta^3 d_{\mathfrak{C}}^{3/2} \to 0. \tag{G.2}$$

Proof. See Appendix G.3.1. □

Since $\kappa_* \geq \kappa_\Delta \geq 1$, Condition (15)(i) implies that (G.2) holds, thus the first term in (G.1) goes to zero.

Second term in (G.1)**.**

Lemma 27. *Assume Assumption 1 holds and* $\|\mathfrak{C}\|_{\mathrm{HS}} > 0$. *If* $\eta\Delta_{\max} \in (0,1)$, *then*

$$d_K\Big(\Big\|\sum_{i=1}^n g_i\Big\|_{\mathrm{F}}^2,\ \|\mathsf{G}\|_{\mathrm{F}}^2\Big) \lesssim \left[\epsilon_n(\log n)\kappa_\Delta \frac{C_\psi^4\lambda_{1\sim r}^2\nu_2^2}{\|\mathfrak{C}\|_{\mathrm{HS}}}\right]^{1/2} + n^{-20},$$

where $\epsilon_n := (1-\widetilde{\eta})^{2n} + \widetilde{\eta}\kappa_\Delta^2$.

Proof. See Appendix G.3.2. □

Note that for $\widetilde{\eta} = n^{-1}c_\eta \log n$ with $c_\eta \geq \frac{1}{2}$, we have $(1-\widetilde{\eta})^{2n} \leq e^{-\log n} = n^{-1} \lesssim \widetilde{\eta}$. Thus, in order for $d_K\Big(\Big\|\sum_{i=1}^n g_i\Big\|_{\mathrm{F}}^2,\ \|\mathsf{G}\|_{\mathrm{F}}^2\Big) \to 0$, it suffices that

$$\widetilde{\eta}\kappa_\Delta^3(\log n)\frac{\lambda_\times^2}{\|\mathfrak{C}\|_{\mathrm{HS}}} \to 0.$$

which is implied by (15)(i), $\widetilde{\eta} \leq 1$ and $d_{\mathfrak{C}} \gtrsim 1$.

Step 2: Gaussian approximation of $\eta^{-1}\|\mathcal{T}_n\|_{\mathrm{F}}^2$**.**

Lemma 28. *Suppose the assumptions of Theorem 5 hold. Then*

$$d_K\bigg(\Big\|\frac{1}{\sqrt{\eta}}\mathcal{T}_n\Big\|_{\mathrm{F}}^2,\ \|\mathsf{G}\|_{\mathrm{F}}^2\bigg) \lesssim d_K\Big(\Big\|\sum_{i=1}^n \xi_i\Big\|_{\mathrm{F}}^2,\ \|\mathsf{G}\|_{\mathrm{F}}^2\Big) + \sqrt{\frac{e_n(e_n+L_n)}{\|\Gamma\|_{\mathrm{HS}}}} + \delta_p,$$

where $e_n = C_e\Delta_{\min}^{1/2}\kappa_*^2 r^{3/2}\sqrt{\widetilde{\eta}}\ (\log n)\bar{\nu}\,[1+\bar{\nu}]$, $L_n = [C_L \operatorname{tr}(\Gamma)\log n]^{1/2}$ *for some constants* $C_e, C_L > 0$ *large enough.*

Proof. See Appendix G.3.3. □

We already have $d_K\Big(\Big\|\sum_{i=1}^n \xi_i\Big\|_{\mathrm{F}}^2,\ \|\mathsf{G}\|_{\mathrm{F}}^2\Big) \to 0$ from Step 1. It suffices to show $\frac{e_n(e_n+L_n)}{\|\Gamma\|_{\mathrm{HS}}} \to 0$, or equivalently,

$$\frac{e_n^2}{\|\Gamma\|_{\mathrm{HS}}} \asymp \kappa_*^4 r^3\widetilde{\eta}\ (\log n)^2\bar{\nu}^2\,[1+\bar{\nu}]^2\,\frac{\Delta_{\min}}{\|\Gamma\|_{\mathrm{HS}}} \to 0,$$
$$\frac{e_n L_n}{\|\Gamma\|_{\mathrm{HS}}} \asymp \kappa_*^2 r^{3/2}\sqrt{\widetilde{\eta}}\ (\log n)^{3/2}\bar{\nu}\,[1+\bar{\nu}]\,\frac{\Delta_{\min}^{1/2}\sqrt{\operatorname{tr}\Gamma}}{\|\Gamma\|_{\mathrm{HS}}} \to 0.$$

By Lemma 17(ii) and (vi), we have $\|\Gamma\|_{\mathrm{HS}} \gtrsim \frac{\|\mathfrak{C}\|_{\mathrm{HS}}}{\Delta_{\max}}$ and

$$\operatorname{tr}\Gamma \lesssim \frac{\operatorname{tr}\mathfrak{C}}{\Delta_{\min}} = \frac{\mathbb{E}\left\|ZY^\top\right\|_{\mathrm{F}}^2}{\Delta_{\min}} \lesssim \frac{\lambda_{1\sim r}\lambda_{r+1\sim d}}{\Delta_{\min}}$$

Thus, it suffices that

$$\Phi_{2,1} := \kappa_*^4 r^3\widetilde{\eta}\ (\log n)^2\bar{\nu}^2\,[1+\bar{\nu}]^2\,\frac{\Delta_{\min}\Delta_{\max}}{\|\mathfrak{C}\|_{\mathrm{HS}}} \to 0$$
$$\Phi_{2,2} := \kappa_*^2 r^{3/2}\sqrt{\widetilde{\eta}}\ (\log n)^{3/2}\bar{\nu}\,[1+\bar{\nu}]\,\frac{\Delta_{\max}\lambda_{1\sim r}\nu_2}{\|\mathfrak{C}\|_{\mathrm{HS}}} \to 0$$

We now show $\Phi_{2,1}, \Phi_{2,2} \to 0$.

(i) $\Phi_{2,1}$. First, note that $\lambda_1^2\bar\nu^2 \le \max(\lambda_1^2\nu_\infty^2, \lambda_1^2\nu_2^2) \le \lambda_\times^2$, thus using $\Delta_{\min} \le \Delta_{\max} \le \lambda_1$ and $(1+\bar\nu)^2 \lesssim 1+\bar\nu^2$ gives

$$\begin{aligned}\Phi_{2,1} &\le \kappa_*^4 r^2 \widetilde\eta (\log n)^2 \left[1+\bar\nu\right]^2 \frac{\lambda_\times^2}{\|\mathfrak{C}\|_{\mathrm{HS}}} \\ &\lesssim \kappa_*^4 r^2 \widetilde\eta (\log n)^2 d_{\mathfrak{C}}^{1/2} + \kappa_*^4 r^2 \widetilde\eta (\log n)^2 \bar\nu^2 d_{\mathfrak{C}}^{1/2}.\end{aligned}$$

where (15)(i) and (15)(ii) ensure each term $\to 0$ on the last line.

(ii) $\Phi_{2,2}$. Using $\Delta_{\max}\bar\nu \le \lambda_\times$, we have

$$\begin{aligned}\Phi_{2,2} &\le \kappa_*^2 r^{3/2} \sqrt{\widetilde\eta}\,(\log n)^{3/2} \left[1+\bar\nu\right] d_{\mathfrak{C}}^{1/2} \\ &= \kappa_*^2 r^{3/2} \sqrt{\widetilde\eta}\,(\log n)^{3/2} d_{\mathfrak{C}}^{1/2} + \kappa_*^2 r^{3/2} \sqrt{\widetilde\eta}\,(\log n)^{3/2} \bar\nu d_{\mathfrak{C}}^{1/2}\end{aligned}$$

where (15)(i) and (15)(ii) ensure each term $\to 0$ on the last line.

Step 3: Gaussian approximation of $\eta^{-1}\left\|\sin\Theta(U_n, U_*)\right\|_{\mathrm{F}}^2$.

Lemma 29. *Suppose the assumptions of Theorem 5 hold. Then*

$$d_K\Big(\Big\|\frac{1}{\sqrt{\eta}}\sin\Theta(U_n,U_*)\Big\|_{\mathrm{F}}^2,\ \|\mathsf{G}\|_{\mathrm{F}}^2\Big) \lesssim d_K\Big(\Big\|\frac{1}{\sqrt{\eta}}\mathcal{T}_n\Big\|_{\mathrm{F}}^2,\ \|\mathsf{G}\|_{\mathrm{F}}^2\Big) + \tau_*\sqrt{\frac{L_n^2}{\|\Gamma\|_{\mathrm{HS}}}} + \delta_p,$$

where $L_n = \sqrt{C_L\left[\operatorname{tr}\Gamma\right]\log n}$ for some constant $C_L > 0$ large enough.

Proof. See Appendix G.3.4. □

We have $d_K\big(\|\frac{1}{\sqrt{\eta}}\mathcal{T}_n\|_{\mathrm{F}}^2,\ \|\mathsf{G}\|_{\mathrm{F}}^2\big) \to 0$ from step 2. It suffices to show $\tau_*^2\operatorname{tr}(\Gamma)(\log n)/\|\Gamma\|_{\mathrm{HS}} \to 0$.

Recall $\tau_* \asymp \kappa_*\bar\nu\sqrt{\widetilde\eta r\log n}$. By Lemma 17(ii) and (vi), we have $\|\Gamma\|_{\mathrm{HS}} \gtrsim \frac{\|\mathfrak{C}\|_{\mathrm{HS}}}{\Delta_{\max}}$ and

$$\operatorname{tr}\Gamma \lesssim \frac{\operatorname{tr}\mathfrak{C}}{\Delta_{\min}} = \frac{\mathbb{E}\left\|ZY^\top\right\|_{\mathrm{F}}^2}{\Delta_{\min}} \lesssim \frac{\lambda_{1\sim r}\lambda_{r+1\sim d}}{\Delta_{\min}}$$

Thus, it suffices that

$$\kappa_*^2\bar\nu^2\widetilde\eta r(\log n)^2\cdot\kappa_\Delta\frac{\lambda_\times^2}{\|\mathfrak{C}\|_{\mathrm{HS}}} \to 0.$$

which is implied by (15)(ii).

G.2 Proof of Theorem 6

For notation simplicity, we drop the subscript n and write U_n as U, $\mathcal{C}_n$ as $\mathcal{C}$, and $\mathcal{T}_n$ as $\mathcal{T}$. Recall the definition of ξ_i from (F.2)

$$\xi_i := \sqrt{\eta}(\mathcal{I}-\eta\mathcal{L})^{n-i}(Z_iY_i^\top) \in \mathbb{R}^{(d-r)\times r}.$$

Denote

$$E = \mathcal{T} - \sqrt{\eta}\sum_{i=1}^n \xi_i \in \mathbb{R}^{(d-r)\times r}.$$

Then Theorem 3 implies that

$$\|E\| \lesssim \frac{\tau_*}{n^\varepsilon\log n}.$$

By Lemma 39, we can obtain the decomposition

$$U\operatorname{sgn}(U^\top U_*) - U_* = \mathcal{W} + \Psi$$

where the main term $\mathcal{W}$ and remainder Ψ are defined as

$$\mathcal{W} := U_{*,\perp}\sqrt{\eta}\sum_{i=1}^{n}\xi_i \in \mathbb{R}^{d\times r},$$
$$\Psi := U_*\big[(\mathcal{C}\mathcal{C}^\top)^{1/2} - I_r\big] + U_{*,\perp}\mathcal{T}\big[(\mathcal{C}\mathcal{C}^\top)^{1/2} - I_r\big] + U_{*,\perp}E \in \mathbb{R}^{d\times r}.$$

and

$$\left\|(\mathcal{C}\mathcal{C}^\top)^{1/2} - I_r\right\| \lesssim \|\mathcal{T}\|^2 \lesssim \tau_*^2 \lesssim \kappa_*^2 r\widetilde{\eta}(\log n)\bar{\nu}^2. \tag{G.3}$$

Fix $m \in [d]$ and define

$$u_m := U_{*,\perp}^\top e_m \in \mathbb{R}^{d-r}, \qquad \widetilde{u}_m := \Lambda_{*,\perp}^{1/2}u_m, \qquad \Sigma_{U,m}^* := \eta\Gamma_{u_m}.$$

where Γ_{u_m} and $\widetilde{\Gamma}_{\widetilde{u}_m}$ are defined as in (F.3) and (F.4), respectively.

Let e_m denote the m-th standard basis vector in $\mathbb{R}^d$. Then the m-th row of $U\operatorname{sgn}(U^\top U_*) - U_*$, identified as a column vector via

$$\big[U\operatorname{sgn}(U^\top U_*) - U_*\big]_{m,\cdot} \leftrightarrow \big[U\operatorname{sgn}(U^\top U_*) - U_*\big]^\top e_m \in \mathbb{R}^r,$$

satisfies

$$\begin{aligned}
&\big[U\operatorname{sgn}(U^\top U_*) - U_*\big]^\top e_m = \mathcal{W}^\top e_m + \Psi^\top e_m \\
\text{where}\qquad &\mathcal{W}^\top e_m = \sqrt{\eta}\sum_{i=1}^{n}\xi_i^\top U_{*,\perp}^\top e_m = \sqrt{\eta}\sum_{i=1}^{n}\xi_i^\top u_m =: \mathcal{W}_{u_m}, \\
&\Psi^\top e_m = \big[(\mathcal{C}\mathcal{C}^\top)^{1/2} - I_r\big]U_*^\top e_m + \big[(\mathcal{C}\mathcal{C}^\top)^{1/2} - I_r\big]\mathcal{T}^\top u_m + E^\top u_m =: \Psi_{u_m}.
\end{aligned} \tag{G.4}$$

Step 1: Gaussian approximation of $\mathcal{W}_{u_m}$.

Lemma 30. *Suppose $\widetilde{u}_{u_m} = \Lambda_{*,\perp}^{1/2}u_{u_m} \neq 0$, $\lambda_{\min}(\widetilde{\mathfrak{C}}) > 0$, $n\widetilde{\eta} \geq \frac{\log n}{2}$. Assume*

$$\widetilde{\eta}^{1/2}\kappa_*^3 r^{7/4}\lambda_{\min}^{-3/2}(\widetilde{\mathfrak{C}}) \to 0 \tag{G.5}$$

Then, with $\Sigma_{U,m}^* = \eta\Gamma_{u_m}$, we have

$$\sup_{\mathcal{A}\in\mathscr{A}^r}\left|\mathbb{P}\left\{\mathcal{W}_{u_m} \in \mathcal{A}\right\} - \mathcal{N}(0, \Sigma_{U,m}^*)\left\{\mathcal{A}\right\}\right| = o(1).$$

Proof. See Appendix G.3.5. $\square$

Step 2: Accounting for higher-order term Ψ_{u_m}.

Lemma 31. *Assume conditions of Theorem 6 hold. Then*

$$\mathbb{P}\left\{\left\|\Sigma_{U,m}^{*-1/2}\Psi_{u_m}\right\| \leq \zeta\right\} \geq 1 - O(\delta_p) \tag{G.6}$$

where ζ is defined as

$$\begin{aligned}
\zeta &= C_\zeta\left(\zeta_1 + \zeta_2\right), \\
\zeta_1 &= \kappa_*^2\widetilde{\eta}^{1/2}r(\log n)\bar{\nu}^2\frac{\lambda_1^{1/2}\left\|U_*^\top e_m\right\|}{\|\widetilde{u}_m\|}\lambda_{\min}^{-1/2}(\widetilde{\mathfrak{C}}), \\
\zeta_2 &= \kappa_*^2\widetilde{\eta}^{1/2}r(\log n)\bar{\nu}\,(1+\bar{\nu})\frac{\lambda_1^{1/2}\left\|U_{*,\perp}^\top e_m\right\|}{\|\widetilde{u}_m\|}\lambda_{\min}^{-1/2}(\widetilde{\mathfrak{C}}).
\end{aligned} \tag{G.7}$$

for some sufficiently large constant $C_\zeta > 0$. Moreover, $r^{1/4}\zeta \to 0$.

Proof. See Appendix G.3.6. $\square$

Let ζ be defined as in (G.7). Recall the definition of $\mathcal{A}^t$ in (I.6) from Appendix I for any convex set $\mathcal{A} \in \mathscr{A}^r$ and any t. We have

$$\begin{aligned}
\mathbb{P}\left\{\Sigma_{U,m}^{*-1/2}\mathcal{W}_{u_m} \in \mathcal{A}^{-\zeta}\right\} &= \mathbb{P}\left\{\Sigma_{U,m}^{*-1/2}\mathcal{W}_{u_m} \in \mathcal{A}^{-\zeta}, \left\|\Sigma_{U,m}^{*-1/2}\Psi_{u_m}\right\| \le \zeta\right\} \\
&\quad + \mathbb{P}\left\{\Sigma_{U,m}^{*-1/2}\mathcal{W}_{u} \in \mathcal{A}^{-\zeta}, \left\|\Sigma_{U,m}^{*-1/2}\Psi_{u}\right\| > \zeta\right\} \\
&\le \mathbb{P}\left\{\Sigma_{U,m}^{*-1/2}(\mathcal{W}_{u}+\Psi_u) \in \mathcal{A}\right\} + \mathbb{P}\left\{\left\|\Sigma_{U,m}^{*-1/2}\Psi_{u}\right\| > \zeta\right\} \\
&\le \mathbb{P}\left\{\Sigma_{U,m}^{*-1/2}\left[U\operatorname{sgn}(U^\top U_*) - U_*\right]^\top e_m \in \mathcal{A}\right\} + O(\delta_p) \qquad \text{(G.8)}
\end{aligned}$$

where the last line follows from the decomposition in (G.4) and the bound in (G.6).

Similarly,

$$\mathbb{P}\left\{\Sigma_{U,m}^{*-1/2}\left[U\operatorname{sgn}(U^\top U_*) - U_*\right]^\top e_m \in \mathcal{A}\right\} \le \mathbb{P}\left\{\Sigma_{U,m}^{*-1/2}\mathcal{W}_{u} \in \mathcal{A}^{\zeta}\right\} + O(\delta_p). \qquad \text{(G.9)}$$

Moreover, we can obtain

$$\begin{aligned}
\mathbb{P}\left\{\Sigma_{U,m}^{*-1/2}\mathcal{W}_{u_m} \in \mathcal{A}^{\zeta}\right\} &= \mathbb{P}\left\{\mathcal{W}_{u_m} \in \Sigma_{U,m}^{*1/2}\mathcal{A}^{\zeta}\right\} \\
&\overset{(i)}{\le} \mathbb{P}\left\{\mathcal{N}(0,\Sigma_{U,m}^*) \in \Sigma_{U,m}^{*1/2}\mathcal{A}^{\zeta}\right\} + o(1) \\
&= \mathbb{P}\left\{\mathcal{N}(0,I_r) \in \mathcal{A}^{\zeta}\right\} + o(1) \\
&\overset{(ii)}{\le} \mathbb{P}\left\{\mathcal{N}(0,I_r) \in \mathcal{A}\right\} + \zeta(0,59r^{1/4}+0.21) + o(1) \\
&\overset{(iii)}{=} \mathbb{P}\left\{\mathcal{N}(0,I_r) \in \mathcal{A}\right\} + o(1) \qquad \text{(G.10)}
\end{aligned}$$

where (i) follows from Lemma 30; (ii) is due to Theorem 14 and (iii) is a result of $r^{1/4}\zeta \to 0$ from Lemma 31. Similarly, we have

$$\mathbb{P}\left\{\Sigma_{U,m}^{*-1/2}\mathcal{W}_{u_m} \in \mathcal{A}^{-\zeta}\right\} \ge \mathbb{P}\left\{\mathcal{N}(0,I_r) \in \mathcal{A}\right\} - o(1). \qquad \text{(G.11)}$$

Finally, combining (G.8), (G.9), (G.10) and (G.11) gives

$$\left|\mathbb{P}\left\{\mathcal{N}(0,I_r) \in \mathcal{A}\right\} - \mathbb{P}\left\{\Sigma_{U,m}^{*-1/2}\left[U\operatorname{sgn}(U^\top U_*) - U_*\right]^\top e_m \in \mathcal{A}\right\}\right| \to 0 \qquad \text{(G.12)}$$

Taking supremum over all convex set $\mathcal{A} \in \mathscr{A}^r$ gives

$$\begin{aligned}
&\sup_{\mathcal{A}\in\mathscr{A}^r}\left|\mathbb{P}\left\{\left[U\operatorname{sgn}(U^\top U_*) - U_*\right]^\top e_m \in \mathcal{A}\right\} - \mathcal{N}(0,\Sigma_{U,m}^*)\left\{\mathcal{A}\right\}\right| \\
&= \sup_{\mathcal{A}\in\mathscr{A}^r}\left|\mathbb{P}\left\{\left[U\operatorname{sgn}(U^\top U_*) - U_*\right]^\top e_m \in \Sigma_{U,m}^{*1/2}\mathcal{A}\right\} - \mathcal{N}(0,\Sigma_{U,m}^*)\left\{\Sigma_{U,m}^{*1/2}\mathcal{A}\right\}\right| \\
&= \sup_{\mathcal{A}\in\mathscr{A}^r}\left|\mathbb{P}\left\{\Sigma_{U,m}^{*-1/2}\left[U\operatorname{sgn}(U^\top U_*) - U_*\right]^\top e_m \in \mathcal{A}\right\} - \mathcal{N}(0,I_r)\left\{\mathcal{A}\right\}\right| \\
&\to 0.
\end{aligned}$$

where the last line follows from (G.12). This completes the proof of Theorem 6.

G.3 Proof of auxiliary lemmas

G.3.1 Proof of Lemma 26

Term L_q^ξ. Recall $L_q^\xi = L_{q,1}^\xi + L_{q,2}^\xi$ where

$$L_{q,1}^\xi := \sum_{i=1}^n \frac{\mathbb{E}\left\langle \xi_i,\, \Gamma_n \xi_i\right\rangle_{\mathrm{F}}^{q/2}}{\|\Gamma_n\|_{\mathrm{HS}}^q}$$

$$L_{q,2}^{\xi} := \frac{n^2}{\binom{n}{2}} \sum_{1\le i<j\le n} \frac{\mathbb{E}\left|\langle \xi_i,\, \xi_j\rangle_{\mathrm{F}}\right|^q}{\|\Gamma_n\|_{\mathrm{HS}}^q}.$$

Bound for $L_{q,1}^{\xi}$. Since $\Gamma_n \succeq 0$,

$$\langle \xi_i, \Gamma_n \xi_i\rangle_{\mathrm{F}} \le \|\Gamma_n\|_{\mathrm{op}} \|\xi_i\|_{\mathrm{F}}^2 \le \|\Gamma_n\|_{\mathrm{HS}} \|\xi_i\|_{\mathrm{F}}^2.$$

Therefore,

$$L_{q,1}^{\xi} \le \frac{\sum_{i=1}^n \mathbb{E}\|\xi_i\|_{\mathrm{F}}^q}{\|\Gamma_n\|_{\mathrm{HS}}^{q/2}}.$$

By Lemma 17, we have

$$\|\Gamma_n\|_{\mathrm{HS}} \gtrsim \frac{\|\mathfrak{C}\|_{\mathrm{HS}}}{\Delta_{\max}}, \qquad \sum_{i=1}^n \mathbb{E}\|\xi_i\|_{\mathrm{F}}^q \lesssim \frac{\widetilde{\eta}^{q/2-1}}{\Delta_{\min}^{q/2}} \lambda_\times^q.$$

Hence

$$L_{q,1}^{\xi} \lesssim \widetilde{\eta}^{q/2-1} \kappa_\Delta^{q/2} \frac{\lambda_\times^q}{\|\mathfrak{C}\|_{\mathrm{HS}}^{q/2}} = \widetilde{\eta}^{q/2-1} \kappa_\Delta^{q/2} d_{\mathfrak{C}}^{q/4}.$$

Bound for $L_{q,2}^{\xi}$. Note that for $i \ne j$,

$$\mathbb{E}\left|\langle \xi_i,\, \xi_j\rangle_{\mathrm{F}}\right|^q \le \eta^q (1-\widetilde{\eta})^{q(n-i)} (1-\widetilde{\eta})^{q(n-j)} \mathbb{E}\left\|ZY^\top\right\|_{\mathrm{F}}^{2q}$$

Since $\sum_{i<j} a_i a_j \le 2^{-1}(\sum_{i=1}^n a_i)^2$, we have

$$\sum_{1\le i<j\le n} \mathbb{E}\left|\langle \xi_i,\, \xi_j\rangle_{\mathrm{F}}\right|^q \le \frac{1}{2}\eta^q \left[\sum_{i=1}^n (1-\widetilde{\eta})^{q(n-i)}\right]^2 \cdot \mathbb{E}\left\|ZY^\top\right\|_{\mathrm{F}}^{2q} \lesssim \eta^q \widetilde{\eta}^{-2} \cdot \mathbb{E}\left\|ZY^\top\right\|_{\mathrm{F}}^{2q}$$

Therefore, applying Lemma 17 and Lemma 38 gives

$$L_{q,2}^{\xi} \lesssim \eta^q \widetilde{\eta}^{-2} \cdot \frac{\mathbb{E}\left\|ZY^\top\right\|_{\mathrm{F}}^{2q}}{\|\Gamma_n\|_{\mathrm{HS}}^q} \lesssim \widetilde{\eta}^{q-2} \kappa_\Delta^q \cdot \frac{\lambda_\times^{2q}}{\|\mathfrak{C}\|_{\mathrm{HS}}^q}$$

Combining two bounds gives

$$L_q^{\xi} \lesssim \widetilde{\eta}^{q/2-1} \kappa_\Delta^{q/2} d_C^{q/4} + \widetilde{\eta}^{q-2} \kappa_\Delta^q d_C^{q/2}.$$

For $q = 3$,

$$L_3^{\xi} \lesssim \widetilde{\eta}^{1/2} \kappa_\Delta^{3/2} d_C^{3/4} + \widetilde{\eta} \kappa_\Delta^3 d_C^{3/2}.$$

