Error Estimation for Sketched SVD via the Bootstrap — Supplement —

Organization of supplement. The main aspects of the proof of Theorem 1 are presented in Section A, and the arguments in this section will refer to lower level results that are stated and proved in Sections B, C, D, and E. Next, in Section F, we provide detailed examples of matrices that satisfy both of the assumptions RP and RS. Lastly, in Section G, we present additional experimental results that go beyond the settings considered in the main text. For ease of reference, we include here statements of the theoretical setup, Assumptions RP and RS, and Theorem 1.

Theoretical setup. Our result is formulated in terms of a sequence of deterministic matrices $A_n \in \mathbb{R}^{n\times d}$ indexed by $n = 1, 2, \ldots$, such that d remains fixed as $n \to \infty$. Likewise, the number $k \in \{1, \ldots, d\}$ and the set of indices $\mathcal{J} \subset \{1, \ldots, k\}$ remain fixed as well. In addition, for each n, there is an associated random sketching matrix $S_n \in \mathbb{R}^{t_n\times n}$ and a number of bootstrap samples B_n such that $t_n \to \infty$ and $B_n \to \infty$ as $n \to \infty$. Here, it is important to note that we make no restriction on the sizes of t_n and B_n relative to n, and hence we allow $t_n/n \to 0$ and $B_n/n \to 0$. Lastly, in order to lighten notation in Theorem 1, we will suppress dependence on n for the outputs of Algorithm 1, as well the exact singular vectors/values (u_j, v_j, σ_j) of A_n and their sketched versions $(\tilde{u}_j, \tilde{v}_j, \tilde{\sigma}_j)$.

Assumption RP. *There is a positive definite matrix G_∞ in $\mathbb{R}^{d\times d}$ such that $\frac{1}{n}A_n^\top A_n \to \mathsf{G}_\infty$ as $n \to \infty$, and the eigenvalues of G_∞ each have multiplicity 1.*

To state Assumption RS, let $(p_1, \ldots, p_n)$ denote the row-sampling probabilities for S_n, and let $a_l \in \mathbb{R}^d$ denote the lth row of A_n. In addition, let $\tilde{r}_n \in \mathbb{R}^d$ denote the first row of the re-scaled sketch $\frac{\sqrt{t}}{\sqrt{n}}S_nA_n$, and let $\mathsf{v}_1, \mathsf{v}_2 \in \mathbb{R}^d$ denote the top two eigenvectors of G_∞.

Assumption RS. *The following conditions hold in addition to Assumption RP. For any fixed matrix $C \in \mathbb{R}^{d\times d}$, the sequence $\operatorname{var}(\tilde{r}_n^\top C\tilde{r}_n)$ converges to a finite limit $\ell(C)$, possibly zero, as $n \to \infty$. Furthermore, if C is chosen as $C = \mathsf{v}_1\mathsf{v}_1^\top$ or $C = \mathsf{v}_1\mathsf{v}_2^\top$, then the limit $\ell(C)$ is positive. Lastly, the condition $\max_{1\leq l\leq n} \|\frac{1}{\sqrt{np_l}}a_l\|_2 = o(t_n^{1/8})$ holds as $n \to \infty$.*

Theorem 1. *Suppose that Assumption RP holds when S_n is a Gaussian random projection, or that Assumption RS holds when S_n is a row-sampling matrix. Also, let $\widehat{q}_{\scriptscriptstyle U}(t_n)$, $\widehat{q}_{\scriptscriptstyle \Sigma}(t_n)$, and $\widehat{q}_{\scriptscriptstyle V}(t_n)$ denote the outputs of Algorithm 1. Then, for any fixed set $\mathcal{J} \subset \{1, \ldots, k\}$ containing 1, and any $\alpha \in (0, 1)$, the following three limits hold as $n \to \infty$,*

$$\mathbb{P}\Big(\max_{j\in\mathcal{J}} \rho_{\sin}(\tilde{u}_j, u_j) \;\leq\; \widehat{q}_{\scriptscriptstyle U}(t_n)\Big) \;\longrightarrow\; 1-\alpha, \tag{0.1}$$

$$\mathbb{P}\Big(\max_{j\in\mathcal{J}} |\tilde{\sigma}_j - \sigma_j| \;\leq\; \widehat{q}_{\scriptscriptstyle \Sigma}(t_n)\Big) \;\longrightarrow\; 1-\alpha, \tag{0.2}$$

$$\mathbb{P}\Big(\max_{j\in\mathcal{J}} \rho_{\sin}(\tilde{v}_j, v_j) \;\leq\; \widehat{q}_{\scriptscriptstyle V}(t_n)\Big) \;\longrightarrow\; 1-\alpha. \tag{0.3}$$

A Proof of Theorem 1

We decompose the main parts of the proof into Sections A.2, A.3, and A.4 corresponding to the three limits (in the order of 0.3, (0.1), and (0.2)). In addition, we provide a summary of the notation and terminology for the proofs immediately below.

A.1 Notation and terminology for proofs

Items related to matrices. For any real matrices L and M of the same size, we frequently use the inner product $\langle\!\langle L, M\rangle\!\rangle := \mathrm{tr}(L^\top M)$. Also, for any real matrix M, the Frobenius norm $\|M\|_F$ is equal to $\sqrt{\mathrm{tr}(M^\top M)}$, and the operator norm $\|M\|_{\mathrm{op}}$ is equal to the maximum singular value of M. The set of symmetric matrices in $\mathbb{R}^{d\times d}$ is denoted $\mathcal{S}^{d\times d}$, and for any $M \in \mathcal{S}^{d\times d}$, its ordered eigenvalues are written as $\lambda_j(M) \geq \lambda_{j+1}(M)$. Likewise, the ordered singular values of a general real (possibly rectangular) matrix R are denoted $\sigma_j(R) \geq \sigma_{j+1}(R)$. For a sketch of A_n, we write $\tilde{A}_n = S_n A_n$, and similarly, the matrix $\tilde{A}_n^*$ is defined as having rows that are sampled with replacement from the rows of $\tilde{A}_n$. When the context is clear, we will sometimes use the shorthand notation

$$
\begin{aligned}
\sigma_j &= \sigma_j(A_n),\\
\tilde{\sigma}_j &= \sigma_j(\tilde{A}_n),\\
\tilde{\sigma}_j^* &= \sigma_j(\tilde{A}_n^*).
\end{aligned}
$$

Similarly, the jth left and right singular vectors of $\tilde{A}_n$ are denoted as $\tilde{u}_j$ and $\tilde{v}_j$, and likewise for $\tilde{u}_j^*$ and $\tilde{v}_j^*$ with respect to $\tilde{A}_n^*$. Hence, the dependence on n will be generally suppressed for these vectors. In addition