Term $L_q^{\mathfrak{g}}$. Recall

$$L_q^{\mathfrak{g}} = \sum_{i=1}^n \frac{\mathbb{E}\,\langle \mathfrak{g}_i,\, \Gamma_n \mathfrak{g}_i\rangle_{\mathrm{F}}^{q/2}}{\|\Gamma_n\|_{\mathrm{HS}}^q}$$

Write $\mathfrak{g}_i = \Xi_i^{1/2} \widetilde{\mathfrak{g}}_i$ where $\widetilde{\mathfrak{g}}_i \sim \mathcal{N}(\mathbf{0}_{d-r,r}, \mathcal{I})$, then $\langle \mathfrak{g}_i,\, \Gamma_n \mathfrak{g}_i\rangle_{\mathrm{F}} = \left\langle \widetilde{\mathfrak{g}}_i,\, \Xi_i^{1/2} \Gamma_n \Xi_i^{1/2} \widetilde{\mathfrak{g}}_i \right\rangle_{\mathrm{F}}$ and we have

$$\mathbb{E}\left|\langle \mathfrak{g}_i,\, \Gamma_n \mathfrak{g}_i\rangle_{\mathrm{F}}\right|^{q/2} \lesssim_q \left[\mathrm{tr}\left(\Xi_i^{1/2} \Gamma_n \Xi_i^{1/2}\right)\right]^{q/2} = \left[\mathrm{tr}\left(\Gamma_n \Xi_i\right)\right]^{q/2} \le \left[\|\Gamma_n\|_{\mathrm{HS}} \|\Xi_i\|_{\mathrm{HS}}\right]^{q/2}$$

Therefore,

$$L_q^{\mathfrak{g}} \lesssim \frac{1}{\|\Gamma_n\|_{\mathrm{HS}}^q} \sum_{i=1}^n \left[\|\Gamma_n\|_{\mathrm{HS}} \|\Xi_i\|_{\mathrm{HS}}\right]^{q/2} = \frac{1}{\|\Gamma_n\|_{\mathrm{HS}}^{q/2}} \sum_{i=1}^n \|\Xi_i\|_{\mathrm{HS}}^{q/2}$$

Note that

$$\|\Xi_i\|_{\mathrm{HS}} \le \eta \left\|(\mathcal{I}-\eta\mathcal{L})^{n-i}\mathfrak{C}(\mathcal{I}-\eta\mathcal{L})^{n-i}\right\|_{\mathrm{HS}} \le \eta(1-\eta\Delta_{\min})^{2(n-i)}\|\mathfrak{C}\|_{\mathrm{HS}}$$

Then

$$\sum_{i=1}^{n}\|\Xi_i\|_{\mathrm{HS}}^{q/2} \lesssim \eta^{q/2}\|\mathfrak{C}\|_{\mathrm{HS}}^{q/2}\sum_{i=1}^{n}(1-\eta\Delta_{\min})^{q(n-i)} \lesssim \frac{\eta^{q/2-1}}{\Delta_{\min}}\|\mathfrak{C}\|_{\mathrm{HS}}^{q/2}$$

Note that Lemma 17 gives $\|\Gamma_n\|_{\mathrm{HS}} \ge (1-\epsilon_n)\|\Gamma\|_{\mathrm{HS}} \gtrsim \frac{\|\mathfrak{C}\|_{\mathrm{HS}}}{\Delta_{\max}}$, thus

$$L_q^{\mathfrak{g}} \lesssim \frac{\widetilde{\eta}^{q/2-1}}{\Delta_{\min}^{q/2}}\cdot\frac{\|\mathfrak{C}\|_{\mathrm{HS}}^{q/2}}{\|\Gamma_n\|_{\mathrm{HS}}^{q/2}} \lesssim \widetilde{\eta}^{q/2-1}\kappa_\Delta^{q/2}.$$

Term K_β^β. Recall

$$K_\beta^\beta = \frac{n^\beta}{n}\sum_{i=1}^{n}\mathbb{E}\left|\frac{\langle\xi_i,\,\xi_i\rangle_{\mathrm{F}} - \mathbb{E}\langle\xi_i,\,\xi_i\rangle_{\mathrm{F}}}{\|\Gamma_n\|_{\mathrm{HS}}}\right|^\beta$$

Note that for $\beta \ge 2$,

$$\mathbb{E}\left|\langle\xi_i,\,\xi_i\rangle_{\mathrm{F}} - \mathbb{E}\langle\xi_i,\,\xi_i\rangle_{\mathrm{F}}\right|^\beta \lesssim \mathbb{E}\left|\langle\xi_i,\,\xi_i\rangle_{\mathrm{F}}\right|^\beta + \left|\mathbb{E}\langle\xi_i,\,\xi_i\rangle_{\mathrm{F}}\right|^\beta \lesssim \mathbb{E}\|\xi_i\|_{\mathrm{F}}^{2\beta}$$

Applying Lemma 17 and Lemma 38 gives

$$n^{1-\beta}K_\beta^\beta \lesssim \sum_{i=1}^{n}\frac{\mathbb{E}\|\xi_i\|_{\mathrm{F}}^{2\beta}}{\|\Gamma_n\|_{\mathrm{HS}}^\beta} \lesssim \widetilde{\eta}^{\beta-1}\kappa_\Delta^\beta\frac{\lambda_\times^{2\beta}}{\|\mathfrak{C}\|_{\mathrm{HS}}^\beta}$$

For $\beta = 3$, we have

$$n^{-2}K_3^3 \lesssim \widetilde{\eta}^2\kappa_\Delta^3\cdot\frac{\lambda_\times^6}{\|\mathfrak{C}\|_{\mathrm{HS}}^3}$$

Term J_n. Recall

$$J_n = \frac{\sum_{i=1}^n \mathfrak{v}_i}{\|\Gamma_n\|_{\mathrm{HS}}^2}, \qquad \text{where} \quad \mathfrak{v}_i = \mathrm{Var}(\|\xi_i\|_{\mathrm{F}}^2)$$

Note that $\mathfrak{v}_i \le \mathbb{E}\|\xi_i\|_{\mathrm{F}}^4$, then applying Lemma 17 and Lemma 38 gives

$$J_n \le \frac{\sum_{i=1}^n \mathbb{E}\|\xi_i\|_{\mathrm{F}}^4}{\|\Gamma_n\|_{\mathrm{HS}}^2} \lesssim \widetilde{\eta}\kappa_\Delta^2\cdot\frac{\lambda_\times^4}{\|\mathfrak{C}\|_{\mathrm{HS}}^2}$$

Sufficient condition for $q = \beta = 3$. Set

$$\rho_n := \widetilde{\eta}\kappa_\Delta^3 d_C^{3/2}.$$

Since $d_C \gtrsim 1$, $\kappa_\Delta \ge 1$, and $\widetilde{\eta} \le \widetilde{\eta}\kappa_\Delta = \eta\Delta_{\max} < 1$, the preceding bounds give

$$L_{3,1}^{\xi} \lesssim \rho_n^{1/2}, \qquad L_{3,2}^{\xi} \lesssim \rho_n, \qquad L_3^{g} \lesssim \rho_n^{1/2}, \qquad n^{-2}K_3^3 \lesssim \widetilde{\eta}\rho_n, \qquad J_n \lesssim \rho_n.$$

Hence a sufficient condition for all four quantities in Lemma 23 to vanish is

$$\widetilde{\eta}\kappa_\Delta^3 d_C^{3/2} \to 0.$$

G.3.2 Proof of Lemma 27

Apply Lemma 22 with $\Delta_\Gamma = \Gamma_n - \Gamma$. For any $t > 0$,

$$d_K\left(\|\mathfrak{g}\|_{\mathrm{F}}^2, \|\widetilde{\mathfrak{g}}\|_{\mathrm{F}}^2\right) \lesssim \sqrt{\frac{|\mathrm{tr}\,\Delta_\Gamma| + t}{\|\Gamma\|_{\mathrm{HS}}}} + \exp\left\{-c\left(\frac{t^2}{\|\Delta_\Gamma\|_{\mathrm{HS}}^2}\bigwedge\frac{t}{\|\Delta_\Gamma\|_{\mathrm{op}}}\right)\right\}$$

Since $\|\Delta_\Gamma\|_{\mathrm{op}} \le \|\Delta_\Gamma\|_{\mathrm{HS}} \le \|\Delta_\Gamma\|_*$, we choose $t = C(\log n)\|\Delta_\Gamma\|_*$ with $C > 0$ sufficiently large. Then

$$\frac{t^2}{\|\Delta_\Gamma\|_{\mathrm{HS}}^2} \ge C^2(\log n)^2, \qquad \frac{t}{\|\Delta_\Gamma\|_{\mathrm{op}}} \ge C\log n.$$

Thus, choosing C large enough, the exponential term is bounded by n^{-20}.

For the first term, using Lemma 17 gives

$$|\operatorname{tr}(\Delta_\Gamma)| + t \le \epsilon_n \operatorname{tr}(\Gamma) + C(\log n)\epsilon_n \operatorname{tr}(\Gamma) \lesssim \epsilon_n(\log n)\operatorname{tr}(\Gamma).$$

Therefore

$$d_K\Big(\Big\|\sum_{i=1}^n g_i\Big\|_{\mathrm{F}}^2,\ \|G\|_{\mathrm{F}}^2\Big) \lesssim \left(\epsilon_n(\log n)\kappa_\Delta \frac{\operatorname{tr}(\mathfrak{C})}{\|\mathfrak{C}\|_{\mathrm{HS}}}\right)^{1/2} + n^{-20}.$$

Finally, noticing that $\operatorname{tr}\mathfrak{C} = \mathbb{E}\left\|ZY^\top\right\|_{\mathrm{F}}^2$ and applying Lemma 38 gives $\operatorname{tr}\mathfrak{C} \lesssim C_\psi^4 \lambda_{1\sim r}^2 \nu_2^2$, we have

$$d_K\Big(\Big\|\sum_{i=1}^n g_i\Big\|_{\mathrm{F}}^2,\ \|G\|_{\mathrm{F}}^2\Big) \lesssim \left(\epsilon_n(\log n)\kappa_\Delta \frac{C_\psi^4 \lambda_{1\sim r}^2 \nu_2^2}{\|\mathfrak{C}\|_{\mathrm{HS}}}\right)^{1/2} + n^{-20}.$$

This proves the claim.

G.3.3 Proof of Lemma 28

We apply Lemma 24 with L_n and e_n as defined in the lemma statement, and set

$$\mathsf{A}_n = \sum_{i=1}^n \xi_i, \qquad \mathsf{E}_n = \eta^{-1/2}\Big[\mathcal{T}_n - \eta\sum_{i=1}^n (\mathcal{I} - \eta\mathcal{L})^{n-i}(Z_i Y_i^\top)\Big]$$

Then it suffices to bound the two probabilities and $w_n(h_n)$ in Lemma 24 as follows.

(i) Theorem 3 implies that with probability at least $1 - \delta_p$,

$$\|\mathsf{E}_n\| \lesssim \frac{1}{\sqrt{\eta}} C_1 \tau_* \left[C_5 C_\psi^2 \kappa_* \sqrt{r\widetilde{\eta}\log n} + \tau_*\right]$$

Since $\tau_* \asymp \sqrt{\widetilde{\eta}}\rho_\sigma \bar{\nu} \asymp C_5 C_\psi^2 \kappa_* \sqrt{\widetilde{\eta} r \log n}\,\bar{\nu}$ and $\eta^{-1/2} = \widetilde{\eta}^{-1/2}\Delta_{\min}^{1/2}$, we have

$$\|\mathsf{E}_n\| \le C_e \Delta_{\min}^{1/2} \kappa_*^2 r \sqrt{\widetilde{\eta}}\,(\log n)\bar{\nu}\,[1 + \bar{\nu}]$$

for some constant $C_e > 0$ large enough. Since E_n has at most r columns, $\|\mathsf{E}_n\|_{\mathrm{F}} \le \sqrt{r}\,\|\mathsf{E}_n\|$. Therefore, on the event that the above bound holds, we have

$$\|E_n\|_{\mathrm{F}} \le C_e \Delta_{\min}^{1/2} \kappa_*^2 r^{3/2} \sqrt{\widetilde{\eta}}(\log n)\bar{\nu}(1 + \bar{\nu}) =: e_n.$$

Hence

$$\mathbb{P}\left\{\|\mathsf{E}_n\|_{\mathrm{F}} > e_n\right\} \le \delta_p.$$

(ii) Lemma 21 implies that with probability at least $1 - n^{-20}$,

$$\|\mathsf{G}_n\|^2 \lesssim \operatorname{tr}\Gamma + \|\Gamma\|_{\mathrm{HS}}\sqrt{\log n} + \|\Gamma\|_{\mathrm{op}}\log n \le C_L(\operatorname{tr}\Gamma)\log n$$

for some constant $C_L > 0$ large enough. Hence

$$\mathbb{P}\left\{\|\mathsf{G}_n\| > L_n\right\} \le n^{-20}.$$

(iii) Define $h_n = e_n(2L_n + e_n)$. Lemma 20 implies that

$$w_n(h_n) \lesssim \sqrt{\frac{h_n}{\|\Gamma\|_{\mathrm{HS}}}} \lesssim \sqrt{\frac{e_n(e_n + L_n)}{\|\Gamma\|_{\mathrm{HS}}}}.$$

G.3.4 Proof of Lemma 29

Lemma 39 implies that

$$\left\|\frac{1}{\sqrt{\eta}}\sin\Theta(U_n, U_*)\right\|_{\mathrm{F}} = \left\|\frac{1}{\sqrt{\eta}}\mathcal{T}_n(\mathcal{C}_n\mathcal{C}_n^\top)^{1/2}\right\|_{\mathrm{F}}$$

with $\left\|(\mathcal{C}_n\mathcal{C}_n^\top)^{1/2} - I_r\right\| \le \|\mathcal{T}_n\|^2$.

Thus, we apply Lemma 25 with L_n being defined as in the lemma statement, $e_n = C_e\tau_*^2$ fo $C_e > 0$ for some sufficiently large constant $C_e > 0$, and

$$\mathsf{A}_n = \frac{1}{\sqrt{\eta}}\mathcal{T}_n, \qquad\qquad \mathsf{E}_n = (\mathcal{C}_n\mathcal{C}_n^\top)^{1/2} - I_r.$$

Then it suffices to bound the two probabilities and $w_n(h_n)$ in Lemma 25 as follows.

(i) Lemma 8 implies that

$$\mathbb{P}\left\{\|\mathsf{E}_n\| > e_n\right\} \le \delta_p.$$

By the sample size condition in (C.17), we have $e_n \le \frac{1}{2}$.

(ii) Similar to the proof of Lemma 28, we have

$$\mathbb{P}\left\{\|\mathsf{G}_n\| > L_n\right\} \le n^{-20}.$$

(iii) Define $h_n := \left[\frac{1}{(1-e_n)^2} - 1\right]L_n^2$. Lemma 20 implies that

$$w_n(h_n) \lesssim \sqrt{\frac{h_n}{\|\Gamma\|_{\mathrm{HS}}}} \lesssim \sqrt{\frac{e_n L_n^2}{\|\Gamma\|_{\mathrm{HS}}}},$$

where the last step follows from the fact that $\frac{1}{(1-e_n)^2} - 1 \lesssim e_n$ for $e_n \in [0, 1/2]$.

G.3.5 Proof of Lemma 30

Recall notation from Appendix F.1

$$\xi_{u,i} = \sqrt{\eta}\left[(\mathcal{I} - \eta\mathcal{L})^{n-i}(Z_iY_i^\top)\right]^\top u = \xi_i^\top u \in \mathbb{R}^r, \qquad \Gamma_{u,n} = \sum_{i=1}^n \mathrm{cov}(\xi_{u,i}).$$

Then $\mathcal{W}_u = \sqrt{\eta}\sum_{i=1}^n \xi_{u,i}$ and $\mathrm{cov}(\mathcal{W}_u) = \eta\Gamma_{u,n}$.

Note that by definition, $\Sigma_{U,m}^* = \eta\Gamma_u$. By triangle inequality, we have

$$\begin{aligned}
&\sup_{\mathcal{A}\in\mathscr{A}^r}\left|\mathbb{P}\left\{\mathcal{W}_u \in \mathcal{A}\right\} - \mathcal{N}(0, \Sigma_{U,m}^*)\left\{\mathcal{A}\right\}\right| \\
&\le \sup_{\mathcal{A}\in\mathscr{A}^r}\left|\mathbb{P}\left\{\mathcal{W}_u \in \mathcal{A}\right\} - \mathcal{N}(0, \eta\Gamma_{u,n})\left\{\mathcal{A}\right\}\right| + \sup_{\mathcal{A}\in\mathscr{A}^r}\left|\mathcal{N}(0, \eta\Gamma_{u,n})\left\{\mathcal{A}\right\} - \mathcal{N}(0, \eta\Gamma_u)\left\{\mathcal{A}\right\}\right|
\end{aligned}$$

The proof proceeds in two steps. In Step 1, we show that the first term vanishes via a Berry–Esseen bound for a sum of independent random vectors. In Step 2, we show the second term vanishes since $\Gamma_{u,n} \to \Gamma_u$.

Preliminaries. We first gather some useful results. Lemma 17(vii) implies that $\lambda_{\min}(\widetilde{\mathfrak{C}}) \lesssim r$, which combined with condition (G.5) gives

$$\eta\Delta_{\max} \le \widetilde{\eta}\kappa_* \lesssim \widetilde{\eta}\kappa_*^6 r^{7/2}\lambda_{\min}^{-3}(\widetilde{\mathfrak{C}}) \to 0.$$

and

$$\widetilde{\eta}\kappa_*^5 r^{5/2}\lambda_{\min}^{-2}(\widetilde{\mathfrak{C}}) \to 0 \tag{G.13}$$

Step 1. Gaussian approximation of $\mathcal{W}_u$ using the Berry-Esseen theorem. The Berry-Esseen theorem (Theorem 13) implies that

$$\sup_{\mathcal{A}\in\mathscr{A}^r} \left|\mathbb{P}\left\{\mathcal{W}_u \in \mathcal{A}\right\} - \mathbb{P}\left\{\mathcal{N}(0, \eta\Gamma_{u,n}) \in \mathcal{A}\right\}\right| \lesssim r^{1/4}\gamma_u$$

where

$$\begin{aligned}\gamma_u &:= \sum_{i=1}^n \mathbb{E}\left\|[\eta\Gamma_{u,n}]^{-1/2}\sqrt{\eta}\xi_{u,i}\right\|^3 = \sum_{i=1}^n \mathbb{E}\left\|\Gamma_{u,n}^{-1/2}\xi_{u,i}\right\|^3 \\ &\le \lambda_{\min}(\Gamma_{u,n})^{-3/2}\sum_{i=1}^n \mathbb{E}\left\|\xi_{u,i}\right\|^3\end{aligned}$$

To upper bound $\mathbb{E}\left\|\xi_{u,i}\right\|^3$, we first introduce some more notation. For any $j \in [r]$, denote $\mathfrak{D}_j = \operatorname{diag}\left\{1 - \eta\Delta_{lj}\right\}_{l\in[d-r]}$. Note that $\mathbb{R}^r \ni \xi_{u,i} = (\xi_{u,i1}, \cdots, \xi_{u,ir})$ with entries

$$\mathbb{R} \ni \xi_{u,ij} = \eta^{1/2}u^\top \begin{pmatrix} Y_{i1}\mathfrak{D}_1^{n-i}Z_i & Y_{i2}\mathfrak{D}_2^{n-i}Z_i & \cdots & Y_{ir}\mathfrak{D}_r^{n-i}Z_i\end{pmatrix}$$

Applying the power mean inequality gives

$$\left\|\xi_{u,i}\right\|^3 = \Big[\sum_{j=1}^r \xi_{u,ij}^2\Big]^{3/2} \le r^{3/2-1}\sum_{j=1}^r \left|\xi_{u,ij}\right|^3 = r^{1/2}\sum_{j=1}^r \left|\xi_{u,ij}\right|^3$$

Taking expectation and applying Cauchy-Schwarz inequality gives

$$\begin{aligned}\mathbb{E}\left\|\xi_{u,i}\right\|^3 &\le \eta^{3/2}r^{1/2}\sum_{j=1}^r \mathbb{E}\left|Y_{ij}u^\top\mathfrak{D}_j^{n-i}Z_i\right|^3 \\ &\le \eta^{3/2}r^{1/2}\sum_{j=1}^r \sqrt{\mathbb{E}Y_{ij}^6}\cdot\sqrt{\mathbb{E}\left[u^\top\mathfrak{D}_j^{n-i}Z_i\right]^6} \\ &\overset{(i)}{=} \eta^{3/2}r^{1/2}\sum_{j=1}^r \sqrt{\mathbb{E}Y_{ij}^6}\cdot\sqrt{\mathbb{E}\left[(\mathfrak{D}_j^{n-i}\Lambda_{*,\perp}^{1/2}u)^\top\widetilde{Z}_i\right]^6} \\ &\overset{(ii)}{\lesssim} \eta^{3/2}r^{1/2}\sum_{j=1}^r \lambda_j^{3/2}\cdot\left\|\mathfrak{D}_j^{n-i}\Lambda_{*,\perp}^{1/2}u\right\|_2^3 \\ &\le \eta^{3/2}r^{1/2}\sum_{j=1}^r \lambda_j^{3/2}\cdot(1-\widetilde{\eta})^{3(n-i)}\cdot\left\|\Lambda_{*,\perp}^{1/2}u\right\|_2^3\end{aligned}$$

where (i) follows from the commutative property $\mathfrak{D}_j^{n-i}\Lambda_{*,\perp}^{1/2} = \Lambda_{*,\perp}^{1/2}\mathfrak{D}_j^{n-i}$ since both are diagonal; (ii) is due to sub-Gaussianity.

Summing over i gives

$$\sum_{i=1}^n \mathbb{E}\left\|\xi_{u,i}\right\|^3 \lesssim \frac{\widetilde{\eta}^{1/2}}{\Delta_{\min}^{3/2}}r^{1/2}\sum_{j=1}^r \lambda_j^{3/2}\cdot\left\|\Lambda_{*,\perp}^{1/2}u\right\|_2^3 \lesssim \frac{\widetilde{\eta}^{1/2}}{\Delta_{\min}^{3/2}}r^{3/2}\lambda_1^{3/2}\cdot\left\|\Lambda_{*,\perp}^{1/2}u\right\|_2^3$$

Meanwhile, we have lower bound from Lemma 18(iii) that

$$\lambda_{\min}(\Gamma_{u,n}) \gtrsim \frac{\lambda_r}{\Delta_{\max}}\lambda_{\min}(\widetilde{\mathfrak{C}})\left\|\Lambda_{*,\perp}^{1/2}u\right\|_2^2$$

Combining the bounds above gives

$$\gamma_u = \sum_{i=1}^n \mathbb{E}\left\|\Gamma_{u,n}^{-1/2}\xi_{u,i}\right\|^3 \lesssim \widetilde{\eta}^{1/2}\kappa_\Delta^{3/2}r^{3/2}\kappa_\lambda^{3/2}\cdot\lambda_{\min}(\widetilde{\mathfrak{C}})^{-3/2}$$

Thus, we have

$$\begin{aligned}\sup_{\mathcal{A}\in\mathscr{A}^r} \left|\mathbb{P}\left\{\mathcal{W}_u \in \mathcal{A}\right\} - \mathbb{P}\left\{\mathcal{N}(0, \eta\Gamma_{u,n}) \in \mathcal{A}\right\}\right| &\lesssim \widetilde{\eta}^{1/2}\kappa_\Delta^{3/2}r^{7/4}\kappa_\lambda^{3/2}\cdot\lambda_{\min}(\widetilde{\mathfrak{C}})^{-3/2} \\ &\lesssim \widetilde{\eta}^{1/2}\kappa_*^3 r^{7/4}\lambda_{\min}(\widetilde{\mathfrak{C}})^{-3/2}\end{aligned}$$

which $\to 0$ under (G.5).

Step 2. bound TV-distance between Gaussian distributions.

$$\begin{aligned}
&\sup_{\mathcal{A}\in\mathscr{A}^r}\left|\mathbb{P}\left\{\mathcal{N}(0,\eta\Gamma_{u,n})\in\mathcal{A}\right\}-\mathbb{P}\left\{\mathcal{N}(0,\eta\Gamma_u)\in\mathcal{A}\right\}\right|\le\mathsf{TV}\left(\mathcal{N}(0,\eta\Gamma_{u,n}),\mathcal{N}(0,\eta\Gamma_u)\right)\\
&\overset{(i)}{\lesssim}\left\|\Gamma_u^{-1/2}\Gamma_{u,n}\Gamma_u^{-1/2}-I_d\right\|_{\mathrm{F}}=\left\|\Gamma_u^{-1/2}(\Gamma_{u,n}-\Gamma_u)\Gamma_u^{-1/2}\right\|_{\mathrm{F}}\\
&\le\left\|\Gamma_u^{-1/2}\right\|\left\|\Gamma_{u,n}-\Gamma_u\right\|_{\mathrm{F}}\left\|\Gamma_u^{-1/2}\right\|\le\sqrt{r}\underbrace{\left\|\Gamma_u^{-1}\right\|}_{=:\alpha_1}\underbrace{\left\|\Gamma_{u,n}-\Gamma_u\right\|}_{=:\alpha_2}
\end{aligned}$$

Here (i) follows from Lemma 42. Next, we bound α_1 and α_2 as follows.

(a) α_1: Lemma 18(iii) gives $\alpha_1\lesssim\frac{\Delta_{\max}}{\lambda_r\lambda_{\min}(\widetilde{\mathfrak{C}})\|\widetilde{u}\|^2}$.

(b) α_2: Since $\Gamma_u=\Lambda_*^{1/2}\widetilde{\Gamma}_{\widetilde{u}}\Lambda_*^{1/2}$ and $\Gamma_{u,n}=\Lambda_*^{1/2}\widetilde{\Gamma}_{\widetilde{u},n}\Lambda_*^{1/2}$ (Lemma 18(ii)), we have

$$\left\|\Gamma_{u,n}-\Gamma_u\right\|\le\lambda_1\left\|\widetilde{\Gamma}_{\widetilde{u},n}-\widetilde{\Gamma}_{\widetilde{u}}\right\|\overset{(ii)}{\le}\lambda_1\epsilon_n\kappa_\Delta\frac{\|\widetilde{\mathfrak{C}}\|_{\mathrm{op}}}{\lambda_{\min}(\widetilde{\mathfrak{C}})}\left\|\widetilde{\Gamma}_{\widetilde{u}}\right\|$$

Here (ii) follows from Lemma 18(iv).

From (F.4), we have $\widetilde{\Gamma}_{\widetilde{u}}=T_{\widetilde{u}}\widetilde{\Gamma}T_{\widetilde{u}}^*$ with $\|T_{\widetilde{u}}\|,\|T_{\widetilde{u}}^*\|\le\|\widetilde{u}\|$. Thus, we have $\left\|\widetilde{\Gamma}_{\widetilde{u}}\right\|\le\|\widetilde{u}\|^2\left\|\widetilde{\Gamma}\right\|$, which combined with Lemma 17(ii) and (vii) gives

$$\left\|\Gamma_{u,n}-\Gamma_u\right\|\lesssim\lambda_1\epsilon_n\kappa_\Delta\frac{\|\widetilde{\mathfrak{C}}\|_{\mathrm{op}}^2}{\Delta_{\min}\lambda_{\min}(\widetilde{\mathfrak{C}})}\|\widetilde{u}\|^2\lesssim\lambda_1\epsilon_n\kappa_\Delta\frac{r^2}{\Delta_{\min}\lambda_{\min}(\widetilde{\mathfrak{C}})}\|\widetilde{u}\|^2.$$

Combining (a) and (b) gives

$$\begin{aligned}
&\sup_{\mathcal{A}\in\mathscr{A}^r}\left|\mathbb{P}\left\{\mathcal{N}(0,\eta\Gamma_{u,n})\in\mathcal{A}\right\}-\mathbb{P}\left\{\mathcal{N}(0,\eta\Gamma_u)\in\mathcal{A}\right\}\right|\\
\le&\sqrt{r}\frac{\Delta_{\max}}{\lambda_r\lambda_{\min}(\widetilde{\mathfrak{C}})\|\widetilde{u}\|^2}\cdot\lambda_1\epsilon_n\kappa_\Delta\frac{r^2}{\Delta_{\min}\lambda_{\min}(\widetilde{\mathfrak{C}})}\|\widetilde{u}\|^2\\
=&\sqrt{r}\epsilon_n\kappa_\Delta\frac{\lambda_1}{\lambda_r}\frac{\Delta_{\max}}{\Delta_{\min}}\frac{r^2}{\lambda_{\min}^2(\widetilde{\mathfrak{C}})}=r^{5/2}\epsilon_n\kappa_\Delta^2\kappa_\lambda\frac{1}{\lambda_{\min}^2(\widetilde{\mathfrak{C}})}.
\end{aligned}$$

Finally, for $\epsilon_n=(1-\widetilde{\eta})^{2n}+\widetilde{\eta}\kappa_\Delta^2$ and $n\widetilde{\eta}\ge\frac{\log n}{2}$, we have $(1-\widetilde{\eta})^{2n}\le\frac{1}{n}\lesssim\widetilde{\eta}$, thus $\epsilon_n\lesssim\widetilde{\eta}\kappa_\Delta^2$, and we have

$$\begin{aligned}
\sup_{\mathcal{A}\in\mathscr{A}^r}\left|\mathbb{P}\left\{\mathcal{N}(0,\eta\Gamma_{u,n})\in\mathcal{A}\right\}-\mathbb{P}\left\{\mathcal{N}(0,\eta\Gamma_u)\in\mathcal{A}\right\}\right|&\lesssim\widetilde{\eta}\kappa_\Delta^4\kappa_\lambda r^{5/2}\lambda_{\min}^{-2}(\widetilde{\mathfrak{C}})\\
&\lesssim\widetilde{\eta}\kappa_*^5r^{5/2}\lambda_{\min}^{-2}(\widetilde{\mathfrak{C}}).
\end{aligned}$$

which $\to 0$ due to (G.13).

G.3.6 Proof of Lemma 31

Theorem 3 implies that with probability at least $1-\delta_p$,

$$\|E\|\lesssim\tau_*^2+\tau_*\kappa_*\sqrt{\widetilde{\eta}r\log n}\lesssim\kappa_*^2r\widetilde{\eta}(\log n)\bar{\nu}\left(1+\bar{\nu}\right)$$

which combined with $\left\|(\mathcal{C}\mathcal{C}^\top)^{1/2}-I_r\right\|\lesssim\|\mathcal{T}\|^2\lesssim\tau_*^2\lesssim\kappa_*^2r\widetilde{\eta}(\log n)\bar{\nu}^2$ gives

$$\begin{aligned}
\left\|\Psi_{u_m}\right\|&\lesssim\left\|U_*^\top e_m\right\|\tau_*^2+\left\|U_{*,\perp}^\top e_m\right\|\left[\tau_*^2+\tau_*\kappa_*\sqrt{\widetilde{\eta}r\log n}\right]\\
&\lesssim\left\|U_*^\top e_m\right\|\kappa_*^2\widetilde{\eta}r(\log n)\bar{\nu}^2+\left\|U_{*,\perp}^\top e_m\right\|\kappa_*^2\widetilde{\eta}r(\log n)\bar{\nu}\left(1+\bar{\nu}\right).
\end{aligned}$$

By Lemma 18, we have

$$\lambda_{\min}(\Sigma_{U,m}^*)=\eta\cdot\lambda_{\min}(\Gamma_u)\gtrsim\frac{\widetilde{\eta}}{\Delta_{\min}}\cdot\frac{\lambda_r\left\|\widetilde{u}_{u_m}\right\|^2}{\Delta_{\max}}\lambda_{\min}(\widetilde{\mathfrak{C}})\ge\widetilde{\eta}\frac{\left\|\widetilde{u}_{u_m}\right\|^2}{\lambda_1}\lambda_{\min}(\widetilde{\mathfrak{C}}).$$

Here we used $\Delta_{\min} = \lambda_r - \lambda_{r+1} \le \lambda_r$ and $\Delta_{\max} \le \lambda_1$.

Thus, we can obtain

$$\begin{aligned}
&\left\| \Sigma_{U,m}^{*-1/2} \Psi_{u_m} \right\| \lesssim \lambda_{\min}^{-1/2}(\Sigma_{U,m}^{*}) \left\| \Psi_{u_m} \right\| \\
&\lesssim \left[\widetilde{\eta} \frac{\left\| \widetilde{u}_{u_m} \right\|^2}{\lambda_1} \lambda_{\min}(\widetilde{\mathfrak{C}}) \right]^{-1/2} \cdot \left[\left\| U_*^\top e_m \right\| \kappa_*^2 \widetilde{\eta} r (\log n) \bar{\nu}^2 + \left\| U_{*,\perp}^\top e_m \right\| \kappa_*^2 \widetilde{\eta} r (\log n) \bar{\nu} \, (1 + \bar{\nu}) \right].
\end{aligned}$$

Expanding the last line gives (G.6) and (G.7).

H Proofs for Section 3.3

This appendix proves the bootstrap results in Section 3.3. The proof has two main components.

First, Appendix H.1 establishes stability and linearization of the bootstrap tangent process. We first work under the joint law of the truncated observations constructed in Appendix D and the multiplier weights. The resulting bounds are then transferred to the original observations through the truncation coupling.

Second, Appendices H.2 and H.3 condition on $\mathcal{F}_X$, i.e., on the initialization and the original observations, and analyze the resulting frozen multiplier sums. Appendix H.2 proves the global bootstrap approximation in Theorem 7, while Appendix H.3 proves the row-wise bootstrap approximation in Theorem 8.

We will also use the deterministic two-subspace alignment identity in Lemma 40, which relates the tangent difference of two subspaces to their Procrustes-aligned difference. It will be invoked in both Appendices H.2 and H.3.

H.1 Bootstrap stability and linearization

Joint coupling and truncated bootstrap process. For the proof, extend the data–bootstrap probability space to include the truncation coupling from Appendix D. Let $\mathbb{P}_{\text{joint}}$ denote the joint law of

$$\left(U_0, \{(X_i, \overline{X}_i)\}_{i=1}^n, \{w_i^{\mathsf{bs}}\}_{i=1}^n \right),$$

where $\{(X_i, \overline{X}_i)\}_{i=1}^n$ are the i.i.d. coupled original and truncated observations constructed in (D.7) Appendix D, and $\{w_i^{\mathsf{bs}}\}_{i=1}^n$ is an independent multiplier sequence satisfying Assumption 2. Let $\mathbb{E}_{\text{joint}}$ denote expectation under $\mathbb{P}_{\text{joint}}$. The $(U_0, X_1, \dots, X_n, w_1^{\mathsf{bs}}, \dots, w_n^{\mathsf{bs}})$ marginal agrees with the data–bootstrap law of Section 3.3; in particular, $\mathbb{P}_X$ is its $(U_0, X_1, \dots, X_n)$ marginal, and conditional on $\mathcal{F}_X$ the multiplier sequence has probability law $\mathbb{P}^*$.

Define the truncated bootstrap iterates

$$\overline{U}_0^{\mathsf{bs}} = U_0, \qquad \overline{U}_k^{\mathsf{bs}} = \mathsf{QR}(\overline{U}_{k-1}^{\mathsf{bs}} + \eta w_k^{\mathsf{bs}} \overline{X}_k \overline{X}_k^\top \overline{U}_{k-1}^{\mathsf{bs}}), \qquad k \in [n],$$

and let $\overline{\mathcal{T}}_k^{\mathsf{bs}}$ denote the corresponding tangent matrix.

Define the truncated bootstrap filtration

$$\overline{\mathcal{F}}_i := \sigma\left(U_0, \overline{X}_1, w_1^{\mathsf{bs}}, \dots, \overline{X}_i, w_i^{\mathsf{bs}} \right), \qquad \overline{\mathbb{E}}_i[\cdot] := \mathbb{E}_{\text{joint}}\left[\cdot \mid \overline{\mathcal{F}}_i \right].$$

We work under $\mathbb{P}_{\text{joint}}$ throughout the joint-law analysis below and pass to conditional statements given $\mathcal{F}_X$ only after establishing the corresponding joint-law bounds.

For brevity, write

$$\overline{\mathcal{T}}^{\mathsf{bs}} = \overline{\mathcal{T}}_i^{\mathsf{bs}}, \qquad \overline{\mathcal{T}}_+^{\mathsf{bs}} = \overline{\mathcal{T}}_{i+1}^{\mathsf{bs}}, \qquad \overline{Y}_+ = \overline{Y}_{i+1}, \qquad \overline{Z}_+ = \overline{Z}_{i+1}, \qquad w_+^{\mathsf{bs}} = w_{i+1}^{\mathsf{bs}}.$$

Similar to (D.8), the Sherman–Morrison formula gives

$$\overline{\mathcal{T}}_+^{\mathsf{bs}} - \overline{\mathcal{T}}^{\mathsf{bs}} = \frac{\eta w_+^{\mathsf{bs}} \overline{F}_+^{\mathsf{bs}}}{1 + \eta w_+^{\mathsf{bs}} \overline{\zeta}_+^{\mathsf{bs}}} = \eta w_+^{\mathsf{bs}} \overline{F}_+^{\mathsf{bs}} - (\eta w_+^{\mathsf{bs}})^2 \frac{\overline{\zeta}_+^{\mathsf{bs}} \overline{F}_+^{\mathsf{bs}}}{1 + \eta w_+^{\mathsf{bs}} \overline{\zeta}_+^{\mathsf{bs}}}. \tag{H.1}$$

where

$$\overline{F}_{+}^{\mathsf{bs}} := \left(-\overline{\mathcal{T}}^{\mathsf{bs}}\overline{Y}_{+} + \overline{Z}_{+}\right)\left(\overline{Y}_{+}^{\top} + \overline{Z}_{+}^{\top}\overline{\mathcal{T}}^{\mathsf{bs}}\right) \in \mathbb{R}^{(d-r)\times r}, \qquad \overline{\zeta}_{+}^{\mathsf{bs}} := \left(\overline{Y}_{+}^{\top} + \overline{Z}_{+}^{\top}\overline{\mathcal{T}}^{\mathsf{bs}}\right)\overline{Y}_{+} \in \mathbb{R}.$$

Let $\overline{\mathcal{L}}$ be the drift operator of the truncated process defined in (D.9). Since w_{+}^{bs} is independent of $\overline{\mathcal{F}}_i \vee \sigma(\overline{X}_{i+1})$ and $\mathbb{E}w_{+}^{\mathsf{bs}} = 1$, we have $\overline{\mathbb{E}}_i[w_{+}^{\mathsf{bs}}\overline{F}_{+}^{\mathsf{bs}}] = \overline{\mathbb{E}}_i[\overline{F}_{+}^{\mathsf{bs}}] = -\overline{\mathcal{L}}\overline{\mathcal{T}}^{\mathsf{bs}}$, and the following decomposition similar to (D.9):

$$\overline{\mathcal{T}}_{+}^{\mathsf{bs}} = (\mathcal{I} - \eta\overline{\mathcal{L}})\overline{\mathcal{T}}^{\mathsf{bs}} + \overline{D}_{+}^{\mathsf{bs}} + \overline{R}_{+}^{\mathsf{bs}} \tag{H.2}$$

where

$$\begin{aligned}
\overline{D}_{+}^{\mathsf{bs}} &:= (\overline{\mathcal{T}}_{+}^{\mathsf{bs}} - \overline{\mathcal{T}}^{\mathsf{bs}}) - \overline{\mathbb{E}}_i[\overline{\mathcal{T}}_{+}^{\mathsf{bs}} - \overline{\mathcal{T}}^{\mathsf{bs}}] = \overline{D}_{+}^{\mathsf{bs},(1)} + \overline{D}_{+}^{\mathsf{bs},(2)},\\
\overline{D}_{+}^{\mathsf{bs},(1)} &:= \eta w_{+}^{\mathsf{bs}}\overline{F}_{+}^{\mathsf{bs}} - \overline{\mathbb{E}}_i[\eta w_{+}^{\mathsf{bs}}\overline{F}_{+}^{\mathsf{bs}}],\\
\overline{D}_{+}^{\mathsf{bs},(2)} &:= -(\eta w_{+}^{\mathsf{bs}})^2\frac{\overline{\zeta}_{+}^{\mathsf{bs}}\overline{F}_{+}^{\mathsf{bs}}}{1+\eta w_{+}^{\mathsf{bs}}\overline{\zeta}_{+}^{\mathsf{bs}}} + \overline{\mathbb{E}}_i\left[(\eta w_{+}^{\mathsf{bs}})^2\frac{\overline{\zeta}_{+}^{\mathsf{bs}}\overline{F}_{+}^{\mathsf{bs}}}{1+\eta w_{+}^{\mathsf{bs}}\overline{\zeta}_{+}^{\mathsf{bs}}}\right]\\
\overline{R}_{+}^{\mathsf{bs}} &:= -\overline{\mathbb{E}}_i\left[(\eta w_{+}^{\mathsf{bs}})^2\frac{\zeta_{+}^{\mathsf{bs}}F_{+}^{\mathsf{bs}}}{1+\eta w_{+}^{\mathsf{bs}}\zeta_{+}^{\mathsf{bs}}}\right]
\end{aligned}$$

As in (D.10) of Appendix D, absorb the covariance perturbation into

$$\overline{R}_{+}^{\mathsf{bs},H} := \overline{R}_{+}^{\mathsf{bs}} + \eta\left(\mathcal{L} - \overline{\mathcal{L}}\right)\overline{\mathcal{T}}^{\mathsf{bs}}.$$

Then

$$\overline{\mathcal{T}}_{+}^{\mathsf{bs}} = (\mathcal{I} - \eta\mathcal{L})\overline{\mathcal{T}}^{\mathsf{bs}} + \overline{D}_{+}^{\mathsf{bs}} + \overline{R}_{+}^{\mathsf{bs},H}. \tag{H.3}$$

Lemma 32 (One-step bounds for the truncated bootstrap process)**.** *Assume Assumptions 1 and 2 hold. Suppose $\overline{\mathcal{T}}^{\mathsf{bs}}$ is well-defined with $\|\overline{\mathcal{T}}^{\mathsf{bs}}\| = \tau$. If*

$$C_w C_{\psi,n}^2\widetilde{\eta}\kappa_* r(1+\tau\nu_2) \le \frac{1}{2}, \tag{H.4}$$

then the following hold, where the implicit constants may depend only on C_w.

(I) *The full centered noise satisfies $\|\overline{D}_{+}^{\mathsf{bs}}\| \lesssim C_{\psi,n}^2\widetilde{\eta}\kappa_* r(1+\tau\nu_2)(\tau+\nu_2)$ almost surely, and*

$$\max\left\{\left\|\overline{\mathbb{E}}_i\overline{D}_{+}^{\mathsf{bs}}(\overline{D}_{+}^{\mathsf{bs}})^\top\right\|^{1/2}, \left\|\overline{\mathbb{E}}_i(\overline{D}_{+}^{\mathsf{bs}})^\top\overline{D}_{+}^{\mathsf{bs}}\right\|^{1/2}\right\} \lesssim C_\psi^2\widetilde{\eta}\kappa_*\sqrt{r}(1+\tau\nu_\infty)(\tau+\overline{\nu}).$$

(II) *The centered second-order noise satisfies $\|\overline{D}_{+}^{\mathsf{bs},(2)}\| \lesssim C_{\psi,n}^4\widetilde{\eta}^2\kappa_*^2 r^2(1+\tau\nu_2)^2(\tau+\nu_2)$ almost surely, and*

$$\max\left\{\left\|\overline{\mathbb{E}}_i\overline{D}_{+}^{\mathsf{bs},(2)}\overline{D}_{+}^{\mathsf{bs},(2)\top}\right\|^{1/2}, \left\|\overline{\mathbb{E}}_i\overline{D}_{+}^{\mathsf{bs},(2)\top}\overline{D}_{+}^{\mathsf{bs},(2)}\right\|^{1/2}\right\} \lesssim C_\psi^4\widetilde{\eta}^2\kappa_*^2 r^{3/2}(1+\tau\nu_\infty)^2(\tau+\overline{\nu}).$$

(III) *The modified drift remainder satisfies*

$$\|\overline{R}_{+}^{\mathsf{bs},H}\| \lesssim C_\psi^4\widetilde{\eta}^2\kappa_*^2 r^{3/2}(1+\tau\nu_\infty)^2(\tau+\nu_\infty).$$

Proof. The proof follows the same argument as that of Lemma 12, after enlarging the constants by quantities depending only on C_w. □

Lemma 32 provides the same inputs as Lemma 12: the linear part of the recursion is $\mathcal{I} - \eta\mathcal{L}$, the centered increments satisfy the same almost-sure and conditional-variance bounds, and the modified drift remainder has the same order. Hence, after enlarging constants by factors depending only on C_w, the convergence argument of Appendix D applies verbatim to the truncated bootstrap process.

Lemma 33 (Convergence of the truncated bootstrap process)**.** *Suppose Assumptions 1, 2, 4, and 5 hold. Assume*

$$\nu_\infty \geq c_\nu n^{-C_\nu} \qquad \text{for constants } C_\nu \geq 0, c_\nu > 0.$$

Let $\varepsilon \in [0, 1/2)$ *and set*

$$\eta = \frac{c_\eta \log n}{n\Delta_{\min}}, \quad \text{where} \quad c_\eta \geq c_\eta^* := \left(1 - \varepsilon + \frac{3C_\nu}{2}\right)\left(1 + \frac{\log^{-3} n + \log 2}{\varepsilon \log n + \log\log n}\right).$$

If

$$n^{1-2\varepsilon} \geq C_{\mathsf{bs}} c_\eta \kappa_*^2 r^3 (\log n)^4 \frac{\log(1/\delta_p)}{\delta_p^2} \max(1, \nu_\infty d)$$

for a sufficiently large constant $C_{\mathsf{bs}} > 0$ *depending only on* $C_\psi, C_\nu, c_\nu, C_d$*, and* C_w*, then, under* $\mathbb{P}_{\mathrm{joint}}$*, with probability at least* $1 - n^{-19}$*,*

$$\left\|\overline{\mathcal{T}}_k^{\mathsf{bs}}\right\| \leq 2\overline{\tau}_*^{\mathsf{bs}} \asymp \kappa_* \bar{\nu}\sqrt{\widetilde{\eta} r \log n} = \frac{\lambda_1}{\Delta_{\min}}\sqrt{c_\eta r}\,\bar{\nu}\frac{\log n}{\sqrt{n}}, \qquad \text{for all } \left\lfloor \tfrac{c_\eta^*}{c_\eta} n \right\rfloor \leq k \leq n,$$

where

$$\overline{\tau}_*^{\mathsf{bs}} := \overline{C}_{\mathsf{bs}} \max\left(C_\psi^2 \kappa_* \bar{\nu}\sqrt{\widetilde{\eta} r \log n}, C_\psi^2 \kappa_* \nu_2 \widetilde{\eta} r (\log n)^2\right)$$

for some constant $\overline{C}_{\mathsf{bs}} > 0$ *(depending only on* C_d*).*

Proof. After enlarging the constants by quantities depending only on C_w, the stopped-process, matrix-Freedman, epoch-chaining, and iteration-budget arguments in Lemmas 5–8 apply verbatim to $\{\overline{\mathcal{T}}_k^{\mathsf{bs}}\}_{k \leq n}$. $\square$

Linearization of the truncated bootstrap process. We next establish the bootstrap counterpart of Lemma 15. By the definition of $\overline{D}_+^{(1)}$ and $\overline{D}_+^{\mathsf{bs},(1)}$ in (D.9) and (H.2), we have

$$\begin{aligned} \overline{D}_+^{(1)} &= \eta(\overline{F}_+ - \overline{\mathbb{E}}_i \overline{F}_+), \\ \overline{D}_+^{\mathsf{bs},(1)} &= \eta w_+^{\mathsf{bs}}(\overline{F}_+^{\mathsf{bs}} - \overline{\mathbb{E}}_i \overline{F}_+^{\mathsf{bs}}) + \eta(w_+^{\mathsf{bs}} - 1)\overline{\mathbb{E}}_i \overline{F}_+^{\mathsf{bs}} \end{aligned} \tag{H.5}$$

Thus, the first term in $\overline{D}_+^{\mathsf{bs},(1)}$ has the same centered structure as $\overline{D}_+^{(1)}$, with the bounded multiplier w_+^{bs} affecting only the constants in the corresponding concentration bounds. The second term is new: since $\mathbb{E}(w_+^{\mathsf{bs}} - 1) = 0$, it gives a conditionally centered multiplier martingale. As shown in the proof below, its accumulated contribution is of the same order as the existing linearization remainder and can be absorbed into it. Consequently, the linearization argument of Lemma 15 carries over to the bootstrap process up to constant factors.

Lemma 34 (Linearization of the truncated bootstrap process)**.** *Under the assumptions of Lemma 33, suppose additionally that* $c_\eta \geq c_\eta^* + 1/2$*, and let* $N_* = \lfloor n c_\eta^*/c_\eta \rfloor$*. Then, with probability at least* $1 - O(n^{-18})$*,*

$$\left\|\overline{\mathcal{T}}_n^{\mathsf{bs}} - \eta\sum_{i=1}^n w_i^{\mathsf{bs}}(\mathcal{I} - \eta\mathcal{L})^{n-i}(\overline{Z}_i \overline{Y}_i^\top)\right\| \lesssim \overline{\tau}_*^{\mathsf{bs}}\left[C_\psi^2 \kappa_* \sqrt{r\widetilde{\eta}\log n} + \overline{\tau}_*^{\mathsf{bs}}\right] + e^{-\widetilde{\eta}(n-N_*)}\kappa_* \nu_\infty \|H\|.$$

In particular, by (D.3)*,*

$$\left\|\overline{\mathcal{T}}_n^{\mathsf{bs}} - \eta\sum_{i=1}^n w_i^{\mathsf{bs}}(\mathcal{I} - \eta\mathcal{L})^{n-i}(\overline{Z}_i \overline{Y}_i^\top)\right\| \lesssim \overline{\tau}_*^{\mathsf{bs}}\left[C_\psi^2 \kappa_* \sqrt{r\widetilde{\eta}\log n} + \overline{\tau}_*^{\mathsf{bs}}\right].$$

Proof. See Appendix H.4. $\square$

Transfer to the original process under $\mathbb{P}_{\mathrm{joint}}$. Use the same initialization and multiplier sequence for the original and truncated bootstrap processes. On the truncation coupling event

$$\mathcal{E}_{\mathrm{tr}} := \{X_i = \overline{X}_i,\ i \in [n]\},$$

we have simultaneously

$$U_k = \overline{U}_k, \qquad U_k^{\mathsf{bs}} = \overline{U}_k^{\mathsf{bs}}, \qquad k \in [n].$$

Moreover,

$$\mathbb{P}_{\mathrm{joint}}(\mathcal{E}_{\mathrm{tr}}^c) \le n\delta_{\mathrm{tr}} \le n^{-19}.$$

The preceding results and the truncation coupling give the following joint-law stability and linearization event.

Corollary 10 (Joint stability and linearization)**.** *Under the assumptions of Lemma 34, define*

$$\tau_*^{\dagger} o := \tau_* \vee \tau_*^{\mathsf{bs}} \lesssim \tau_*$$

and linearization remainder

$$r_n^{\mathrm{lin}} := C_{\mathrm{lin}}(C_w)\, \tau_*^{\dagger} \left[C_\psi^2 \kappa_* \sqrt{r\widetilde{\eta} \log n} + \tau_*^{\dagger}\right].$$

Then there exists an event $\mathcal{A}_n^{\mathrm{lin}}$ on the joint data–bootstrap probability space such that

$$\mathbb{P}_{\mathrm{joint}}(\mathcal{A}_n^{\mathrm{lin}}) \lesssim n^{-18},$$

and on $(\mathcal{A}_n^{\mathrm{lin}})^c$,

$$\|\mathcal{T}_n\| \le 2\tau_*^{\dagger}, \qquad \|\mathcal{T}_n^{\mathsf{bs}}\| \le 2\tau_*^{\dagger},$$

and

$$\left\|\mathcal{T}_n^{\mathsf{bs}} - \mathcal{T}_n - \eta \sum_{i=1}^{n} (w_i^{\mathsf{bs}} - 1)(\mathcal{I} - \eta\mathcal{L})^{n-i}(Z_i Y_i^{\top})\right\| \le r_n^{\mathrm{lin}}.$$

Proof. Combine Lemma 15, Lemma 34, and the truncation coupling (D.7). □

Transfer to the conditional bootstrap law. Recall from Section 3.3 that $\mathcal{F}_X = \sigma(U_0, X_1, ..., X_n)$, that $\mathbb{P}_X$ denotes the law of $(U_0, X_1, \ldots, X_n)$, and that $\mathbb{P}^*$ and $\mathbb{E}^*$ denote the conditional bootstrap probability and expectation given $\mathcal{F}_X$. Under the joint construction above, equivalently,

$$\mathbb{P}^*(\cdot) = \mathbb{P}_{\mathrm{joint}}(\,\cdot \mid \mathcal{F}_X), \qquad \mathbb{E}^*[\cdot] = \mathbb{E}_{\mathrm{joint}}[\,\cdot \mid \mathcal{F}_X].$$

For any event A on the joint probability space and any $a > 0$, the Markov–Fubini argument gives

$$\mathbb{P}_X\{\mathbb{P}^*(\mathcal{A}) > a\} \le \frac{\mathbb{P}_{\mathrm{joint}}(\mathcal{A})}{a}. \tag{H.6}$$

Indeed, $\mathbb{E}_X[\mathbb{P}^*(\mathcal{A})] = \mathbb{P}_{\mathrm{joint}}(\mathcal{A})$, and the claim follows from Markov's inequality.

Corollary 11 (Conditional bootstrap stability and linearization)**.** *Under the assumptions of Corollary 10, there exists an $\mathcal{F}^X$-measurable event $\mathcal{G}_n^{\mathrm{lin}}$ satisfying*

$$\mathbb{P}_X(\mathcal{G}_n^{\mathrm{lin}}) \ge 1 - O(n^{-9}),$$

such that, on $\mathcal{G}_n^{\mathrm{lin}}$,

$$\|\mathcal{T}_n\| \le 2\tau_*^{\dagger}, \qquad \mathbb{P}^*\big\{\|\mathcal{T}_n^{\mathsf{bs}}\| > 2\tau_*^{\dagger}\big\} \le n^{-9},$$

and

$$\mathbb{P}^*\Big\{\Big\|\mathcal{T}_n^{\mathsf{bs}} - \mathcal{T}_n - \eta \sum_{i=1}^{n} (w_i^{\mathsf{bs}} - 1)(\mathcal{I} - \eta\mathcal{L})^{n-i}(Z_i Y_i^{\top})\Big\| > r_n^{\mathrm{lin}}\Big\} \le n^{-9}.$$

Proof. Let $\mathcal{A}_n^{\mathrm{lin}}$ be the event in Corollary 10, and define

$$\mathcal{G}_n^{\mathrm{lin}} := \left\{\mathbb{P}_{\mathrm{joint}}(\mathcal{A}_n^{\mathrm{lin}} \mid \mathcal{F}_X) \le n^{-9}\right\}.$$

Since $\mathbb{P}_{\mathrm{joint}}(\mathcal{A}_n^{\mathrm{lin}}) \lesssim n^{-18}$, (H.6) gives

$$\mathbb{P}_X\big((\mathcal{G}_n^{\mathrm{lin}})^c\big) \lesssim n^{-9}.$$

The claimed conditional bounds follow directly from the definition of $\mathcal{A}_n^{\mathrm{lin}}$.

Finally, since $\|\mathcal{T}_n\|$ is $\mathcal{F}^X$-measurable, if $\|\mathcal{T}_n\| > 2\tau_*^{\dagger}$, then $\mathbb{P}^*(\mathcal{A}_n^{\mathrm{lin}}) = 1$. Hence $\|\mathcal{T}_n\| \le 2\tau_*^{\dagger}$ on $\mathcal{G}_n^{\mathrm{lin}}$. □

H.2 Proof of Theorem 7

Recall $\mathcal{F}_X$, $\mathbb{P}_X$, $\mathbb{P}^*$, and $\mathbb{E}^*$ from Section 3.3. Write the Kolmogorov distance between conditional laws (on $\mathcal{F}_X$) of two random variables V_1, V_2 as

$$d_K^*(V_1, V_2) := \sup_{t\in\mathbb{R}} \left|\mathbb{P}^*\{V_1 \le t\} - \mathbb{P}^*\{V_2 \le t\}\right|.$$

Note that when the law of V is independent of $\mathcal{F}_X$, $\mathbb{P}^*\{V \le t\} = \mathbb{P}\{V \le t\}$.

Under the assumptions of Theorem 7, the conditions of Proposition 3 hold with $\delta_p \to 0$. As in the proof of Theorem 1, condition (11) implies Assumption 4, while the Haar initialization satisfies Assumption 5 with probability $1 - o(1)$. Thus, after intersecting with this initialization event, Corollary 11 yields an $\mathcal{F}_X$-measurable event $\mathcal{G}_n^{\mathrm{lin}}$ such that $\mathbb{P}_X(\mathcal{G}_n^{\mathrm{lin}}) \to 1$ and, on $\mathcal{G}_n^{\mathrm{lin}}$,

$$\|\mathcal{T}_n\| \le 2\tau_*^\dagger, \qquad \mathbb{P}^*\{\|\mathcal{T}_n^{\mathsf{bs}}\| > 2\tau_*^\dagger\} \le n^{-9},$$

and

$$\mathbb{P}^*\Big\{\Big\|\mathcal{T}_n^{\mathsf{bs}} - \mathcal{T}_n - \eta\sum_{i=1}^n (w_i^{\mathsf{bs}} - 1)(\mathcal{I} - \eta\mathcal{L})^{n-i}(Z_iY_i^\top)\Big\| > r_n^{\mathrm{lin}}\Big\} \le n^{-9}.$$

For $i \in [n]$, define the $\mathcal{F}_X$-measurable matrices

$$M_i := (\mathcal{I} - \eta\mathcal{L})^{n-i}(Z_iY_i^\top), \qquad \xi_i^{\mathsf{bs}} := \sqrt{\eta}\,(w_i^{\mathsf{bs}} - 1)M_i, \qquad \Xi_n^{\mathsf{bs}} := \sum_{i=1}^n \xi_i^{\mathsf{bs}}.$$

Also define the conditional covariance operator

$$\widehat{\Gamma}_n := \eta\sum_{i=1}^n M_i \otimes M_i.$$

Since the multiplier weights are independent of $\mathcal{F}_X$ with $\mathbb{E}(w_i^{\mathsf{bs}} - 1) = 0$ and $\mathbb{E}(w_i^{\mathsf{bs}} - 1)^2 = 1$, conditionally on $\mathcal{F}_X$ the ξ_i^{bs} are independent and centered, with

$$\mathrm{cov}^*(\Xi_n^{\mathsf{bs}}) = \widehat{\Gamma}_n, \qquad \mathbb{E}_X\widehat{\Gamma}_n = \Gamma_n = \eta\sum_{i=1}^n (\mathcal{I} - \eta\mathcal{L})^{n-i}\mathfrak{C}(\mathcal{I} - \eta\mathcal{L})^{n-i}.$$

Let $\mathsf{G} \sim N(0, \Gamma)$, where Γ is defined in Appendix F. It suffices to prove

$$d_K^*\left(\eta^{-1}\|\sin\Theta(U_n^{\mathsf{bs}}, U_n)\|_{\mathrm{F}}^2, \|\mathsf{G}\|_{\mathrm{F}}^2\right) \xrightarrow{\mathbb{P}_X} 0,$$

because Theorem 5 and the triangle inequality then give the claim. Similar to the proof of Theorem 5, the proof proceeds in three steps, progressing right-to-left through the chain

$$\eta^{-1}\left\|\sin\Theta(U_n^{\mathsf{bs}}, U_n)\right\|_{\mathrm{F}}^2 \overset{\text{Step 3.}}{\approx} \eta^{-1}\left\|\mathcal{T}_n^{\mathsf{bs}} - \mathcal{T}_n\right\|_{\mathrm{F}}^2 \overset{\text{Step 2.}}{\approx} \left\|\Xi_n^{\mathsf{bs}}\right\|_{\mathrm{F}}^2 \overset{\text{Step 1.}}{\approx} \|\mathsf{G}\|_{\mathrm{F}}^2.$$

Step 1: Conditional Gaussian approximation of $\|\Xi_n^{\mathsf{bs}}\|_{\mathrm{F}}^2$. We first record an empirical moment bound. For each fixed $p \in \{2, 3, 4, 6\}$,

$$S_p := \sum_{i=1}^n \|M_i\|_{\mathrm{F}}^p = O_{\mathbb{P}_X}\left(\frac{\lambda_\times^p}{\widetilde{\eta}}\right). \tag{H.7}$$

Indeed, by Lemma 36, $\|M_i\|_{\mathrm{F}} \le (1 - \widetilde{\eta})^{n-i}\|Z_iY_i^\top\|_{\mathrm{F}}$, and therefore Lemma 17(vi) and geometric summation give

$$\mathbb{E}_X\sum_{i=1}^n \|M_i\|_{\mathrm{F}}^p \lesssim \lambda_\times^p\sum_{i=1}^n (1 - \widetilde{\eta})^{p(n-i)} \lesssim \frac{\lambda_\times^p}{\widetilde{\eta}}.$$

We next compare $\widehat{\Gamma}_n$ with Γ. Since $\|M \otimes M\|_{\mathrm{HS}} = \|M\|_{\mathrm{F}}^2$, independence gives

$$\mathbb{E}_X\|\widehat{\Gamma}_n - \Gamma_n\|_{\mathrm{HS}}^2 \lesssim \eta^2\sum_{i=1}^n \mathbb{E}\|M_i\|_{\mathrm{F}}^4 \lesssim \frac{\widetilde{\eta}\,\lambda_\times^4}{\Delta_{\min}^2},$$

and similarly since $\operatorname{tr}(M \otimes M) = \|M\|_{\mathrm{F}}^2$, we have

$$\mathbb{E}_X \left|\operatorname{tr}(\widehat{\Gamma}_n - \Gamma_n)\right|^2 \lesssim \frac{\widetilde{\eta}\,\lambda_\times^4}{\Delta_{\min}^2}.$$

Since Lemma 17(ii) gives

$$\|\Gamma\|_{\mathrm{HS}} \gtrsim \frac{\|\mathfrak{C}\|_{\mathrm{HS}}}{\Delta_{\max}} = \frac{\lambda_\times^2}{\Delta_{\max}\sqrt{d_{\mathfrak{C}}}}, \tag{H.8}$$

we obtain

$$\frac{\|\widehat{\Gamma}_n - \Gamma_n\|_{\mathrm{HS}}}{\|\Gamma\|_{\mathrm{HS}}} = O_{\mathbb{P}_X}\left(\kappa_\Delta \sqrt{\widetilde{\eta}\, d_{\mathfrak{C}}}\right),$$

and the same bound holds with $\|\widehat{\Gamma}_n - \Gamma_n\|_{\mathrm{HS}}$ replaced by $|\operatorname{tr}(\widehat{\Gamma}_n - \Gamma_n)|$.

Let $\varepsilon_n := (1-\widetilde{\eta})^{2n} + \widetilde{\eta}\kappa_\Delta^2$. By Lemma 17(iv),

$$\|\Gamma_n - \Gamma\|_{\mathrm{HS}} \le \varepsilon_n \|\Gamma\|_{\mathrm{HS}}, \qquad |\operatorname{tr}(\Gamma_n - \Gamma)| \le \varepsilon_n \operatorname{tr}\Gamma.$$

Moreover, Lemma 17(ii) and (vi) imply

$$\frac{\operatorname{tr}\Gamma}{\|\Gamma\|_{\mathrm{HS}}} \lesssim \kappa_\Delta \sqrt{d_{\mathfrak{C}}}.$$

Consequently, under condition (15),

$$\frac{\|\widehat{\Gamma}_n - \Gamma\|_{\mathrm{HS}}}{\|\Gamma\|_{\mathrm{HS}}} = o_{\mathbb{P}_X}(1), \qquad \frac{|\operatorname{tr}(\widehat{\Gamma}_n - \Gamma)|}{\|\Gamma\|_{\mathrm{HS}}} = o_{\mathbb{P}_X}(1), \tag{H.9}$$

which combined with (H.8) also implies

$$\|\widehat{\Gamma}\|_{\mathrm{HS}} \gtrsim_{\mathbb{P}_X} \frac{\lambda_\times^2}{\Delta_{\max}\sqrt{d_{\mathfrak{C}}}}, \tag{H.10}$$

We now apply Lemma 23 conditionally on $\mathcal{F}_X$, with $q = \beta = 3$. Let $L_3^{\xi,\mathsf{bs}}$, $L_3^{g,\mathsf{bs}}$, K_3^{bs} and J_n^{bs} denote the corresponding conditional quantities. More precisely,

$$\begin{aligned}
L_q^{\xi,\mathsf{bs}} &= \sum_{i=1}^n \frac{\mathbb{E}^*\left\langle \xi_i^{\mathsf{bs}},\, \widehat{\Gamma}_n \xi_i^{\mathsf{bs}}\right\rangle_{\mathrm{F}}^{q/2}}{\left\|\widehat{\Gamma}_n\right\|_{\mathrm{HS}}^q} + \frac{n^2}{\binom{n}{2}} \sum_{1\le i<j\le n} \frac{\mathbb{E}^*\left|\left\langle \xi_i^{\mathsf{bs}},\, \xi_j^{\mathsf{bs}}\right\rangle_{\mathrm{F}}\right|^q}{\left\|\widehat{\Gamma}_n\right\|_{\mathrm{HS}}^q},\\
L_q^{\mathfrak{g},\mathsf{bs}} &= \sum_{i=1}^n \frac{\mathbb{E}^*\left\langle \mathfrak{g}_i^{\mathsf{bs}},\, \widehat{\Gamma}_n \mathfrak{g}_i^{\mathsf{bs}}\right\rangle_{\mathrm{F}}^{q/2}}{\left\|\widehat{\Gamma}_n\right\|_{\mathrm{HS}}^q}, \quad \text{where} \quad \mathfrak{g}_i^{\mathsf{bs}} \sim N(0, \eta M_i \otimes M_i),\\
(K_\beta^{\mathsf{bs}})^\beta &= \frac{n^\beta}{n} \sum_{i=1}^n \mathbb{E}^* \left| \frac{\left\langle \xi_i^{\mathsf{bs}},\, \xi_i^{\mathsf{bs}}\right\rangle_{\mathrm{F}} - \mathbb{E}^*\left\langle \xi_i^{\mathsf{bs}},\, \xi_i^{\mathsf{bs}}\right\rangle_{\mathrm{F}}}{\left\|\widehat{\Gamma}_n\right\|_{\mathrm{HS}}} \right|^\beta\\
J_n^{\mathsf{bs}} &= \frac{\sum_{i=1}^n \mathfrak{v}_i^{\mathsf{bs}}}{\left\|\widehat{\Gamma}_n\right\|_{\mathrm{HS}}^2}, \qquad \text{where} \quad \mathfrak{v}_i = \operatorname{Var}^*(\left\|\xi_i^{\mathsf{bs}}\right\|_{\mathrm{F}}^2)
\end{aligned}$$

Lemma 35. *Under* (H.7), (H.10) *and Assumption 2,*

$$\begin{aligned}
L_3^{\xi,\mathsf{bs}} &= O_{\mathbb{P}_X}\left(\widetilde{\eta}^{1/2}\kappa_\Delta^{3/2} d_{\mathfrak{C}}^{3/4} + \widetilde{\eta}\kappa_\Delta^3 d_{\mathfrak{C}}^{3/2}\right),\\
L_3^{g,\mathsf{bs}} &= O_{\mathbb{P}_X}\left(\widetilde{\eta}^{1/2}\kappa_\Delta^{3/2} d_{\mathfrak{C}}^{3/4}\right),\\
n^{-2}(K_3^{\mathsf{bs}})^3 &= O_{\mathbb{P}_X}\left(\widetilde{\eta}^2\kappa_\Delta^3 d_{\mathfrak{C}}^{3/2}\right),\\
J_n^{\mathsf{bs}} &= O_{\mathbb{P}_X}\left(\widetilde{\eta}\kappa_\Delta^2 d_{\mathfrak{C}}\right).
\end{aligned}$$

Proof. See Appendix H.5. ∎

All four quantities converge to zero under condition (15)(i). Hence the conditional version of Lemma 23 gives

$$d_K^*\left(\|\Xi_n^{\mathsf{bs}}\|_{\mathrm{F}}^2, \|\mathsf{G}_{\widehat{\Gamma}_n}\|_{\mathrm{F}}^2\right) \xrightarrow{\mathbb{P}_X} 0, \tag{H.11}$$

where, conditionally on the data, $\mathsf{G}_{\widehat{\Gamma}_n} \sim N(0, \widehat{\Gamma}_n)$.

It remains to replace $\widehat{\Gamma}_n$ by Γ. Set

$$a_n := \frac{|\operatorname{tr}(\widehat{\Gamma}_n - \Gamma)|}{\|\Gamma\|_{\mathrm{HS}}}, \qquad b_n := \frac{\|\widehat{\Gamma}_n - \Gamma\|_{\mathrm{HS}}}{\|\Gamma\|_{\mathrm{HS}}}.$$

By (H.9), $a_n = o_{\mathbb{P}_X}(1)$ and $b_n = o_{\mathbb{P}_X}(1)$. On $\{b_n > 0\}$, apply Lemma 22 conditionally with

$$t_n := \|\Gamma\|_{\mathrm{HS}}\sqrt{b_n}.$$

Then

$$\frac{|\operatorname{tr}(\widehat{\Gamma}_n - \Gamma)| + t_n}{\|\Gamma\|_{\mathrm{HS}}} = a_n + \sqrt{b_n} = o_{\mathbb{P}_X}(1),$$

while

$$\frac{t_n^2}{\|\widehat{\Gamma}_n - \Gamma\|_{\mathrm{HS}}^2} = \frac{1}{b_n}, \qquad \frac{t_n}{\|\widehat{\Gamma}_n - \Gamma\|_{\mathrm{op}}} \geq \frac{1}{\sqrt{b_n}}.$$

On $\{b_n = 0\}$ the two covariance operators coincide. Therefore,

$$d_K^*\left(\|\mathsf{G}_{\widehat{\Gamma}_n}\|_{\mathrm{F}}^2, \|\mathsf{G}\|_{\mathrm{F}}^2\right) \xrightarrow{\mathbb{P}_X} 0.$$

Together with (H.11), this gives

$$d_K^*\left(\|\Xi_n^{\mathsf{bs}}\|_{\mathrm{F}}^2, \|\mathsf{G}\|_{\mathrm{F}}^2\right) \xrightarrow{\mathbb{P}_X} 0. \tag{H.12}$$

Step 2: Conditional Gaussian approximation of $\eta^{-1}\|\mathcal{T}_n^{\mathsf{bs}} - \mathcal{T}_n\|_{\mathrm{F}}^2$. On $\mathcal{G}_n^{\mathrm{lin}}$, Corollary 11 gives

$$\mathbb{P}^*\left\{\|\mathsf{E}_n^{\mathsf{bs}}\| > \eta^{-1/2} r_n^{\mathrm{lin}}\right\} \leq n^{-9}.$$

Applying Lemma 24 under $\mathbb{P}^*$ with

$$\mathsf{A}_n^{\mathsf{bs}} := \Xi_n^{\mathsf{bs}}, \qquad \mathsf{B}_n^{\mathsf{bs}} := \eta^{-1/2}(\mathcal{T}_n^{\mathsf{bs}} - \mathcal{T}_n), \qquad \mathsf{E}_n^{\mathsf{bs}} := \mathsf{B}_n^{\mathsf{bs}} - \mathsf{A}_n^{\mathsf{bs}}, \qquad \mathsf{G}_n := \mathsf{G}$$

and

$$e_n := \sqrt{r}\eta^{-1/2} r_n^{\mathrm{lin}}, \qquad L_n^2 := C_L(\operatorname{tr}\Gamma)\log n, \qquad h_n := 2L_n e_n + e_n^2,$$

where $C_L > 0$ is a sufficiently large constant so that $\mathbb{P}\{\|\mathsf{G}\|_{\mathrm{F}} > L_n\} \leq n^{-9}$ (see Lemma 21), gives

$$\begin{aligned} d_K^*\left(\|\eta^{-1/2}(\mathcal{T}_n^{\mathsf{bs}} - \mathcal{T}_n)\|_{\mathrm{F}}^2, \|\mathsf{G}\|_{\mathrm{F}}^2\right) &\leq 2d_K^*\left(\|\Xi_n^{\mathsf{bs}}\|_{\mathrm{F}}^2, \|\mathsf{G}\|_{\mathrm{F}}^2\right) + w_n(h_n) + \mathbb{P}^*\{\|\mathsf{E}_n^{\mathsf{bs}}\|_{\mathrm{F}} > e_n\} \\ &\quad + \mathbb{P}^*\{\|\mathsf{G}\|_{\mathrm{F}} > L_n\}. \end{aligned}$$

We bound each term on the right-hand side.

- (H.12) gives that $d_K^*\left(\|\Xi_n^{\mathsf{bs}}\|_{\mathrm{F}}^2, \|\mathsf{G}\|_{\mathrm{F}}^2\right) \xrightarrow{\mathbb{P}_X} 0$.
- Lemma 20 gives

  $$w_n(h_n) = \sup_{t\in\mathbb{R}} \mathbb{P}\{t < \|\mathsf{G}\|_{\mathrm{F}}^2 \leq t + h_n\} \lesssim \sqrt{\frac{h_n}{\|\Gamma\|_{\mathrm{HS}}}} = \sqrt{\frac{e_n(2L_n + e_n)}{\|\Gamma\|_{\mathrm{HS}}}}.$$

  We will show that $e_n L_n/\|\Gamma\|_{\mathrm{HS}} \to 0$ and $e_n^2/\|\Gamma\|_{\mathrm{HS}} \to 0$ in the sequel, which implies that

  $$w_n(h_n) = o(1).$$

  Using the definition of r_n^{lin} and $\tau_*^\dagger \lesssim \tau_*$, we have

  $$e_n \lesssim \Delta_{\min}^{1/2}\kappa_*^2 r^{3/2}\sqrt{\widetilde{\eta}}\,(\log n)\bar{\nu}(1 + \bar{\nu}). \tag{H.13}$$

By Lemma 17(ii) and (vi),

$$\|\Gamma\|_{\mathrm{HS}} \gtrsim \frac{\|\mathfrak{C}\|_{\mathrm{HS}}}{\Delta_{\max}}, \qquad \operatorname{tr}\Gamma \lesssim \frac{\lambda_\times^2}{\Delta_{\min}}.$$

Together with (H.13), $\Delta_{\min}\Delta_{\max} \le \lambda_1^2$, $\lambda_1^2\bar{\nu}^2 \le \lambda_\times^2$, and $\Delta_{\max}\bar{\nu} \le \lambda_\times$, this gives

$$\frac{e_n^2}{\|\Gamma\|_{\mathrm{HS}}} \lesssim \widetilde{\eta}\kappa_*^4 r^3 (\log n)^2 d_{\mathfrak{C}}^{1/2}(1+\bar{\nu}^2) = o(1),$$

and

$$\frac{e_n L_n}{\|\Gamma\|_{\mathrm{HS}}} \lesssim \kappa_*^2 r^{3/2}\sqrt{\widetilde{\eta}}\,(\log n)^{3/2} d_{\mathfrak{C}}^{1/2}(1+\bar{\nu}) = o(1),$$

where the last two relations follow from conditions (15)(i) and (15)(ii).

- Since E_n^{bs} has at most r columns,

$$\mathbb{P}^*\{\|\mathsf{E}_n^{\mathsf{bs}}\|_{\mathrm{F}} > e_n\} \le \mathbb{P}^*\{\|\mathsf{E}_n^{\mathsf{bs}}\|_{\mathrm{op}} > \eta^{-1/2} r_n^{\mathrm{lin}}\} \le n^{-9}.$$

  where the last inequality follows from Corollary 11.
- $\mathbb{P}^*\{\|\mathsf{G}\|_{\mathrm{F}} > L_n\} \le n^{-9}$ by the choice of C_L and Lemma 21.

Therefore,

$$d_K^*\left(\eta^{-1}\|\mathcal{T}_n^{\mathsf{bs}} - \mathcal{T}_n\|_{\mathrm{F}}^2, \|\mathsf{G}\|_{\mathrm{F}}^2\right) \xrightarrow{\mathbb{P}_X} 0. \tag{H.14}$$

Step 3: Conditional Gaussian approximation of $\eta^{-1}\|\sin\Theta(U_n^{\mathsf{bs}}, U_n)\|_F^2$. We first gather some useful facts.

- Define $\mathcal{E}_{\tau,n} := \{\|\mathcal{T}_n\| \le 2\tau_*^\dagger,\ \|\mathcal{T}_n^{\mathsf{bs}}\| \le 2\tau_*^\dagger\}$. On $\mathcal{G}_n^{\mathrm{lin}}$, Corollary 11 gives $\mathbb{P}^*(\mathcal{E}_{\tau,n}^c) \le n^{-9}$.
- $\tau_*^\dagger$ can be taken uniformly small. Indeed, $\tau_*^\dagger \lesssim \tau_*$, and the same calculation as in (C.54) gives

$$\tau_*^\dagger \le \frac{C}{\sqrt{C_{\mathsf{bs}}}}\,\frac{1}{n^\epsilon \log n},$$

  where C_{bs} denotes the sufficiently large constant in Lemma 33. By enlarging C_{bs} if necessary, we may therefore assume

$$\tau_*^\dagger \le \frac{1}{6}. \tag{H.15}$$

  In particular, $4(\tau_*^\dagger)^2 < 1$.
- Since $\sigma_{\min}(\mathcal{C}_n)^2 = \lambda_{\min}(\mathcal{C}_n^\top\mathcal{C}_n) = 1 - \|\mathcal{S}_n^\top\mathcal{S}_n\|$, and similarly for $\mathcal{C}_n^{\mathsf{bs}}$, we have, on $\mathcal{E}_{\tau,n}$,

$$\sigma_{\min}(\mathcal{C}_n)^2 \wedge \sigma_{\min}(\mathcal{C}_n^{\mathsf{bs}})^2 \ge 1 - 4(\tau_*^\dagger)^2 > 0.$$

  Moreover,

$$(U_n^{\mathsf{bs}})^\top U_n = (\mathcal{C}_n^{\mathsf{bs}})^\top\left(I_r + (\mathcal{T}_n^{\mathsf{bs}})^\top\mathcal{T}_n\right)\mathcal{C}_n,$$

  where $\|(\mathcal{T}_n^{\mathsf{bs}})^\top\mathcal{T}_n\| \le 4(\tau_*^\dagger)^2 < 1$. Hence $\mathcal{C}_n$, $\mathcal{C}_n^{\mathsf{bs}}$, and $(U_n^{\mathsf{bs}})^\top U_n$ are invertible on $\mathcal{E}_{\tau,n}$, so Lemma 40 applies, and we can define the self-adjoint positive-semidefinite operator $\mathcal{Q}_n : \mathbb{H} \to \mathbb{H}$ on $\mathcal{E}_{\tau,n}$ through

$$\langle M,\ \mathcal{Q}_n M\rangle_{\mathrm{F}} := \left\|\mathcal{C}_n\mathcal{C}_n^\top\mathcal{T}_n^\top M\mathcal{C}_n^{\mathsf{bs}}\right\|_{\mathrm{F}}^2 + \left\|(I_{d-r} - \mathcal{S}_n\mathcal{S}_n^\top)M\mathcal{C}_n^{\mathsf{bs}}\right\|_{\mathrm{F}}^2, \qquad \forall M \in \mathbb{H}.$$

  Lemma 40(ii), with $(U_1, U_2) = (U_n, U_n^{\mathsf{bs}})$, then gives

$$\eta^{-1}\|\sin\Theta(U_n^{\mathsf{bs}}, U_n)\|_{\mathrm{F}}^2 = \left\|\mathcal{Q}_n^{1/2}\eta^{-1/2}(\mathcal{T}_n^{\mathsf{bs}} - \mathcal{T}_n)\right\|_{\mathrm{F}}^2 \qquad \text{on } \mathcal{E}_{\tau,n}. \tag{H.16}$$

  We next bound $\mathcal{Q}_n^{1/2} - \mathcal{I}$. On $\mathcal{E}_{\tau,n}$, for any $M \in \mathbb{H}$,

$$\left\|\mathcal{C}_n\mathcal{C}_n^\top\mathcal{T}_n^\top M\mathcal{C}_n^{\mathsf{bs}}\right\|_{\mathrm{F}}^2 \le 4(\tau_*^\dagger)^2\|M\|_{\mathrm{F}}^2.$$

Also, $\sigma_{\min}(I_{d-r} - \mathcal{S}_n\mathcal{S}_n^\top) \ge 1 - \|\mathcal{S}_n\|^2 \ge 1 - 4(\tau_*^\dagger)^2$ and $\sigma_{\min}(\mathcal{C}_n^{\mathsf{bs}})^2 \ge 1 - 4(\tau_*^\dagger)^2$. Therefore,
$$(1 - 4(\tau_*^\dagger)^2)^3 \|M\|_{\mathrm{F}}^2 \le \langle M,\, \mathcal{Q}_n M\rangle_{\mathrm{F}} \le (1 + 4(\tau_*^\dagger)^2)\|M\|_{\mathrm{F}}^2.$$
Since $(1-4x)^3 \ge 1 - 12x$ for $x \in [0, 1/4]$, setting $e_n := 12(\tau_*^\dagger)^2$ gives
$$(1 - e_n)\mathcal{I} \preceq \mathcal{Q}_n \preceq (1 + e_n)\mathcal{I},$$
where by (H.15), $e_n \le \frac{1}{3} < \frac{1}{2}$. Thus, on $\mathcal{E}_{\tau,n}$,
$$\|\mathcal{Q}_n^{1/2} - \mathcal{I}\|_{\mathrm{op}} \le \max\{1 - (1 - e_n)^{1/2}, (1 + e_n)^{1/2} - 1\} \le e_n. \tag{H.17}$$

Note that $\mathcal{Q}_n$ is only defined on $\mathcal{E}_{\tau,n}$. Define the random operator $\widetilde{\mathcal{Q}}_n : \mathbb{H} \to \mathbb{H}$ by
$$\widetilde{\mathcal{Q}}_n := \begin{cases} \mathcal{Q}_n, & \text{on } \mathcal{E}_{\tau,n}, \\ \mathcal{I}, & \text{on } \mathcal{E}_{\tau,n}^c. \end{cases}$$

Applying Lemma 25 under $\mathbb{P}^*$ with
$$\mathsf{A}_n^{\mathsf{bs}} := \eta^{-1/2}(\mathcal{T}_n^{\mathsf{bs}} - \mathcal{T}_n), \qquad \mathsf{B}_n^{\mathsf{bs}} := \widetilde{\mathcal{Q}}_n^{1/2}\mathsf{A}_n^{\mathsf{bs}}, \qquad \mathsf{E}_n^{\mathsf{bs}} := \widetilde{\mathcal{Q}}_n^{1/2} - \mathcal{I}, \qquad \mathsf{G}_n := \mathsf{G},$$
and
$$e_n = 12(\tau_*^\dagger)^2, \qquad L_n^2 := C_L(\operatorname{tr}\Gamma)\log n, \qquad h_n := \left(\frac{1}{(1-e_n)^2} - 1\right) L_n^2,$$
where $C_L > 0$ is sufficiently large so that $\mathbb{P}\{\|\mathsf{G}\|_{\mathrm{F}} > L_n\} \le n^{-9}$, gives
$$d_K^*\left(\|\mathsf{B}_n^{\mathsf{bs}}\|_{\mathrm{F}}^2, \|\mathsf{G}\|_{\mathrm{F}}^2\right) \le d_K^*\left(\eta^{-1}\|\mathcal{T}_n^{\mathsf{bs}} - \mathcal{T}_n\|_{\mathrm{F}}^2, \|\mathsf{G}\|_{\mathrm{F}}^2\right) + w_n(h_n) + \mathbb{P}^*\left\{\|\mathsf{E}_n^{\mathsf{bs}}\|_{\mathrm{op}} > e_n\right\} + \mathbb{P}^*\{\|\mathsf{G}\|_{\mathrm{F}} > L_n\}.$$
We consider each term as below.

- By (H.16) and $d_K^* \le 1$,
$$d_K^*\left(\eta^{-1}\|\sin\Theta(U_n^{\mathsf{bs}}, U_n)\|_{\mathrm{F}}^2, \|\mathsf{B}_n^{\mathsf{bs}}\|_{\mathrm{F}}^2\right) \le \mathbb{P}^*(\mathcal{E}_{\tau,n}^c), \qquad \mathbb{P}_X \text{ almost surely.}$$
Thus, on $\mathcal{G}_n^{\mathrm{lin}}$,
$$d_K^*\left(\eta^{-1}\|\sin\Theta(U_n^{\mathsf{bs}}, U_n)\|_{\mathrm{F}}^2, \|\mathsf{B}_n^{\mathsf{bs}}\|_{\mathrm{F}}^2\right) \le n^{-9}.$$
- (H.14) gives $d_K^*\left(\eta^{-1}\|\mathcal{T}_n^{\mathsf{bs}} - \mathcal{T}_n\|_{\mathrm{F}}^2, \|\mathsf{G}\|_{\mathrm{F}}^2\right) \xrightarrow{\mathbb{P}_X} 0$.
- Since $e_n < 1/2$, we have $h_n \lesssim e_n L_n^2 \lesssim (\tau_*^\dagger)^2 L_n^2$. Thus Lemma 20 gives
$$w_n(h_n) = \sup_{t\in\mathbb{R}} \mathbb{P}\{t < \|G\|_{\mathrm{F}}^2 \le t + h_n\} \lesssim \sqrt{\frac{h_n}{\|\Gamma\|_{\mathrm{HS}}}} \lesssim \tau_*^\dagger\sqrt{\frac{L_n^2}{\|\Gamma\|_{\mathrm{HS}}}}.$$
By Lemma 17(ii) and (vi),
$$\frac{\operatorname{tr}\Gamma}{\|\Gamma\|_{\mathrm{HS}}} \lesssim \kappa_\Delta\sqrt{d_{\mathcal{C}}},$$
which combined with $\tau_*^\dagger \lesssim \tau_*$ gives
$$\frac{(\tau_*^\dagger)^2 L_n^2}{\|\Gamma\|_{\mathrm{HS}}} \lesssim \widetilde{\eta}\kappa_*^2\kappa_\Delta r\bar{\nu}^2 d_{\mathfrak{C}}^{1/2}(\log n)^2 = o(1),$$
where the last relation follows from condition (15)(ii). Consequently, $w_n(h_n) = o(1)$.
- (H.17) implies $\mathbb{P}_X$ almost surely, $\mathbb{P}^*\left\{\|\mathsf{E}_n^{\mathsf{bs}}\|_{\mathrm{op}} > e_n\right\} = 0$.
- $\mathbb{P}^*\{\|\mathsf{G}\|_{\mathrm{F}} > L_n\} \le n^{-9}$ by the choice of C_L and Lemma 21.

Combining the above bounds gives
$$d_K^*\left(\eta^{-1}\|\sin\Theta(U_n^{\mathsf{bs}}, U_n)\|_{\mathrm{F}}^2, \|\mathsf{G}\|_{\mathrm{F}}^2\right) \xrightarrow{\mathbb{P}_X} 0.$$

Finally, Theorem 5 gives
$$d_K\left(\eta^{-1}\|\sin\Theta(U_n, U_*)\|_{\mathrm{F}}^2, \|\mathsf{G}\|_{\mathrm{F}}^2\right) \to 0.$$
Triangle inequality then proves Theorem 7.

H.3 Proof of Theorem 8

For notation simplicity, write

$$U := U_n, \qquad U^{\mathsf{bs}} := U_n^{\mathsf{bs}}, \qquad \mathcal{T} := \mathcal{T}_n, \qquad \mathcal{T}^{\mathsf{bs}} := \mathcal{T}_n^{\mathsf{bs}},$$

and define

$$R := \operatorname{sgn}(U^\top U_*), \qquad R^{\mathsf{bs}} := \operatorname{sgn}((U^{\mathsf{bs}})^\top U_*), \qquad Q^{\mathsf{bs}} := \operatorname{sgn}((U^{\mathsf{bs}})^\top U).$$

Set

$$V_1 := UR, \qquad V_2 := U^{\mathsf{bs}} R^{\mathsf{bs}}.$$

Thus V_1 and V_2 are both Procrustes-aligned with U_*.

We apply Lemma 40 with $(U_1, U_2) = (V_1, V_2)$. Note the following facts.

- Let $M_{21} := V_2^\top V_1 = (R^{\mathsf{bs}})^\top (U^{\mathsf{bs}})^\top U R$. By orthogonal equivariance of the polar factor,
$$\operatorname{sgn}(M_{21}) = (R^{\mathsf{bs}})^\top \operatorname{sgn}(U^{\mathsf{bs}\top} U) R = (R^{\mathsf{bs}})^\top Q^{\mathsf{bs}} R.$$
Hence,
$$V_2 \operatorname{sgn}(M_{21}) - V_1 = U^{\mathsf{bs}} R^{\mathsf{bs}} (R^{\mathsf{bs}})^\top Q^{\mathsf{bs}} R - UR = (U^{\mathsf{bs}} Q^{\mathsf{bs}} - U)R.$$
- Let $\mathcal{C}_1, \mathcal{S}_1$ and $\mathcal{C}_2, \mathcal{S}_2$ denote the cosine and sine matrices of V_1 and V_2 relative to U_*. Right multiplication by an orthogonal matrix does not change the tangent matrix. Hence the tangent matrices of V_1 and V_2 are respectively $\mathcal{T}$ and $\mathcal{T}^{\mathsf{bs}}$.

Then Lemma 40(i) gives

$$(U^{\mathsf{bs}} Q^{\mathsf{bs}} - U)R = U_* A + U_{*,\perp} B, \tag{H.18}$$

where

$$\begin{aligned} A &= \mathcal{C}_1 \left[(M_{21}^\top M_{21})^{1/2} - I_r \right] - \mathcal{C}_1 \mathcal{C}_1^\top \mathcal{T}^\top (\mathcal{T}^{\mathsf{bs}} - \mathcal{T}) \mathcal{C}_2 \operatorname{sgn}(M_{21}), \\ B &= \mathcal{S}_1 \left[(M_{21}^\top M_{21})^{1/2} - I_r \right] + (I_{d-r} - \mathcal{S}_1 \mathcal{S}_1^\top)(\mathcal{T}^{\mathsf{bs}} - \mathcal{T}) \mathcal{C}_2 \operatorname{sgn}(M_{21}). \end{aligned}$$

We provide some intuition before we proceed. Recall from Appendix H.2 that $\mathcal{T}^{\mathsf{bs}} - \mathcal{T} \approx \sqrt{\eta} \sum_{i=1}^n \xi_i^{\mathsf{bs}}$, where

$$\xi_i^{\mathsf{bs}} := \sqrt{\eta} (w_i^{\mathsf{bs}} - 1)(\mathcal{I} - \eta \mathcal{L})^{n-i} (Z_i Y_i^\top) \in \mathbb{H}, \qquad i \in [n].$$

Meanwhile, since V_1 and V_2 are already aligned with U_*, $\mathcal{C}_1 \approx \mathcal{C}_2 \approx \operatorname{sgn}(M_{21}) \approx I_r$. These observations motivate the following decomposition of $(U^{\mathsf{bs}} Q^{\mathsf{bs}} - U)R$:

$$(U^{\mathsf{bs}} Q^{\mathsf{bs}} - U)R = W^{\mathsf{bs}} + \Psi^{\mathsf{bs}}, \tag{H.19}$$

where

$$\begin{aligned} W^{\mathsf{bs}} &= U_{*,\perp} \sqrt{\eta} \sum_{i=1}^n \xi_i^{\mathsf{bs}}, \\ \Psi^{\mathsf{bs}} &= U_* A + U_{*,\perp} [B - (\mathcal{T}^{\mathsf{bs}} - \mathcal{T})] + U_{*,\perp} E^{\mathsf{bs}}, \qquad E^{\mathsf{bs}} := (\mathcal{T}^{\mathsf{bs}} - \mathcal{T}) - \sqrt{\eta} \sum_{i=1}^n \xi_i^{\mathsf{bs}}. \end{aligned}$$

Fix $m \in [d]$ and define

$$u := u_m := U_{*,\perp}^\top e_m, \qquad \widetilde{u} := \Lambda_{*,\perp}^{1/2} u, \qquad \Sigma_{U,m}^* := \eta \Gamma_u.$$

Identifying the m-th row with a column vector as in the proof of Theorem 6, (H.19) gives

$$\left[(U^{\mathsf{bs}} Q^{\mathsf{bs}} - U) R \right]_{m,\cdot}^\top = W_u^{\mathsf{bs}} + \Psi_u^{\mathsf{bs}}, \tag{H.20}$$

where

$$\begin{aligned} W_u^{\mathsf{bs}} &:= \sqrt{\eta} \sum_{i=1}^n (\xi_i^{\mathsf{bs}})^\top u, \\ \Psi_u^{\mathsf{bs}} &:= A^\top U_*^\top e_m + [B - (\mathcal{T}^{\mathsf{bs}} - \mathcal{T})]^\top u + (E^{\mathsf{bs}})^\top u. \end{aligned}$$

Step 1: Gaussian approximation of W_u^{bs}. Define the projected bootstrap summands

$$\xi_{u,i}^{\mathsf{bs}} := (\xi_i^{\mathsf{bs}})^\top u = \sqrt{\eta}(w_i^{\mathsf{bs}} - 1) \left[(\mathcal{I} - \eta\mathcal{L})^{n-i}(Z_i Y_i^\top)\right]^\top u \in \mathbb{R}^r.$$

Then

$$W_u^{\mathsf{bs}} = \sqrt{\eta} \sum_{i=1}^n \xi_{u,i}^{\mathsf{bs}}.$$

Conditionally on $\mathcal{F}_X$, the variables $\{\xi_{u,i}^{\mathsf{bs}}\}_{i=1}^n$ are independent and centered. Define their conditional cumulative covariance by

$$\widehat{\Gamma}_{u,n} := \sum_{i=1}^n \operatorname{cov}^*(\xi_{u,i}^{\mathsf{bs}}).$$

Then $\operatorname{cov}^*(W_u^{\mathsf{bs}}) = \eta\widehat{\Gamma}_{u,n}$. Since $\mathbb{E}(w_i^{\mathsf{bs}} - 1)^2 = 1$, we have $\mathbb{E}_X \widehat{\Gamma}_{u,n} = \Gamma_{u,n}$ and

$$\widehat{\Gamma}_{u,n} = \eta \sum_{i=1}^n \left[(\mathcal{I} - \eta\mathcal{L})^{n-i}(Z_i Y_i^\top)\right]^\top u u^\top \left[(\mathcal{I} - \eta\mathcal{L})^{n-i}(Z_i Y_i^\top)\right].$$

We first compare $\widehat{\Gamma}_{u,n}$ with Γ_u. For $p \in \{3, 4\}$, we claim that

$$\sum_{i=1}^n \left\| \left[(\mathcal{I} - \eta\mathcal{L})^{n-i}(Z_i Y_i^\top)\right]^\top u \right\|^p = O_{\mathbb{P}_X}\left(\frac{r^{p/2}\lambda_1^{p/2}\|\widetilde{u}\|^p}{\widetilde{\eta}} \right).$$

Indeed, for each $j \in [r]$, let

$$D_j := \operatorname{diag}\{1 - \eta\Delta_{\ell j}\}_{\ell \in [d-r]}.$$

The j-th coordinate of the vector inside the norm is $[Y_i]_j u^\top D_j^{n-i} Z_i$. By the power-mean inequality, Cauchy–Schwarz, and Lemma 37,

$$\begin{aligned}
\mathbb{E}\left|[Y_i]_j\, u^\top D_j^{n-i} Z_i\right|^p &\le \left(\mathbb{E}|[Y_i]_j|^{2p}\right)^{1/2} \left(\mathbb{E}|u^\top D_j^{n-i} Z_i|^{2p}\right)^{1/2} \\
&\lesssim \lambda_j^{p/2} \left\| \Lambda_{*,\perp}^{1/2} D_j^{n-i} u \right\|^p \\
&\le \lambda_1^{p/2} (1 - \widetilde{\eta})^{p(n-i)} \|\widetilde{u}\|^p.
\end{aligned}$$

Therefore,

$$\mathbb{E}\left\| \left[(\mathcal{I} - \eta\mathcal{L})^{n-i}(Z_i Y_i^\top)\right]^\top u \right\|^p \lesssim r^{p/2}\lambda_1^{p/2}(1 - \widetilde{\eta})^{p(n-i)} \|\widetilde{u}\|^p.$$

Summing the geometric series and applying Markov's inequality proves the claim.

Using the $p = 4$ case and independence gives

$$\mathbb{E}_X \|\widehat{\Gamma}_{u,n} - \Gamma_{u,n}\|_{\mathrm{F}}^2 \lesssim \eta^2 \sum_{i=1}^n \mathbb{E}\left\| \left[(\mathcal{I} - \eta\mathcal{L})^{n-i}(Z_i Y_i^\top)\right]^\top u \right\|^4 \lesssim \widetilde{\eta}\kappa_*^2 r^2 \|\widetilde{u}\|^4.$$

Hence

$$\|\widehat{\Gamma}_{u,n} - \Gamma_{u,n}\|_{\mathrm{F}} = O_{\mathbb{P}_X}\left(\sqrt{\widetilde{\eta}}\, \kappa_* r \|\widetilde{u}\|^2 \right).$$

By Lemma 18(iii),

$$\lambda_{\min}(\Gamma_u) \gtrsim \frac{\lambda_r \|\widetilde{u}\|^2}{\Delta_{\max}} \lambda_{\min}(\widetilde{\mathfrak{C}}).$$

Combining this with the approximation in Lemma 18(iv), as in Step 2 of the proof of Lemma 30, gives

$$\left\| \Gamma_u^{-1/2}(\widehat{\Gamma}_{u,n} - \Gamma_u)\Gamma_u^{-1/2} \right\|_{\mathrm{F}} = O_{\mathbb{P}_X}\left(\sqrt{\widetilde{\eta}}\, \kappa_*^2 r \lambda_{\min}^{-1}(\widetilde{\mathfrak{C}}) + \widetilde{\eta}\kappa_*^5 r^{5/2} \lambda_{\min}^{-2}(\widetilde{\mathfrak{C}}) \right).$$

The first condition in Theorem 6, $\widetilde{\eta}\kappa_*^6 r^{7/2}\lambda_{\min}^{-3}(\widetilde{\mathfrak{C}}) \to 0$, implies that the right-hand side is $o_{\mathbb{P}_X}(1)$. Thus, with $\mathbb{P}_X$-probability tending to one, $\widehat{\Gamma}_{u,n}$ is positive definite and

$$\lambda_{\min}(\widehat{\Gamma}_{u,n}) \gtrsim \lambda_{\min}(\Gamma_u).$$

We now apply Theorem 13 conditionally on $\mathcal{F}_X$. The corresponding Berry–Esseen quantity satisfies

$$\begin{aligned}\gamma_u^{\mathsf{bs}} &:= \sum_{i=1}^n \mathbb{E}^* \left\|(\eta\widehat{\Gamma}_{u,n})^{-1/2}\sqrt{\eta}\,\xi_{u,i}^{\mathsf{bs}}\right\|^3 \\ &= \sum_{i=1}^n \mathbb{E}^* \left\|\widehat{\Gamma}_{u,n}^{-1/2}\xi_{u,i}^{\mathsf{bs}}\right\|^3 \\ &= O_{\mathbb{P}_X}\left(\sqrt{\widetilde{\eta}}\,\kappa_*^3 r^{3/2}\lambda_{\min}^{-3/2}(\widetilde{\mathfrak{C}})\right),\end{aligned}$$

where the last line follows from Assumption 2, the $p=3$ empirical moment bound above, and the lower bound for $\lambda_{\min}(\widehat{\Gamma}_{u,n})$. Therefore, Theorem 13 yields

$$\sup_{A\in\mathcal{A}_r}\left|\mathbb{P}^*\{W_u^{\mathsf{bs}}\in A\} - N(0,\eta\widehat{\Gamma}_{u,n})\{A\}\right| = O_{\mathbb{P}_X}\left(\sqrt{\widetilde{\eta}}\,\kappa_*^3 r^{7/4}\lambda_{\min}^{-3/2}(\widetilde{\mathfrak{C}})\right) = o_{\mathbb{P}_X}(1).$$

The last convergence follows from the first condition in Theorem 6.

Moreover, the preceding relative covariance bound and Lemma 42 give

$$\sup_{A\in\mathcal{A}_r}\left|N(0,\eta\widehat{\Gamma}_{u,n})\{A\} - N(0,\eta\Gamma_u)\{A\}\right| = o_{\mathbb{P}_X}(1).$$

Consequently,

$$\sup_{A\in\mathcal{A}_r}\left|\mathbb{P}^*\{W_u^{\mathsf{bs}}\in A\} - N(0,\Sigma_{U,m}^*)\{A\}\right| \xrightarrow{\mathbb{P}_X} 0.$$

Step 2: Accounting for the higher-order term Ψ_u^{bs}. Under the assumptions of Theorem 8, the conditions of Proposition 3 hold with $\delta_p \to 0$. As in the proof of Theorem 7, condition (11) implies Assumption 4, while the Haar initialization satisfies Assumption 5 with probability $1-o(1)$. Thus, after intersecting with this initialization event, Corollary 11 gives an $\mathcal{F}_X$-measurable event $\mathcal{G}_n^{\mathrm{lin}}$ such that $\mathbb{P}_X(\mathcal{G}_n^{\mathrm{lin}})\to 1$ and, on $\mathcal{G}_n^{\mathrm{lin}}$,

$$\|\mathcal{T}\| \le 2\tau_*^\dagger, \qquad \mathbb{P}^*\{\|\mathcal{T}^{\mathsf{bs}}\| > 2\tau_*^\dagger\} \le n^{-9}, \qquad \mathbb{P}^*\{\|E^{\mathsf{bs}}\| > r_n^{\mathrm{lin}}\} \le n^{-9},$$

where

$$\tau_*^\dagger \lesssim \tau_* \asymp \kappa_*\bar{\nu}\sqrt{r\widetilde{\eta}\log n}, \qquad r_n^{\mathrm{lin}} \lesssim \tau_*^\dagger\left[C_\psi^2\kappa_*\sqrt{r\widetilde{\eta}\log n} + \tau_*^\dagger\right].$$

Define

$$\mathcal{E}_{\tau,n} := \{\|\mathcal{T}\| \le 2\tau_*^\dagger,\ \|\mathcal{T}^{\mathsf{bs}}\| \le 2\tau_*^\dagger\}.$$

On $\mathcal{G}_n^{\mathrm{lin}}$, Corollary 11 gives $\mathbb{P}^*(\mathcal{E}_{\tau,n}^c) \le n^{-9}$.

To control A and $B - (\mathcal{T}^{\mathsf{bs}} - \mathcal{T})$ defined in (H.18) on $\mathcal{E}_{\tau,n}$, we use the following bounds.

- Since V_1 and V_2 are Procrustes-aligned with U_*, their cosine matrices $\mathcal{C}_1$ and $\mathcal{C}_2$ are symmetric positive definite. Using $\mathcal{S}_i = \mathcal{T}_i\mathcal{C}_i$ and $\mathcal{C}_i^\top\mathcal{C}_i + \mathcal{S}_i^\top\mathcal{S}_i = I_r$ gives

  $$\mathcal{C}_1 = (I_r + \mathcal{T}^\top\mathcal{T})^{-1/2}, \qquad \mathcal{C}_2 = (I_r + (\mathcal{T}^{\mathsf{bs}})^\top\mathcal{T}^{\mathsf{bs}})^{-1/2}.$$

  Therefore,

  $$\|\mathcal{C}_i - I_r\| \lesssim (\tau_*^\dagger)^2, \qquad \|\mathcal{S}_i\| \lesssim \tau_*^\dagger, \qquad i=1,2.$$

  Moreover, since $M_{21} = \mathcal{C}_2^\top\mathcal{C}_1 + \mathcal{S}_2^\top\mathcal{S}_1$, we have

  $$\|M_{21} - I_r\| \lesssim (\tau_*^\dagger)^2.$$

- Lemma 40(iii) gives

  $$\|(M_{21}^\top M_{21})^{1/2} - I_r\| \le \|\sin\Theta(V_2,V_1)\|^2 \le (1+\|\mathcal{T}\|^2)\|\Delta\|^2.$$

  On $\mathcal{E}_{\tau,n}$,

  $$\|\mathcal{T}^{\mathsf{bs}} - \mathcal{T}\| \le \|\mathcal{T}\| + \|\mathcal{T}^{\mathsf{bs}}\| \le 4\tau_*^\dagger,$$

  and hence

  $$\|(M_{21}^\top M_{21})^{1/2} - I_r\| \lesssim (\tau_*^\dagger)^2.$$

- Since $\operatorname{sgn}(M_{21}) = M_{21}(M_{21}^\top M_{21})^{-1/2}$, we have
$$\|\operatorname{sgn}(M_{21}) - I_r\| = \|(M_{21} - I_r)(M_{21}^\top M_{21})^{-1/2} + (M_{21}^\top M_{21})^{-1/2} - I_r\| \lesssim (\tau_*^\dagger)^2$$
and consequently
$$\|\mathcal{C}_2 \operatorname{sgn}(M_{21}) - I_r\| \le \|(\mathcal{C}_2 - I_r)\operatorname{sgn}(M_{21})\| + \|\operatorname{sgn}(M_{21}) - I_r\| \lesssim (\tau_*^\dagger)^2.$$

It follows from the above bounds and the definition in (H.18) that
$$\begin{aligned}
\|A\| &\lesssim (\tau_*^\dagger)^2 + \|\mathcal{T}\|\,\|\Delta\| \lesssim (\tau_*^\dagger)^2;\\
\|B - (\mathcal{T}^{\mathsf{bs}} - \mathcal{T})\| &= \|\mathcal{S}_1\big[(M_{21}^\top M_{21})^{1/2} - I_r\big]\| + \|\mathcal{S}_1\mathcal{S}_1^\top(\mathcal{T}^{\mathsf{bs}} - \mathcal{T})\mathcal{C}_2 \operatorname{sgn}(M_{21})\|\\
&\quad + \|(\mathcal{T}^{\mathsf{bs}} - \mathcal{T})(\mathcal{C}_2 \operatorname{sgn}(M_{21}) - I_r)\|\\
&\lesssim (\tau_*^\dagger)^3.
\end{aligned}$$

Recall
$$\Psi_u^{\mathsf{bs}} = A^\top U_*^\top e_m + [B - (\mathcal{T}^{\mathsf{bs}} - \mathcal{T})]^\top u + (E^{\mathsf{bs}})^\top u.$$
Hence, on $\mathcal{G}_n^{\text{lin}}$, with conditional probability at least $1 - O(n^{-9})$,
$$\|\Psi_u^{\mathsf{bs}}\| \lesssim (\tau_*^\dagger)^2\|U_*^\top e_m\| + (\tau_*^\dagger)^3\|u\| + r_n^{\text{lin}}\|u\| \lesssim \tau_*^2\|U_*^\top e_m\| + r_n^{\text{lin}}\|u\|.$$
Here the last line uses $\tau_*^\dagger \lesssim \tau_*$, $\tau_*^\dagger = o(1)$, and the fact that the definition of r_n^{lin} contains a term of order $(\tau_*^\dagger)^2$.

By Lemma 18(iii),
$$\lambda_{\min}(\Sigma_{U,m}^*) = \eta\lambda_{\min}(\Gamma_u) \gtrsim \eta\frac{\lambda_r\|\widetilde{u}\|^2}{\Delta_{\max}}\lambda_{\min}(\widetilde{\mathfrak{C}}) \gtrsim \widetilde{\eta}\frac{\|\widetilde{u}\|^2}{\lambda_1}\lambda_{\min}(\widetilde{\mathfrak{C}}).$$

Consequently,
$$\mathbb{P}^*\big\{\big\|(\Sigma_{U,m}^*)^{-1/2}\Psi_u^{\mathsf{bs}}\big\| \le \zeta^{\mathsf{bs}}\big\} \ge 1 - O(n^{-9})$$
on $\mathcal{G}_n^{\text{lin}}$, where $\zeta^{\mathsf{bs}} := C(\zeta_1 + \zeta_2)$ for a sufficiently large constant C, with
$$\begin{aligned}
\zeta_1 &:= \kappa_*^2\widetilde{\eta}^{1/2} r(\log n)\bar{\nu}^2\frac{\lambda_1^{1/2}\|U_*^\top e_m\|}{\|\widetilde{u}\|}\lambda_{\min}^{-1/2}(\widetilde{\mathcal{C}}),\\
\zeta_2 &:= \kappa_*^2\widetilde{\eta}^{1/2} r(\log n)\bar{\nu}(1+\bar{\nu})\frac{\lambda_1^{1/2}\|u\|}{\|\widetilde{u}\|}\lambda_{\min}^{-1/2}(\widetilde{\mathfrak{C}}).
\end{aligned}$$

These are of the same orders as ζ_1 and ζ_2 in Lemma 31. In particular, condition (17) gives
$$r^{1/4}\zeta^{\mathsf{bs}} \to 0.$$

Indeed,
$$\begin{aligned}
r^{1/2}\zeta_1^2 &\lesssim \widetilde{\eta}\kappa_*^4 r^{5/2}(\log n)^2\bar{\nu}^2(1+\bar{\nu})^2\frac{\lambda_1\|U_*^\top e_m\|^2}{\|\widetilde{u}\|^2}\lambda_{\min}^{-1}(\widetilde{\mathfrak{C}}),\\
r^{1/2}\zeta_2^2 &\lesssim \widetilde{\eta}\kappa_*^4 r^{5/2}(\log n)^2\bar{\nu}^2(1+\bar{\nu})^2\frac{\lambda_1\|u\|^2}{\|\widetilde{u}\|^2}\lambda_{\min}^{-1}(\widetilde{\mathfrak{C}}),
\end{aligned}$$
where for the first bound we used $\bar{\nu}^4 \le \bar{\nu}^2(1+\bar{\nu})^2$. Both terms vanish by condition (17).

We now proceed as in the proof of Theorem 6. For any $A \in \mathcal{A}_r$, let A^ζ denote the signed enlargement defined before Theorem 14. The decomposition
$$\big[(U^{\mathsf{bs}}Q^{\mathsf{bs}} - U)R\big]_{m,\cdot}^\top = W_u^{\mathsf{bs}} + \Psi_u^{\mathsf{bs}}$$
and the preceding remainder bound imply
$$\begin{aligned}
\mathbb{P}^*\big\{(\Sigma_{U,m}^*)^{-1/2}W_u^{\mathsf{bs}} \in A^{-\zeta^{\mathsf{bs}}}\big\} - O(n^{-9}) &\le \mathbb{P}^*\big\{(\Sigma_{U,m}^*)^{-1/2}\big[(U^{\mathsf{bs}}Q^{\mathsf{bs}} - U)R\big]_{m,\cdot}^\top \in A\big\}\\
&\le \mathbb{P}^*\big\{(\Sigma_{U,m}^*)^{-1/2}W_u^{\mathsf{bs}} \in A^{\zeta^{\mathsf{bs}}}\big\} + O(n^{-9}).
\end{aligned}$$

By Step 1,

$$\sup_{A\in\mathcal{A}_r}\left|\mathbb{P}^*\left\{(\Sigma^*_{U,m})^{-1/2}W_u^{\mathsf{bs}}\in A\right\}-\mathbb{P}\{N(0,I_r)\in A\}\right|\xrightarrow{\mathbb{P}_X}0.$$

Hence, by Theorem 14,

$$\begin{aligned}\mathbb{P}^*\left\{(\Sigma^*_{U,m})^{-1/2}W_u^{\mathsf{bs}}\in A^{\zeta^{\mathsf{bs}}}\right\}&\le\mathbb{P}\{N(0,I_r)\in A^{\zeta^{\mathsf{bs}}}\}+o_{\mathbb{P}_X}(1)\\&\le\mathbb{P}\{N(0,I_r)\in A\}+(0.59r^{1/4}+0.21)\zeta^{\mathsf{bs}}+o_{\mathbb{P}_X}(1)\\&=\mathbb{P}\{N(0,I_r)\in A\}+o_{\mathbb{P}_X}(1).\end{aligned}$$

The corresponding lower bound follows in the same way using $A^{-\zeta^{\mathsf{bs}}}$. Therefore,

$$\sup_{A\in\mathcal{A}_r}\left|\mathbb{P}^*\left\{\left[(U^{\mathsf{bs}}Q^{\mathsf{bs}}-U)R\right]_{m,\cdot}\in A\right\}-N(0,\Sigma^*_{U,m})\{A\}\right|\xrightarrow{\mathbb{P}_X}0.$$

Finally, Theorem 6 gives

$$\sup_{A\in\mathcal{A}_r}\left|\mathbb{P}\{[UR-U_*]_{m,\cdot}\in A\}-N(0,\Sigma^*_{U,m})\{A\}\right|\to 0.$$

Combining the last two displays by the triangle inequality yields

$$\sup_{A\in\mathcal{A}_r}\left|\mathbb{P}^*\left\{\left[(U^{\mathsf{bs}}Q^{\mathsf{bs}}-U)R\right]_{m,\cdot}\in A\right\}-\mathbb{P}\{[UR-U_*]_{m,\cdot}\in A\}\right|\xrightarrow{\mathbb{P}_X}0.$$

Recalling

$$U=U_n,\qquad U^{\mathsf{bs}}=U_n^{\mathsf{bs}},\qquad R=\operatorname{sgn}(U_n^\top U_*),\qquad Q^{\mathsf{bs}}=\operatorname{sgn}((U_n^{\mathsf{bs}})^\top U_n),$$

this is exactly the conclusion of Theorem 8.

H.4 Proof of Lemma 34

Proof. Starting from (H.3) and (H.5), the same decomposition as in Lemma 15 gives

$$\overline{\mathcal{T}}_n^{\mathsf{bs}}=(\mathcal{I}-\eta\mathcal{L})^{n-N_*}\overline{\mathcal{T}}_{N_*}^{\mathsf{bs}}+\overline{\mathcal{D}}_{N_*,n}^{\mathsf{bs},(1,0)}+\overline{\mathcal{D}}_{N_*,n}^{\mathsf{bs},(1,1)}+\overline{\mathcal{D}}_{N_*,n}^{\mathsf{bs},(1,2)}+\overline{\mathcal{D}}_{N_*,n}^{\mathsf{bs},(2)}+\overline{\mathcal{R}}_{N_*,n}^{\mathsf{bs},H},$$

where

$$\begin{aligned}\overline{\mathcal{D}}_{N_*,n}^{\mathsf{bs},(1,0)}&:=\eta\sum_{i=N_*+1}^{n}w_i^{\mathsf{bs}}(\mathcal{I}-\eta\mathcal{L})^{n-i}(\overline{Z}_i\overline{Y}_i^\top-\overline{\Lambda}_{*,\times});\\
\overline{\mathcal{D}}_{N_*,n}^{\mathsf{bs},(1,1)}&:=\overline{S}_1^{\mathsf{bs}}+\overline{S}_2^{\mathsf{bs}}+\overline{S}_3^{\mathsf{bs}},\\
\text{where}\quad\overline{S}_1^{\mathsf{bs}}&:=\eta\sum_{i=N_*+1}^{n}w_i^{\mathsf{bs}}(\mathcal{I}-\eta\mathcal{L})^{n-i}[-\overline{\mathcal{T}}_{i-1}^{\mathsf{bs}}(\overline{Y}_i\overline{Y}_i^\top-\overline{\Lambda}_*)],\\
\overline{S}_2^{\mathsf{bs}}&:=\eta\sum_{i=N_*+1}^{n}w_i^{\mathsf{bs}}(\mathcal{I}-\eta\mathcal{L})^{n-i}[(\overline{Z}_i\overline{Z}_i^\top-\overline{\Lambda}_{*,\perp})\overline{\mathcal{T}}_{i-1}^{\mathsf{bs}}],\\
\overline{S}_3^{\mathsf{bs}}&:=\eta\sum_{i=N_*+1}^{n}w_i^{\mathsf{bs}}(\mathcal{I}-\eta\mathcal{L})^{n-i}[-\overline{\mathcal{T}}_{i-1}^{\mathsf{bs}}(\overline{Y}_i\overline{Z}_i^\top-\overline{\Lambda}_{*,\times}^\top)\overline{\mathcal{T}}_{i-1}^{\mathsf{bs}}];\\
\overline{\mathcal{D}}_{N_*,n}^{\mathsf{bs},(1,2)}&:=-\eta\sum_{i=N_*+1}^{n}(w_i^{\mathsf{bs}}-1)(\mathcal{I}-\eta\mathcal{L})^{n-i}(\mathcal{L}\overline{\mathcal{T}}_{i-1}^{\mathsf{bs}});\\
\overline{\mathcal{D}}_{N_*,n}^{\mathsf{bs},(2)}&:=\sum_{i=N_*+1}^{n}(\mathcal{I}-\eta\mathcal{L})^{n-i}\overline{D}_i^{\mathsf{bs},(2)};\\
\overline{\mathcal{R}}_{N_*,n}^{\mathsf{bs},H}&:=\sum_{i=N_*+1}^{n}(\mathcal{I}-\eta\mathcal{L})^{n-i}\overline{R}_i^{\mathsf{bs},H}.\end{aligned}$$

Then

$$
\begin{aligned}
&\Big\|\overline{\mathcal{T}}_n^{\mathsf{bs}} - \eta\sum_{i=1}^{n} w_i^{\mathsf{bs}}(\mathcal{I}-\eta\mathcal{L})^{n-i}(\overline{Z}_i\overline{Y}_i^\top)\Big\| \\
&\le \underbrace{\big\|(\mathcal{I}-\eta\mathcal{L})^{n-N_*}\overline{\mathcal{T}}_{N_*}^{\mathsf{bs}}\big\|}_{=:\overline{\alpha}_1^{\mathsf{bs}}} + \underbrace{\big\|\eta\sum_{i=1}^{N_*} w_i^{\mathsf{bs}}(\mathcal{I}-\eta\mathcal{L})^{n-i}(\overline{Z}_i\overline{Y}_i^\top - \overline{\Lambda}_{*,\times})\big\|}_{=:\overline{\alpha}_2^{\mathsf{bs}}} \\
&+ \underbrace{\big\|\overline{\mathcal{D}}_{N_*,n}^{\mathsf{bs},(1,1)}\big\|}_{=:\overline{\alpha}_3^{\mathsf{bs}}} + \underbrace{\big\|\overline{\mathcal{D}}_{N_*,n}^{\mathsf{bs},(2)}\big\|}_{=:\overline{\alpha}_4^{\mathsf{bs}}} + \underbrace{\big\|\overline{\mathcal{R}}_{N_*,n}^{\mathsf{bs},H} - \eta\sum_{i=N_*+1}^{n} w_i^{\mathsf{bs}}(\mathcal{I}-\eta\mathcal{L})^{n-i}\overline{\Lambda}_{*,\times}\big\|}_{=:\overline{\alpha}_5^{\mathsf{bs}}} \\
&+ \underbrace{\big\|\eta\sum_{i=1}^{N_*} w_i^{bs}(\mathcal{I}-\eta\mathcal{L})^{n-i}\overline{\Lambda}_{*,\times}\big\|}_{=:\overline{\alpha}_6^{\mathsf{bs}}} + \underbrace{\big\|\overline{\mathcal{D}}_{N_*,n}^{\mathsf{bs},(1,2)}\big\|}_{=:\overline{\alpha}_7^{\mathsf{bs}}}.
\end{aligned}
$$

The terms $\overline{\alpha}_1^{\mathsf{bs}},\dots,\overline{\alpha}_6^{\mathsf{bs}}$ are controlled exactly as their counterparts $\alpha_1,\dots,\alpha_6$ in the proof of Lemma 15, up to constants depending only on C_w. In particular, we have

$$
\sum_{l=1}^{6}\overline{\alpha}_l^{\mathsf{bs}} \lesssim \overline{\tau}_*^{\mathsf{bs}}\big[C_\psi^2\kappa_*\sqrt{r\widetilde{\eta}\log n} + \overline{\tau}_*^{\mathsf{bs}}\big] + e^{-\widetilde{\eta}(n-N_*)}\kappa_*\nu_\infty\|H\|. \tag{H.21}
$$

It remains to control the new multiplier term $\overline{\alpha}_7^{\mathsf{bs}}$. Note that $\|\mathcal{L}\overline{\mathcal{T}}_{i-1}^{\mathsf{bs}}\| \le 2\lambda_1\|\overline{\mathcal{T}}_{i-1}^{\mathsf{bs}}\|$. Moreover, similar to (C.54), we have $\overline{\tau}_*^{\mathsf{bs}} \le \frac{1}{n^\varepsilon \log n}$ for sufficiently large constant C_{bs}. Then on event $\{\|\overline{\mathcal{T}}_{i-1}^{\mathsf{bs}}\| \le 2\tau_*^{\mathsf{bs}}\}$, (D.3) and (D.13) give

$$
\begin{aligned}
\|(\overline{\mathcal{L}}-\mathcal{L})\overline{\mathcal{T}}_{i-1}^{\mathsf{bs}}\| &\le \|H\|\,\lambda_1(1+\|\overline{\mathcal{T}}_{i-1}^{\mathsf{bs}}\|\nu_\infty)(\|\overline{\mathcal{T}}_{i-1}^{\mathsf{bs}}\|+\nu_\infty) \\
&\lesssim \|H\|\,\lambda_1(\overline{\tau}_*^{\mathsf{bs}}+\nu_\infty) \\
&\lesssim \lambda_1\overline{\tau}_*^{\mathsf{bs}},
\end{aligned}
$$

where the last inequality uses $\nu_\infty\|H\| \lesssim \nu_\infty\widetilde{\eta} \lesssim \tau_*^{\mathsf{bs}}$. By triangle inequality, on event $\{\|\overline{\mathcal{T}}_{i-1}^{\mathsf{bs}}\| \le 2\tau_*^{\mathsf{bs}}\}$, we have

$$
\|\overline{\mathcal{L}\mathcal{T}}_{i-1}^{\mathsf{bs}}\| \lesssim \lambda_1\overline{\tau}_*^{\mathsf{bs}}.
$$

Define, for $i = N_*+1,\dots,n$,

$$
\Xi_i^{\mathsf{bs}} := -\eta(w_i^{\mathsf{bs}}-1)(\mathcal{I}-\eta\mathcal{L})^{n-i}(\overline{\mathcal{L}\mathcal{T}}_{i-1}^{\mathsf{bs}})\mathbb{1}\big\{\|\overline{\mathcal{T}}_{i-1}^{\mathsf{bs}}\| \le 2\tau_*^{\mathsf{bs}}\big\}.
$$

Since $\overline{\mathcal{T}}_{i-1}^{\mathsf{bs}}$ is $\overline{\mathcal{F}}_{i-1}$-measurable and w_i^{bs} is independent of $\overline{\mathcal{F}}_{i-1}$ with $\mathbb{E}(w_i^{\mathsf{bs}}-1)=0$, $\{\Xi_i^{\mathsf{bs}}\}$ is a matrix martingale-difference sequence. Moreover,

$$
\|\Xi_i^{\mathsf{bs}}\| \lesssim \eta\lambda_1\tau_*^{\mathsf{bs}}(1-\widetilde{\eta})^{n-i}.
$$

Therefore,

$$
\max_{i>N_*}\|\Xi_i^{\mathsf{bs}}\| \lesssim \widetilde{\eta}\kappa_*\tau_*^{\mathsf{bs}}.
$$

Similarly,

$$
\begin{aligned}
&\max\Big\{\Big\|\sum_{i=N_*+1}^{n}\overline{\mathbb{E}}_{i-1}\big[\Xi_i^{\mathsf{bs}}(\Xi_i^{\mathsf{bs}})^\top\big]\Big\|^{1/2}, \Big\|\sum_{i=N_*+1}^{n}\overline{\mathbb{E}}_{i-1}\big[(\Xi_i^{\mathsf{bs}})^\top\Xi_i^{\mathsf{bs}}\big]\Big\|^{1/2}\Big\} \\
&\lesssim \eta\lambda_1\tau_*^{\mathsf{bs}}\Big[\sum_{i=N_*+1}^{n}(1-\widetilde{\eta})^{2(n-i)}\Big]^{1/2} \\
&\lesssim \kappa_*\tau_*^{\mathsf{bs}}\sqrt{\widetilde{\eta}}.
\end{aligned}
$$

Thus, matrix Freedman's inequality implies that, with probability at least $1-O(n^{-18})$,

$$\Big\|\sum_{i=N_*+1}^{n}\Xi_i^{\mathsf{bs}}\Big\|\lesssim\kappa_*\tau_*^{\mathsf{bs}}\big(\sqrt{\widetilde{\eta}\log n}+\widetilde{\eta}\log n\big)\lesssim\tau_*^{\mathsf{bs}}C_\psi^2\kappa_*\sqrt{r\widetilde{\eta}\log n}.$$

On the stability event from Lemma 33, the stopping indicators equal one for all $i>N_*$, and therefore

$$\overline{\alpha}_7^{\mathsf{bs}}\lesssim\tau_*^{\mathsf{bs}}C_\psi^2\kappa_*\sqrt{r\widetilde{\eta}\log n}$$

with probability at least $1-O(n^{-18})$. Combining this with (H.21) gives the desired bound.

□

H.5 Proof of Lemma 35

Proof. Recall (H.7), (H.10):

$$S_p:=\sum_{i=1}^{n}\|M_i\|_{\mathrm{F}}^p=O_{\mathbb{P}_X}\left(\frac{\lambda_\times^p}{\widetilde{\eta}}\right),$$
$$\|\widehat{\Gamma}_n\|_{\mathrm{HS}}=\Omega_{\mathbb{P}_X}\left(\frac{\lambda_\times^2}{\Delta_{\max}\sqrt{d_{\mathfrak{C}}}}\right).$$

For $L_3^{\xi,\mathsf{bs}}$, since $\xi_i^{\mathsf{bs}}=\sqrt{\eta}(w_i^{\mathsf{bs}}-1)M_i$ and the multipliers are uniformly bounded,

$$\frac{\sum_{i=1}^{n}\mathbb{E}^*\langle\xi_i^{\mathsf{bs}},\widehat{\Gamma}_n\xi_i^{\mathsf{bs}}\rangle_{\mathrm{F}}^{3/2}}{\|\widehat{\Gamma}_n\|_{\mathrm{HS}}^3}\lesssim\frac{\eta^{3/2}\|\widehat{\Gamma}_n\|_{\mathrm{op}}^{3/2}S_3}{\|\widehat{\Gamma}_n\|_{\mathrm{HS}}^3}\le\frac{\eta^{3/2}S_3}{\|\widehat{\Gamma}_n\|_{\mathrm{HS}}^{3/2}}=O_{\mathbb{P}_X}\left(\widetilde{\eta}^{1/2}\kappa_\Delta^{3/2}d_{\mathfrak{C}}^{3/4}\right).$$

Also, conditional independence gives

$$\frac{\sum_{1\le i<j\le n}\mathbb{E}^*\big|\langle\xi_i^{\mathsf{bs}},\xi_j^{\mathsf{bs}}\rangle_{\mathrm{F}}\big|^3}{\|\widehat{\Gamma}_n\|_{\mathrm{HS}}^3}\lesssim\frac{\eta^3S_3^2}{\|\widehat{\Gamma}_n\|_{\mathrm{HS}}^3}=O_{\mathbb{P}_X}\left(\widetilde{\eta}\kappa_\Delta^3d_{\mathfrak{C}}^{3/2}\right).$$

Therefore,

$$L_3^{\xi,\mathsf{bs}}=O_{\mathbb{P}_X}\left(\widetilde{\eta}^{1/2}\kappa_\Delta^{3/2}d_{\mathfrak{C}}^{3/4}+\widetilde{\eta}\kappa_\Delta^3d_{\mathfrak{C}}^{3/2}\right).$$

For $L_3^{g,\mathsf{bs}}$, conditionally on $\mathcal{F}_X$, $g_i^{\mathsf{bs}}\sim N(0,\eta M_i\otimes M_i)$, so we may write $g_i^{\mathsf{bs}}=\sqrt{\eta}\gamma_iM_i$, $\gamma_i\sim N(0,1)$. Hence

$$L_3^{g,\mathsf{bs}}=\frac{\sum_{i=1}^{n}\mathbb{E}^*\langle g_i^{\mathsf{bs}},\widehat{\Gamma}_ng_i^{\mathsf{bs}}\rangle_{\mathrm{F}}^{3/2}}{\|\widehat{\Gamma}_n\|_{\mathrm{HS}}^3}\lesssim\frac{\eta^{3/2}\|\widehat{\Gamma}_n\|_{\mathrm{op}}^{3/2}S_3}{\|\widehat{\Gamma}_n\|_{\mathrm{HS}}^3}\le\frac{\eta^{3/2}S_3}{\|\widehat{\Gamma}_n\|_{\mathrm{HS}}^{3/2}}=O_{\mathbb{P}_X}\left(\widetilde{\eta}^{1/2}\kappa_\Delta^{3/2}d_{\mathfrak{C}}^{3/4}\right).$$

Next, by the definition of K_3^{bs},

$$n^{-2}(K_3^{\mathsf{bs}})^3\lesssim\frac{\sum_{i=1}^{n}\mathbb{E}^*\|\xi_i^{\mathsf{bs}}\|_{\mathrm{F}}^6}{\|\widehat{\Gamma}_n\|_{\mathrm{HS}}^3}\lesssim\frac{\eta^3S_6}{\|\widehat{\Gamma}_n\|_{\mathrm{HS}}^3}=O_{\mathbb{P}_X}\left(\widetilde{\eta}^2\kappa_\Delta^3d_{\mathfrak{C}}^{3/2}\right).$$

Finally,

$$J_n^{\mathsf{bs}}=\frac{\sum_{i=1}^{n}\mathrm{Var}^*(\|\xi_i^{\mathsf{bs}}\|_{\mathrm{F}}^2)}{\|\widehat{\Gamma}_n\|_{\mathrm{HS}}^2}\le\frac{\sum_{i=1}^{n}\mathbb{E}^*\|\xi_i^{\mathsf{bs}}\|_{\mathrm{F}}^4}{\|\widehat{\Gamma}_n\|_{\mathrm{HS}}^2}\lesssim\frac{\eta^2S_4}{\|\widehat{\Gamma}_n\|_{\mathrm{HS}}^2}=O_{\mathbb{P}_X}\left(\widetilde{\eta}\kappa_\Delta^2d_{\mathfrak{C}}\right).$$

□

I Technical lemmas

Let $\mathcal{I}$ be the identity operator and $\mathcal{L}$ be the gap operator on $\mathbb{R}^{(d-r)\times r}$ defined as

$$[\mathcal{L}M]_{l,j} = \Delta_{lj} M_{l,j}, \quad \text{where} \quad \Delta_{lj} = \lambda_j - \lambda_{r+l} \text{ for any } l \in [d-r], j \in [r].$$

Lemma 36. *Assume $\eta\lambda_1 \in (0,1)$. Let D be a random matrix in $\mathbb{R}^{(d-r)\times r}$. Then the following hold.*

(i) $(\mathbb{E}D)^\top(\mathbb{E}D) \preceq \mathbb{E}D^\top D$ *which implies* $\|\mathbb{E}D\| \le \min(\|\mathbb{E}DD^\top\|^{1/2}, \|\mathbb{E}D^\top D\|^{1/2})$.

(ii) *for any matrix $M \in \mathbb{R}^{(d-r)\times r}$, $\|(\mathcal{I}-\eta\mathcal{L})M\| \le (1-\eta\Delta_{\min})\,\|M\|$.*

(iii) $\left\|\mathbb{E}\left[[(\mathcal{I}-\eta\mathcal{L})^m D]^\top\,[(\mathcal{I}-\eta\mathcal{L})^m D]\right]\right\|^{1/2} \le (1-\eta\Delta_{\min})^m \|\mathbb{E}D^\top D\|^{1/2}$ *for $m = 1, 2, \ldots$.*

(iv) $\left\|\mathbb{E}\left[[(\mathcal{I}-\eta\mathcal{L})^m D]\,[(\mathcal{I}-\eta\mathcal{L})^m D]^\top\right]\right\|^{1/2} \le (1-\eta\Delta_{\min})^m \|\mathbb{E}DD^\top\|^{1/2}$ *for $m = 1, 2, \ldots$.*

Proof. See Appendix I.1.1. □

Lemma 37 (Moment bound)**.** *Let $\widetilde{X} \in \mathbb{R}^d$ be a random vector with $\mathbb{E}\widetilde{X}\widetilde{X}^\top = I_d$ and $\|\widetilde{X}\|_{\psi_2} \le C_\psi$. Then for any $A \in \mathbb{R}^{p\times d}$ and any $k \ge 1$,*

$$\left[\mathbb{E}\left\|A\widetilde{X}\right\|^{2k}\right]^{1/(2k)} \le C\sqrt{k}\,\|A\|_{\mathrm{F}}\,C_\psi \tag{I.1}$$

$$\left[\mathbb{E}\left\|A\widetilde{X}\right\|^{2k}\right]^{1/(2k)} \le C\sqrt{kd}\|A\widetilde{X}\|_{\psi_2} \tag{I.2}$$

Equivalently, let $X = A\widetilde{X}$, we have

$$\left[\mathbb{E}\left\|X\right\|^{2k}\right]^{1/(2k)} \le C\sqrt{k\,\mathrm{tr}\,\mathbb{E}(XX^\top)}\,C_\psi$$

$$\left[\mathbb{E}\left\|X\right\|^{2k}\right]^{1/(2k)} \le C\sqrt{kd}\,\|X\|_{\psi_2}$$

Here $C > 0$ is an absolute constant.

Proof. See Appendix I.1.2. □

Theorem 12. *Consider a matrix martingale $\{Y_k : k = 0, 1, \ldots\}$ whose values are matrices in dimension $d_1 \times d_2$ and $Y_0 = 0$. Let $\{X_k : k = 1, \ldots\}$ be the difference sequence $X_k = Y_k - Y_{k-1}$. Assume that the difference sequence is uniformly bounded:*

$$\|X_k\| \le B \qquad \textit{almost surely} \quad \forall\, k = 1, 2, 3, \ldots$$

Define two predictable quadratic variation processes for the martingale:

$$W_{\mathrm{col},k} := \sum_{j=1}^k \mathbb{E}_{j-1} X_j X_j^\top, \qquad W_{\mathrm{row},k} := \sum_{j=1}^k \mathbb{E}_{j-1} X_j^\top X_j, \qquad \forall\, k = 1, 2, 3, \ldots$$

Then, for all $t \ge 0$ and σ^2,

$$\mathbb{P}\left\{\exists\, k \ge 0 : \|Y_k\| \ge t \textit{ and } \|W_{\mathrm{col},k}\| \vee \|W_{\mathrm{row},k}\| \le \sigma^2\right\} \le (d_1 + d_2)\exp\left(\frac{-t^2/2}{\sigma^2 + Bt/3}\right)$$

Proof. See [28, Corollary 1.3]. □

Lemma 38 (Matrix moments)**.** *Assume Assumption 1 holds. Then the following hold.*

(i) *For any $k \ge 1$,*

$$\left[\mathbb{E}\left\|ZY^\top\right\|_{\mathrm{F}}^{2k}\right]^{1/(2k)} \le C_1 k\,[\mathrm{tr}\,\Lambda_*]^{1/2}\,[\mathrm{tr}\,\Lambda_{*,\perp}]^{1/2}\,\left\|\widetilde{X}\right\|_{\psi_2}^2$$

$$\left[\mathbb{E}\left\|\widetilde{Z}\widetilde{Y}^\top\right\|_{\mathrm{F}}^{2k}\right]^{1/(2k)} \le C_1 k r^{1/2}(d-r)^{1/2}\left\|\widetilde{X}\right\|_{\psi_2}^2$$

for an absolute constant $C_1 > 0$. As a consequence, $d_{\mathfrak{C}} \gtrsim 1$.

(ii) *Let* $\mathfrak{C} = \operatorname{cov}(ZY^\top)$ *and* $\widetilde{\mathfrak{C}} = \operatorname{cov}(\widetilde{Z}\widetilde{Y}^\top)$. *Then*

$$\begin{aligned}\left\|\mathfrak{C}\right\|_{\mathrm{op}} &\le C_2 C_\psi^4 \min(r, d-r)\lambda_1\lambda_{r+1},\\ \left\|\widetilde{\mathfrak{C}}\right\|_{\mathrm{op}} &\le C_2 C_\psi^4 \min(r, d-r)\end{aligned}$$

for an absolute constant $C_2 > 0$.

Proof. See Appendix I.1.3. □

Lemma 39 (Alignment with the target subspace)**.** *Let* $U \in \mathbb{O}_{d,r}$ *with* $\mathcal{C} = U_*^\top U \in \mathbb{R}^{r\times r}$ *invertible,* $\mathcal{S} = U_{*,\perp}^\top U \in \mathbb{R}^{(d-r)\times r}$*, and* $\mathcal{T} = \mathcal{S}\mathcal{C}^{-1}$. *Then*

$$|||\sin\Theta(U, U_*)||| = |||\mathcal{T}\,(\mathcal{C}\mathcal{C}^\top)^{1/2}|||, \tag{I.3}$$

$$U\operatorname{sgn}(U^\top U_*) - U_* = U_*\big[(\mathcal{C}\mathcal{C}^\top)^{1/2} - I_r\big] + U_{*,\perp}\mathcal{T}(\mathcal{C}\mathcal{C}^\top)^{1/2}. \tag{I.4}$$

where $|||\cdot|||$ *is either operator norm or Frobenius norm. Denote* $\tau = \|\mathcal{T}\|_{\mathrm{op}}$. *Then*

$$\left\|(\mathcal{C}\mathcal{C}^\top)^{1/2} - I_r\right\| \le \left\|\mathcal{T}(\mathcal{C}\mathcal{C}^\top)^{1/2}\right\|^2 \le \tau^2 \tag{I.5}$$

If $\tau < 1$, *then* $\left\|(\mathcal{C}\mathcal{C}^\top)^{1/2} - I_r\right\|_{\mathrm{op}} \ge \tau^2/4$.

Proof. See Appendix I.1.4. □

The preceding lemma concerns alignment with the target subspace U_*. We next record its two-subspace analogue, which will be used to compare the original and bootstrap iterates.

Lemma 40 (Relative alignment of two subspaces)**.** *Let* $U_*, U_1, U_2 \in \mathbb{O}_{d,r}$ *with* $U_i = U_*\mathcal{C}_i + U_{*,\perp}\mathcal{S}_i$. *Denote* $M := U_2^\top U_1 = \mathcal{C}_2^\top\mathcal{C}_1 + \mathcal{S}_2^\top\mathcal{S}_1$. *Assume* $\mathcal{C}_1$, $\mathcal{C}_2$ *and* M *are invertible, and let* $\mathcal{T}_i = \mathcal{S}_i\mathcal{C}_i^{-1}$ *for* $i = 1, 2$. *Then the following hold.*

(i) (Decomposition) $U_2\operatorname{sgn}(U_2^\top U_1) - U_1 = U_*A + U_{*,\perp}B$ *where*

$$\begin{aligned}A &= \mathcal{C}_1\big[(M^\top M)^{1/2} - I_r\big] + \mathcal{C}_1\mathcal{C}_1^\top\mathcal{T}_1^\top(\mathcal{T}_1 - \mathcal{T}_2)\mathcal{C}_2\operatorname{sgn}(M),\\ B &= \mathcal{S}_1\big[(M^\top M)^{1/2} - I_r\big] - (I_{d-r} - \mathcal{S}_1\mathcal{S}_1^\top)(\mathcal{T}_1 - \mathcal{T}_2)\mathcal{C}_2\operatorname{sgn}(M).\end{aligned}$$

(ii) $\|\sin\Theta(U_2, U_1)\|_{\mathrm{F}}^2 = \left\|\mathcal{C}_1\mathcal{C}_1^\top\mathcal{T}_1^\top(\mathcal{T}_1 - \mathcal{T}_2)\mathcal{C}_2\right\|_{\mathrm{F}}^2 + \left\|(I_{d-r} - \mathcal{S}_1\mathcal{S}_1^\top)(\mathcal{T}_1 - \mathcal{T}_2)\mathcal{C}_2\right\|_{\mathrm{F}}^2$

(iii) *Let* $\tau_1 = \|\mathcal{T}_1\|$. *Then*

$$\left\|(M^\top M)^{1/2} - I_r\right\| \le \|\sin\Theta(U_2, U_1)\|^2 \le (1 + \tau_1^2)\,\|\mathcal{T}_1 - \mathcal{T}_2\|^2$$

Proof. See Appendix I.2. □

Lemma 41 (Initialization)**.** (i) [13, Lemma 28] and [10, Theorem II.13] *Let* $G \in \mathbb{R}^{d\times r}$ *be a standard Gaussian matrix. For any* $t \ge 0$,

$$\mathbb{P}\big\{\|G\| \ge \sqrt{d} + \sqrt{r} + t\big\} \le 2e^{-t^2/2}.$$

(ii) [13, Lemma 29] and [2, Lemma i.A.3] *Let* $G \in \mathbb{R}^{r\times r}$ *be a standard Gaussian matrix. For any* $\delta \in (0, 1)$,

$$\mathbb{P}\Big\{\left\|G^{-1}\right\| \ge \frac{6\sqrt{r}}{\delta}\Big\} \le \delta.$$

(iii) (Haar initialization) *Fix* $U_* \in \mathbb{O}_{d,r}$ *and let* $U_0 \sim \mathsf{Haar}(\mathbb{O}_{d,r})$. *Then for any* $\delta_p \in (0, 1/e]$, *any* $c_p \ge 1$, *we have*

$$\mathbb{P}\Big\{\left\|(U_{*,\perp}^\top U_0)(U_*^\top U_0)^{-1}\right\| \le \frac{48c_p\sqrt{2\log(4c_p)}\sqrt{rd\log(1/\delta_p)}}{\delta_p}\Big\} \ge 1 - \frac{\delta_p}{c_p}.$$

Proof. See Appendix I.2.1. $\square$

Theorem 13 (Berry-Esseen)**.** *Let $\xi_1, \ldots, \xi_n$ be independent, $\mathbb{R}^d$-valued random variables with zero means. Let $W = \sum_{i=1}^n \xi_i$ and assume $\Sigma = \mathrm{cov}(W)$ is invertible. Let Z be a d-dimensional Gaussian random vector with zero mean and covariance Σ. Then we have*

$$\sup_{\mathcal{A} \in \mathscr{A}^d} |\mathbb{P}\{W \in \mathcal{A}\} - \mathbb{P}\{Z \in \mathcal{A}\}| \leq (42d^{1/4} + 16)\gamma, \qquad \text{where } \gamma = \sum_{i=1}^n \mathbb{E}\|\Sigma^{-1/2}\xi_i\|^3$$

Here $\mathscr{A}^d$ represents the set of all convex sets in $\mathbb{R}^d$.

Proof. See [32, Theorem 11]. $\square$

Lemma 42. *Let $\mu \in \mathbb{R}^d$, and let Σ_1, Σ_2 be positive definite $d \times d$ matrices. Then the total variation distance between $\mathcal{N}(\mu, \Sigma_1)$ and $\mathcal{N}(\mu, \Sigma_2)$ – denoted by $\mathsf{TV}(\mathcal{N}(\mu, \Sigma_1), \mathcal{N}(\mu, \Sigma_2))$–satisfies*

$$\frac{1}{100} \leq \frac{\mathsf{TV}(\mathcal{N}(\mu, \Sigma_1), \mathcal{N}(\mu, \Sigma_2))}{\min\left\{1, \left\|\Sigma_1^{-1/2}\Sigma_2\Sigma_1^{-1/2} - I_d\right\|_{\mathrm{F}}\right\}} \leq \frac{3}{2}$$

Proof. See [32, Theorem 12]. $\square$

For any point $x \in \mathbb{R}^r$ and any non-empty convex set $\mathcal{A} \in \mathscr{A}^r$ satisfying $\mathcal{A} \neq \mathbb{R}^r$, define the signed distance function as

$$\delta_{\mathcal{A}}(x) := \begin{cases} -\mathsf{dist}(x, \mathbb{R}^r \setminus \mathcal{A}), & \text{if } x \in \mathcal{A}; \\ \mathsf{dist}(x, \mathcal{A}), & \text{if } x \notin \mathcal{A}. \end{cases}$$

Here, $\mathsf{dist}(x, \mathcal{A})$ is the Euclidean distance between $x \in \mathbb{R}^r$ and a non-empty set $\mathcal{A} \subset \mathbb{R}^r$. Also, for any $t \in \mathbb{R}$, define

$$\mathcal{A}^t := \{x \in \mathbb{R}^r : \delta_{\mathcal{A}}(x) \leq t\} \tag{I.6}$$

for any non-empty convex set $\mathcal{A} \in \mathscr{A}^r$ satisfying $\mathcal{A} \neq \mathbb{R}^r$, and define $\emptyset^t = \emptyset$ and $(\mathbb{R}^r)^t = \mathbb{R}^r$.

For any convex set $\mathcal{A} \in \mathscr{A}^r$, define the following quantity related to Gaussian distributions

$$\begin{aligned} &\gamma_r := \sup_{\mathcal{A} \in \mathscr{A}^r} \gamma(\mathcal{A}), \\ &\quad \text{where} \quad \gamma(\mathcal{A}) := \sup_{t>0} \max\left\{\frac{1}{t}\mathcal{N}(0, I_r)\left\{\mathcal{A}^t \setminus \mathcal{A}\right\}, \frac{1}{t}\mathcal{N}(0, I_r)\left\{\mathcal{A} \setminus \mathcal{A}^{-t}\right\}\right\}, \end{aligned}$$

Theorem 14. *For all $r \in \mathbb{N}$, we have $\gamma_r < 0.59r^{1/4} + 0.21$.*

Proof. See [32, Theorem 13]. $\square$

I.1 Proof of technical lemmas

I.1.1 Proof of Lemma 36

(i) for any $u \in \mathbb{R}^r$, we have

$$u^\top \left[\mathbb{E}D^\top D - (\mathbb{E}D)^\top(\mathbb{E}D)\right] u = \mathbb{E}\|Du\|^2 - \|\mathbb{E}Du\|^2 \geq 0$$

Thus $(\mathbb{E}D)^\top(\mathbb{E}D) \preceq \mathbb{E}D^\top D$, which combined with $\|M\|^2 = \|MM^\top\| = \|M^\top M\|$ gives the desired result.

(ii) By the definition of $\mathcal{L}$, we have

$$(\mathcal{I} - \eta\mathcal{L})M = \eta\Lambda_{*,\perp}M + M(I_r - \eta\Lambda_*)$$

Then the triangle inequality gives

$$\begin{aligned} \|(\mathcal{I} - \eta\mathcal{L})M\| &\leq \eta\|\Lambda_{*,\perp}M\| + \|M(I_r - \eta\Lambda_*)\| \\ &\leq \eta\lambda_{r+1}\|M\| + (1 - \eta\lambda_r)\|M\| \end{aligned}$$

$$= (1-\eta(\lambda_r-\lambda_{r+1}))\,\|M\| = (1-\eta\Delta_{\min})\,\|M\|\,.$$

(iii) Without loss of generality, assume $m=1$ since $m>1$ follows by iterating the $m=1$ case. For any unit vector $u\in\mathbb{R}^r$, we have

$$\begin{aligned}\Big[u^\top\mathbb{E}\big[[(\mathcal{I}-\eta\mathcal{L})D]^\top\,[(\mathcal{I}-\eta\mathcal{L})D]\big]u\Big]^{1/2} &= \|[(\mathcal{I}-\eta\mathcal{L})D]\,u\|_{L_2}\\ &= \|\eta\Lambda_{*,\perp}Du + D(I_r-\eta\Lambda_*)u\|_{L_2}\\ &\le \|\eta\Lambda_{*,\perp}Du\|_{L_2} + \|D(I_r-\eta\Lambda_*)u\|_{L_2}\end{aligned} \tag{I.7}$$

Note that

$$\|\eta\Lambda_{*,\perp}Du\|_{L_2}^2 \le \eta^2\lambda_{r+1}^2\,\|Du\|_{L_2}^2 = \eta^2\lambda_{r+1}^2\,\big\|\mathbb{E}u^\top D^\top Du\big\| \le \eta^2\lambda_{r+1}^2\,\big\|\mathbb{E}D^\top D\big\|$$

and

$$\begin{aligned}\|D(I_r-\eta\Lambda_*)u\|_{L_2}^2 &= u^\top(I_r-\eta\Lambda_*)(\mathbb{E}D^\top D)(I_r-\eta\Lambda_*)u\\ &\le \|(I_r-\eta\Lambda_*)u\|^2\,\big\|\mathbb{E}D^\top D\big\| \le (1-\eta\lambda_r)^2\,\big\|\mathbb{E}D^\top D\big\|\end{aligned}$$

which combined with (I.7) gives

$$\begin{aligned}\Big[u^\top\mathbb{E}\big[[(\mathcal{I}-\eta\mathcal{L})D]^\top\,[(\mathcal{I}-\eta\mathcal{L})D]\big]u\Big]^{1/2} &\le (1-\eta\lambda_r+\eta\lambda_{r+1})\,\big\|\mathbb{E}D^\top D\big\|^{1/2}\\ &= (1-\eta\Delta_{\min})\|\mathbb{E}D^\top D\|^{1/2}\end{aligned}$$

(iv) is similar to (iii).

I.1.2 Proof of Lemma 37

If $A=0$, the bound is trivial. Assume $A\neq 0$.

Proof of (I.1). Let $A_{i,\cdot}\in\mathbb{R}^{1\times d}$ denote the i-th row of A, then

$$\|A\widetilde{X}\|^2 = \sum_{i=1}^p (A_{i,\cdot}\widetilde{X})^2 = \|A\|_{\mathrm{F}}^2\cdot\sum_{A_{i,\cdot}\neq 0}\frac{\|A_{i,\cdot}\|^2}{\|A\|_{\mathrm{F}}^2}\frac{(A_{i,\cdot}\widetilde{X})^2}{\|A_{i,\cdot}\|^2}$$

Note that $\sum_{A_{i,\cdot}\neq 0}\|A_{i,\cdot}\|^2/\|A\|_{\mathrm{F}}^2 = 1$. Taking k-th power and applying Jensen's inequality with weights $\|A_{i,\cdot}\|^2/\|A\|_{\mathrm{F}}^2$ gives

$$\mathbb{E}\|A\widetilde{X}\|^{2k} \le \|A\|_{\mathrm{F}}^{2k}\sum_{A_{i,\cdot}\neq 0}\frac{\|A_{i,\cdot}\|^2}{\|A\|_{\mathrm{F}}^2}\mathbb{E}\bigg[\frac{(A_{i,\cdot}\widetilde{X})^2}{\|A_{i,\cdot}\|^2}\bigg]^k$$

Since $[\mathbb{E}(v^\top\widetilde{X})^{2k}]^{1/2k}\le C\sqrt{k}C_\psi$ holds uniformly for all unit vector $v\in\mathbb{R}^d$ for some absolute constant $C>0$ [29], taking $2k$-th root gives

$$\big[\mathbb{E}\|A\widetilde{X}\|^{2k}\big]^{1/(2k)} \le C\sqrt{k}\;\|A\|_{\mathrm{F}}\,\big\|\widetilde{X}\big\|_{\psi_2}$$

Proof of (I.2). Denote $X=A\widetilde{X}$. For any $k\ge 1$,

$$\|X\|^{2k} = \Big[\sum_{i=1}^d X_i^2\Big]^k \le d^{k-1}\sum_{i=1}^d X_i^{2k}$$

Since $\|X_i\|_{\psi_2}\le\|X\|_{\psi_2}$ for all $i\in[d]$, taking expectation and $2k$-th root, we have

$$\big[\mathbb{E}\,\|X\|^{2k}\big]^{1/(2k)} \le \Big[\mathbb{E}d^{k-1}\sum_{i=1}^d X_i^{2k}\Big]^{1/(2k)} \le C\sqrt{kd}\;\|X\|_{\psi_2}$$

for the same absolute constant $C>0$ as before.

I.1.3 Proof of Lemma 38

(i) Note that $\left\|ZY^\top\right\|_{\mathrm{F}}^2 = \|Z\|^2\|Y\|^2$, then we have

$$\left\|ZY^\top\right\|_{\mathrm{F}}^{2k} = \left(\left\|\Lambda_{*,\perp}^{1/2}\widetilde{Z}\right\|^2\left\|\Lambda_*^{1/2}\widetilde{Y}\right\|^2\right)^k = \left(\sum_{l=1}^{d-r}\sum_{j=1}^{r}\lambda_{r+l}\lambda_j\widetilde{Z}_l^2\widetilde{Y}_j^2\right)^k$$

Let $S = [\operatorname{tr}\Lambda_{*,\perp}][\operatorname{tr}\Lambda_*]$. Note that $\sum_{lj}\lambda_{r+l}\lambda_j = S$. Since $k \geq 1$, applying Jensen's inequality gives

$$\left\|ZY^\top\right\|_{\mathrm{F}}^{2k} \leq S^k \cdot \sum_{l=1}^{d-r}\sum_{j=1}^{r}\frac{\lambda_{r+l}\lambda_j}{S}\widetilde{Z}_l^{2k}\widetilde{Y}_j^{2k}$$

For sub-exponential norm, we have

$$\left\|\widetilde{Z}_l\widetilde{Y}_j\right\|_{\psi_1} \leq \left\|\widetilde{Z}_l\right\|_{\psi_2}\left\|\widetilde{Y}_j\right\|_{\psi_2} \leq \left\|\widetilde{Z}\right\|_{\psi_2}\left\|\widetilde{Y}\right\|_{\psi_2} \leq \left\|\widetilde{X}\right\|_{\psi_2}^2$$

Hence

$$\begin{aligned}\mathbb{E}\left\|ZY^\top\right\|_{\mathrm{F}}^{2k} &\leq S^k \cdot \sum_{l=1}^{d-r}\sum_{j=1}^{r}\frac{\lambda_{r+l}\lambda_j}{S}\mathbb{E}(\widetilde{Z}_l\widetilde{Y}_j)^{2k} \leq S^k \cdot \sum_{l=1}^{d-r}\sum_{j=1}^{r}\frac{\lambda_{r+l}\lambda_j}{S}\left[Ck\left\|\widetilde{X}\right\|_{\psi_2}^2\right]^{2k} \\ &= S^k \cdot \left[Ck\left\|\widetilde{X}\right\|_{\psi_2}^2\right]^{2k}\end{aligned}$$

thus

$$\left[\mathbb{E}\left\|ZY^\top\right\|_{\mathrm{F}}^{2k}\right]^{1/(2k)} \leq CkS^{1/2}\left\|\widetilde{X}\right\|_{\psi_2}^2 = Ck \cdot [\operatorname{tr}\Lambda_*]^{1/2}[\operatorname{tr}\Lambda_{*,\perp}]^{1/2}\left\|\widetilde{X}\right\|_{\psi_2}^2$$

Applying the above inequality with $k = 1$ gives

$$d_{\mathfrak{C}} := \frac{\lambda_\times^4}{\|\mathfrak{C}\|_{\mathrm{HS}}^2} \gtrsim \left[\frac{\mathbb{E}\|ZY^\top\|_{\mathrm{F}}^2}{\|\mathfrak{C}\|_{\mathrm{HS}}}\right]^2 \geq 1,$$

where the last inequality follows from the fact that $\|\mathfrak{C}\|_{\mathrm{HS}} \leq \operatorname{tr}(\mathfrak{C}) = \mathbb{E}\|ZY^\top\|_{\mathrm{F}}^2$.

(ii) For any $A \in \mathbb{R}^{(d-r)\times r}$ with $\|A\|_{\mathrm{F}} = 1$,

$$\begin{aligned}\langle A, \mathfrak{C}A\rangle_{\mathrm{F}} = \mathbb{E}\left\langle ZY^\top, A\right\rangle_{\mathrm{F}}^2 &= \mathbb{E}\left\langle\widetilde{Z}\widetilde{Y}^\top, \Lambda_{*,\perp}^{1/2}A\Lambda_*^{1/2}\right\rangle_{\mathrm{F}}^2 \\ &\leq \left\|\widetilde{\mathfrak{C}}\right\|_{\mathrm{op}}\left\|\Lambda_{*,\perp}^{1/2}A\Lambda_*^{1/2}\right\|_{\mathrm{F}}^2 \leq \left\|\widetilde{\mathfrak{C}}\right\|_{\mathrm{op}}\lambda_1\lambda_{r+1}\end{aligned}$$

Thus

$$\|\mathfrak{C}\|_{\mathrm{op}} \leq \left\|\widetilde{\mathfrak{C}}\right\|_{\mathrm{op}}\lambda_1\lambda_{r+1}$$

It suffices to show $\left\|\widetilde{\mathfrak{C}}\right\|_{\mathrm{op}} \lesssim C_\psi^4\min(r, d-r)$. For any $A \in \mathbb{R}^{(d-r)\times r}$ with $\|A\|_{\mathrm{F}} = 1$, we have

$$\left\langle\widetilde{Z}\widetilde{Y}^\top, A\right\rangle_{\mathrm{F}} = \widetilde{Z}^\top A\widetilde{Y} = (A^\top\widetilde{Z})^\top\widetilde{Y} = \widetilde{Z}^\top(A\widetilde{Y})$$

(a) By Cauchy-Schwarz inequality, we have

$$\left\langle A, \widetilde{\mathfrak{C}}A\right\rangle_{\mathrm{F}} = \mathbb{E}\left\langle\widetilde{Z}\widetilde{Y}^\top, A\right\rangle_{\mathrm{F}}^2 = \mathbb{E}\left[(A^\top\widetilde{Z})^\top\widetilde{Y}\right]^2 \leq \left[\mathbb{E}\left\|A^\top\widetilde{Z}\right\|^4\right]^{1/2}\left[\mathbb{E}\left\|\widetilde{Y}\right\|^4\right]^{1/2}$$

Applying Lemma 37 to each factor gives

$$\left[\mathbb{E}\left\|A^\top\widetilde{Z}\right\|^4\right]^{1/4} \lesssim C_\psi\left\|A^\top\right\|_{\mathrm{F}} = C_\psi, \quad \left[\mathbb{E}\left\|\widetilde{Y}\right\|^4\right]^{1/4} \lesssim C_\psi\|I_r\|_{\mathrm{F}} = C_\psi\sqrt{r}$$

Thus, we have

$$\left\langle A, \widetilde{\mathfrak{C}}A\right\rangle_{\mathrm{F}} \lesssim C_\psi^4 r.$$

Taking supremum over A gives $\left\|\widetilde{\mathfrak{C}}\right\|_{\mathrm{op}} \lesssim C_\psi^4 r$.

(b) Applying the same argument as in (a) to $\widetilde{Z}^\top(A\widetilde{Y})$ gives $\left\|\widetilde{\mathfrak{C}}\right\|_{\mathrm{op}} \lesssim C_\psi^4(d-r)$.

I.1.4 Proof of Lemma 39

We first establish the following identities for $\mathcal{C}$ and $\mathcal{S}$:

$$\mathcal{C}\,\mathrm{sgn}(\mathcal{C}^\top) = (\mathcal{C}\mathcal{C}^\top)^{1/2}, \qquad \mathcal{S}\,\mathrm{sgn}(\mathcal{C}^\top) = \mathcal{T}(\mathcal{C}\mathcal{C}^\top)^{1/2}. \tag{I.8}$$

To see this, note that for any matrix M with singular value decomposition $M = A\Lambda B^\top$, the sign matrix of M is defined as $\mathrm{sgn}(M) = AB^\top$. Hence $M = \mathrm{sgn}(M)(M^\top M)^{1/2}$. Applying this to $M = \mathcal{C}^\top$ gives $\mathcal{C}^\top = \mathrm{sgn}(\mathcal{C}^\top)\,(\mathcal{C}\mathcal{C}^\top)^{1/2}$. Since $\mathcal{C}$ is invertible by assumption, so is $(\mathcal{C}\mathcal{C}^\top)^{1/2}$, hence we have $\mathrm{sgn}(\mathcal{C}^\top) = \mathcal{C}^\top(\mathcal{C}\mathcal{C}^\top)^{-1/2}$. Thus, for the first identity in (I.8), we have

$$\mathcal{C}\,\mathrm{sgn}(\mathcal{C}^\top) = \mathcal{C}\mathcal{C}^\top(\mathcal{C}\mathcal{C}^\top)^{-1/2} = (\mathcal{C}\mathcal{C}^\top)^{1/2}.$$

For the second, using $\mathcal{S} = \mathcal{T}\mathcal{C}$ gives

$$\mathcal{S}\,\mathrm{sgn}(\mathcal{C}^\top) = \mathcal{T}\mathcal{C}\mathcal{C}^\top(\mathcal{C}\mathcal{C}^\top)^{-1/2} = \mathcal{T}(\mathcal{C}\mathcal{C}^\top)^{1/2}.$$

With (I.8) in hand, we can obtain a representation of $U\,\mathrm{sgn}(U^\top U_*)$. To see this, since $U^\top U_* = \mathcal{C}^\top$, plugging in (I.8) gives

$$U\,\mathrm{sgn}(U^\top U_*) = (U_*\mathcal{C} + U_{*,\perp}\mathcal{S})\,\mathrm{sgn}(\mathcal{C}^\top) = U_*(\mathcal{C}\mathcal{C}^\top)^{1/2} + U_{*,\perp}\mathcal{T}(\mathcal{C}\mathcal{C}^\top)^{1/2}. \tag{I.9}$$

which implies (I.3) and (I.4) directly by noticing that $\|\sin\Theta(U, U_*)\|_{\mathrm{F}} = \|\mathcal{S}\|_{\mathrm{F}}$.

Proof of (I.5)**.** Since $\mathcal{C} \in \mathbb{R}^{r\times r}$ is a square matrix, $\mathcal{C}\mathcal{C}^\top$ and $\mathcal{C}^\top\mathcal{C}$ share eigenvalues. From $\mathcal{C}^\top\mathcal{C} = I_r - \mathcal{S}^\top\mathcal{S}$, the eigenvalues of $\mathcal{C}\mathcal{C}^\top$ are $\{1-\sigma_j^2\}_{j=1}^r$, where $\sigma_1 \geq \cdots \geq \sigma_r \geq 0$ are the singular values of $\mathcal{S}$. Hence $(\mathcal{C}\mathcal{C}^\top)^{1/2}$ has eigenvalues $\{\sqrt{1-\sigma_j^2}\}_{j=1}^r$, all lying in $[\sqrt{1-\sigma_1^2}, 1]$.

For the operator norm bound (I.5), since $1 - \sqrt{1-x} \leq x$ for all $0 \leq x \leq 1$, we have

$$\left\|I_r - (\mathcal{C}\mathcal{C}^\top)^{1/2}\right\|_{\mathrm{op}} = \max_j \left\{1 - \sqrt{1-\sigma_j^2}\right\} \leq \max_j \sigma_j^2 = \|\mathcal{S}\|_{\mathrm{op}}^2 = \left\|\mathcal{T}(\mathcal{C}\mathcal{C}^\top)^{1/2}\right\|_{\mathrm{op}}^2.$$

where the last equality follows since $\|\mathcal{S}\|_{\mathrm{op}} = \left\|\mathcal{S}\,\mathrm{sgn}(\mathcal{C}^\top)\right\|_{\mathrm{op}}$ and (I.8). Finally, $\|\mathcal{S}\|_{\mathrm{op}} = \|\mathcal{T}\mathcal{C}\|_{\mathrm{op}} \leq \|\mathcal{T}\|_{\mathrm{op}}\|\mathcal{C}\|_{\mathrm{op}} \leq \tau$, so $\left\|(\mathcal{C}\mathcal{C}^\top)^{1/2} - I_r\right\|_{\mathrm{op}} \leq \tau^2$.

Proof of Lower bound. Since $1 - \sqrt{1-x} \geq x/2$ for all $x \in [0,1]$, we have

$$\left\|I_r - (\mathcal{C}\mathcal{C}^\top)^{1/2}\right\|_{\mathrm{op}} = 1 - \sqrt{1-\sigma_1^2} \geq \frac{\sigma_1^2}{2} = \frac{\|\mathcal{S}\|_{\mathrm{op}}^2}{2}.$$

Finally, we connect $\|\mathcal{S}\|_{\mathrm{op}}$ to τ via

$$\|\mathcal{S}\|_{\mathrm{op}} = \|\mathcal{C}\mathcal{T}\|_{\mathrm{op}} \geq \sigma_{\min}(\mathcal{C})\tau = \sqrt{1 - \|\mathcal{S}\|_{\mathrm{op}}^2}\,\tau$$

which rearranges to $\|\mathcal{S}\|_{\mathrm{op}}^2 \geq \tau^2/(1+\tau^2)$. Hence

$$\left\|I_r - (\mathcal{C}\mathcal{C}^\top)^{1/2}\right\|_{\mathrm{op}} \geq \frac{\|\mathcal{S}\|_{\mathrm{op}}^2}{2} \geq \frac{\tau^2}{2(1+\tau^2)}.$$

When $\tau \in [0,1]$, we have $(1+\tau^2) \leq 2$, so

$$\left\|I_r - (\mathcal{C}\mathcal{C}^\top)^{1/2}\right\|_{\mathrm{op}} \geq \frac{\tau^2}{4}.$$

I.2 Proof of Lemma 40

Preliminaries. Using $\mathcal{S}_i = \mathcal{T}_i\mathcal{C}_i$, we have $\mathcal{S}_1^\top\mathcal{S}_2 = \mathcal{C}_1^\top\mathcal{T}_1^\top\mathcal{T}_2\mathcal{C}_2$ and

$$\begin{aligned} U_1^\top U_2 = \mathcal{C}_1^\top\mathcal{C}_2 + \mathcal{S}_1^\top\mathcal{S}_2 = \mathcal{C}_1^\top(I_r + \mathcal{T}_1^\top\mathcal{T}_2)\mathcal{C}_2 &= \mathcal{C}_1^\top(I_r + \mathcal{T}_1^\top\mathcal{T}_1)\mathcal{C}_2 - \mathcal{C}_1^\top\mathcal{T}_1^\top(\mathcal{T}_1 - \mathcal{T}_2)\mathcal{C}_2 \\ &\overset{(a)}{=} \mathcal{C}_1^{-1}\mathcal{C}_2 - \mathcal{C}_1^\top\mathcal{T}_1^\top(\mathcal{T}_1 - \mathcal{T}_2)\mathcal{C}_2 \end{aligned} \tag{I.10}$$

Here (a) follows from the fact that since $\mathcal{C}_1^\top\mathcal{C}_1 + \mathcal{S}_1^\top\mathcal{S}_1 = I_r$, we have $\mathcal{C}_1^\top(I_r + \mathcal{T}_1^\top\mathcal{T}_1)\mathcal{C}_1 = I_r$, which implies $I_r + \mathcal{T}_1^\top\mathcal{T}_1 = \mathcal{C}_1^{\top -1}\mathcal{C}_1^{-1}$.

(i) Decomposition. We first decompose the aligned difference into its components parallel and orthogonal to $\mathrm{col}(U_1)$:

$$U_2 \operatorname{sgn}(M) - U_1 = U_1 U_1^\top [U_2 \operatorname{sgn}(M) - U_1] + (I - U_1 U_1^\top) U_2 \operatorname{sgn}(M). \tag{I.11}$$

- First term: Since $U_1^\top U_2 = M^\top$ and $M^\top \operatorname{sgn}(M) = (M^\top M)^{1/2}$, the first term equals $U_1[(M^\top M)^{1/2} - I_r]$. Using $U_1 = U_* \mathcal{C}_1 + U_{*,\perp} \mathcal{S}_1$ gives

$$U_1[(M^\top M)^{1/2} - I_r] = U_* \mathcal{C}_1 [(M^\top M)^{1/2} - I_r] + U_{*,\perp} \mathcal{S}_1 [(M^\top M)^{1/2} - I_r].$$

- Second term: we first compute U_* and $U_{*,\perp}$ components of $(I - U_1 U_1^\top) U_2$:
  - U_*-component is $\mathcal{C}_2 - \mathcal{C}_1(\mathcal{C}_1^\top \mathcal{C}_2 + \mathcal{S}_1^\top \mathcal{S}_2)$. Using (I.10) gives

$$\begin{aligned}\mathcal{C}_2 - \mathcal{C}_1(\mathcal{C}_1^\top \mathcal{C}_2 + \mathcal{S}_1^\top \mathcal{S}_2) &= \underbrace{\mathcal{C}_2 - \mathcal{C}_1 \mathcal{C}_1^\top (I_r + \mathcal{T}_1^\top \mathcal{T}_1) \mathcal{C}_2}_{=0} + \mathcal{C}_1 \mathcal{C}_1^\top \mathcal{T}_1^\top (\mathcal{T}_1 - \mathcal{T}_2) \mathcal{C}_2 \\ &= \mathcal{C}_1 \mathcal{C}_1^\top \mathcal{T}_1^\top (\mathcal{T}_1 - \mathcal{T}_2) \mathcal{C}_2\end{aligned}$$

    where the last line follows from $\mathcal{C}_1^\top \mathcal{C}_1 + \mathcal{S}_1^\top \mathcal{S}_1 = I_r$.
  - $U_{*,\perp}$-component is $\mathcal{S}_2 - \mathcal{S}_1(\mathcal{C}_1^\top \mathcal{C}_2 + \mathcal{S}_1^\top \mathcal{S}_2)$. Using (I.10) gives

$$\begin{aligned}\mathcal{S}_2 - \mathcal{S}_1(\mathcal{C}_1^\top \mathcal{C}_2 + \mathcal{S}_1^\top \mathcal{S}_2) &= \mathcal{T}_2 \mathcal{C}_2 - \mathcal{T}_1(\mathcal{C}_2 - \mathcal{C}_1 \mathcal{C}_1^\top \mathcal{T}_1^\top (\mathcal{T}_1 - \mathcal{T}_2) \mathcal{C}_2) \\ &= -(\mathcal{T}_1 - \mathcal{T}_2) \mathcal{C}_2 + \mathcal{S}_1 \mathcal{S}_1^\top (\mathcal{T}_1 - \mathcal{T}_2) \mathcal{C}_2 \\ &= -(I_{d-r} - \mathcal{S}_1 \mathcal{S}_1^\top)(\mathcal{T}_1 - \mathcal{T}_2) \mathcal{C}_2\end{aligned}$$

Combining the above gives the desired decomposition.

(ii) $\|\sin\Theta(U_2, U_1)\|_{\mathrm{F}}^2$. This follows directly from (I.11) and the fact that $\|\sin\Theta(U_2, U_1)\|_{\mathrm{F}}^2 = \|(I - U_1 U_1^\top) U_2\|_{\mathrm{F}}^2$ [8, Lemma 2.5].

(iii) It remains to prove the operator-norm bounds. Let $U_{1,\perp} \in \mathbb{O}_{d,d-r}$ be an orthogonal complement of U_1, and set $\widetilde{M} = U_{1,\perp}^\top U_2$. Since $[U_1, U_{1,\perp}]$ is orthogonal,

$$U_2^\top U_2 = U_2^\top U_1 U_1^\top U_2 + U_2^\top U_{1,\perp} U_{1,\perp}^\top U_2,$$

that is, $MM^\top + \widetilde{M}^\top \widetilde{M} = I_r$. Thus the singular values of M are $\sqrt{1 - \sigma_j^2(\widetilde{M})}$, while the singular values of $\widetilde{M}$ are $\sin\theta_j(U_2, U_1)$. Therefore

$$\|(M^\top M)^{1/2} - I_r\| = \max_j \{1 - \sqrt{1 - \sin^2\theta_j}\}.$$

Using $\frac{x}{2} \le 1 - \sqrt{1-x} \le x$ for $x \in [0,1]$, with $x = \sin^2\theta_j$, gives

$$\frac{1}{2}\|\sin\Theta(U_2, U_1)\|^2 \le \|(M^\top M)^{1/2} - I_r\| \le \|\sin\Theta(U_2, U_1)\|^2.$$

Finally, since $\|\sin\Theta(U_2, U_1)\| = \|(I - U_1 U_1^\top) U_2\|$ [8, Lemma 2.5], analysis of the second term in (I.11) gives

$$\|\sin\Theta(U_2, U_1)\|^2 \le \|\mathcal{C}_1 \mathcal{C}_1^\top \mathcal{T}_1^\top \Delta \mathcal{C}_2\|^2 + \|(I_{d-r} - \mathcal{S}_1 \mathcal{S}_1^\top)(\mathcal{T}_1 - \mathcal{T}_2) \mathcal{C}_2\|^2,$$

Since $\|\mathcal{C}_i\| \le 1$ and $\|I_{d-r} - \mathcal{S}_1 \mathcal{S}_1^\top\| \le 1$, we have

$$\|\sin\Theta(U_2, U_1)\|^2 \le (1 + \tau_1^2)\|\mathcal{T}_1 - \mathcal{T}_2\|^2.$$

The proof is complete.

I.2.1 Proof of Lemma 41

Proof of Item (ii). This is Lemma 29 of [13], which follows from Lemma i.A.3 of [2] upon setting $p = \delta^2/4$ and observing that $\mathrm{tr}[(G^\top G)^{-1}] = \|G^{-1}\|_{\mathrm{F}}^2$.

Proof of Item (iii). Without loss of generality, we can assume $U_* = [I_r; 0]$ and $U_0 = G(G^\top G)^{-1/2}$ where $G \in \mathbb{R}^{d\times r}$ is a standard Gaussian matrix. Write

$$G = \begin{pmatrix} G_1 \\ G_2 \end{pmatrix}$$

where $G_1 \in \mathbb{R}^{r\times r}$ and $G_2 \in \mathbb{R}^{(d-r)\times r}$ are independent standard Gaussian matrices. Then

$$(U_{*,\perp}^\top U_0)(U_*^\top U_0)^{-1} = G_2 G_1^{-1}.$$

Item (ii) gives $\mathbb{P}(\|G_1^{-1}\| \geq 12\sqrt{r}/\delta) \leq \delta/2$. Item (i) with $t = \sqrt{2\log(4/\delta)}$ gives

$$\mathbb{P}\big\{\|G_2\| \geq \sqrt{d-r} + \sqrt{r} + \sqrt{2\log(4/\delta)}\big\} \leq \delta/2.$$

Combining the two bounds gives

$$\mathbb{P}\Big\{\big\|(U_{*,\perp}^\top U_0)(U_*^\top U_0)^{-1}\big\| \leq \frac{12\sqrt{r}}{\delta}\left(\sqrt{d-r} + \sqrt{r} + \sqrt{2\log(4/\delta)}\right)\Big\} \geq 1-\delta.$$

Since $\sqrt{d-r} + \sqrt{r} \leq \sqrt{2d}$, and

$$\sqrt{2d} + \sqrt{2\log(4/\delta)} \leq 2\sqrt{2d}\sqrt{2\log(4/\delta)} = 4\sqrt{d\log(4/\delta)}$$

We have

$$\mathbb{P}\Big\{\big\|(U_{*,\perp}^\top U_0)(U_*^\top U_0)^{-1}\big\| \leq \frac{48\sqrt{rd\log(4/\delta)}}{\delta}\Big\} \geq 1-\delta.$$

Set $\delta = \delta_p/c_p$ for $c_p \geq 1$, since $\log(4c_p/\delta_p) \leq 2\log(4c_p)\log(1/\delta_p)$ for $\delta_p \leq 1/e$, we have

$$\mathbb{P}\Big\{\big\|(U_{*,\perp}^\top U_0)(U_*^\top U_0)^{-1}\big\| \leq \frac{48c_p\sqrt{2\log(4c_p)}\sqrt{rd\log(1/\delta_p)}}{\delta_p}\Big\} \geq 1-\frac{\delta_p}{c_p}.$$

J Additional experiments

We provide additional row-wise diagnostics for Theorem 8 under the same simulation settings as in Section 5.

Row-specific coverage. Table 4 summarizes the empirical coverage rates across all $d = 100$ rows by reporting their quantiles. The coverage rates are concentrated near the nominal level, particularly for faster spectral decay, while some rows exhibit undercoverage for smaller β. The variation across rows is natural, as Theorem 8 is a fixed-row result with row-dependent conditions, so finite-sample accuracy may differ across $m \in [d]$.

Table 4: Quantiles across rows of the row-specific empirical coverage rates at nominal level 0.95 over 200 Monte Carlo trials.

β	Min	10%	20%	30%	40%	50%	60%	70%	80%	90%	Max
0.5	0.840	0.880	0.890	0.897	0.906	0.910	0.920	0.930	0.930	0.940	0.970
1	0.860	0.899	0.910	0.920	0.920	0.930	0.934	0.940	0.950	0.960	0.990
4	0.900	0.920	0.930	0.930	0.940	0.940	0.950	0.950	0.960	0.970	0.970

Row-specific Q–Q diagnostics. To further assess row-wise calibration, for each row $m \in [d]$ we compare the standardized quadratic statistic

$$\widehat{Q}_m := ([U_* \,\mathrm{sgn}(U_*^\top U_n)]_{m,\cdot} - [U_n]_{m,\cdot})^\top [\widehat{\Sigma}_m^{\mathrm{bs}}]^{-1}([U_* \,\mathrm{sgn}(U_*^\top U_n)]_{m,\cdot} - [U_n]_{m,\cdot})$$

with the χ_r^2 distribution using Q–Q plots. Table 5 summarizes, across all $d = 100$ rows, the quantiles of the fitted Q–Q slopes, with slopes near one indicating good agreement. Figure 2 additionally shows the Q–Q plots for the first two rows under each $\beta \in \{0.5, 1, 4\}$. Overall, the slopes become more concentrated around one as the spectral decay becomes faster, consistent with the coverage results above.

Table 5: Quantiles across rows of the fitted slopes of the row-specific Q–Q plots against the χ_r^2 distribution.

β	Min	10%	20%	30%	40%	50%	60%	70%	80%	90%	Max
0.5	0.943	1.052	1.107	1.121	1.156	1.189	1.214	1.257	1.292	1.348	1.431
1	0.899	1.017	1.049	1.063	1.082	1.107	1.138	1.169	1.197	1.240	1.343
4	0.912	0.964	0.987	1.002	1.018	1.033	1.058	1.075	1.102	1.140	1.228

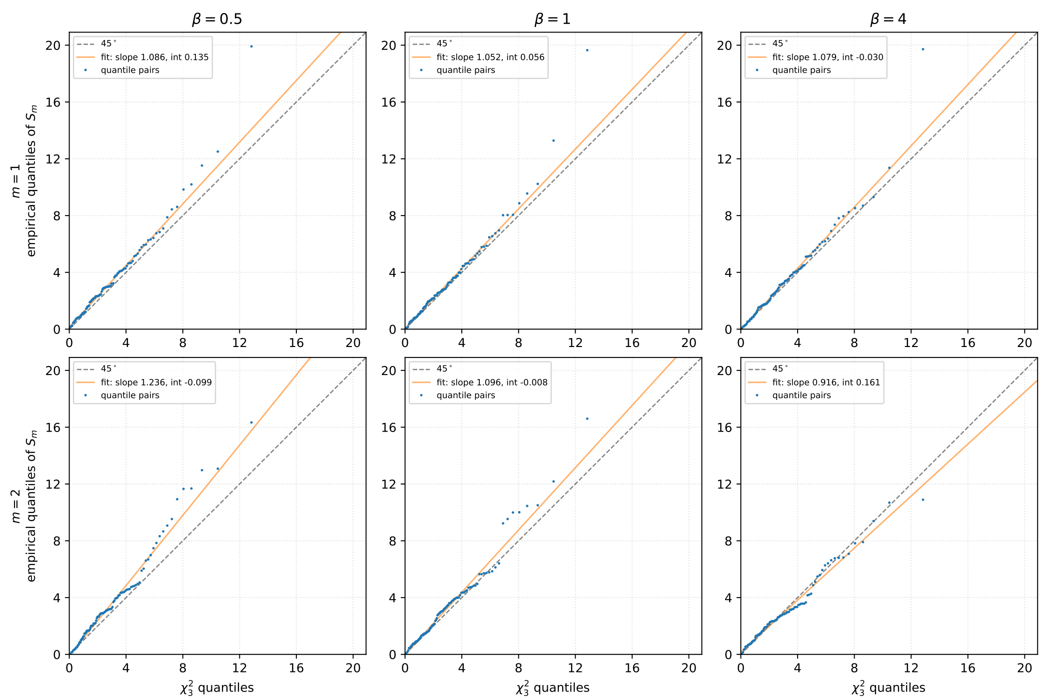


Figure 2: Row-specific Q–Q plots of the standardized quadratic statistic against the χ_r^2 distribution for $m = 1, 2$ and $\beta \in \{0.5, 1, 4\}$. The diagonal line represents exact agreement. Columns correspond to $\beta = 0.5, 1, 4$, and rows correspond to $m = 1, 2$.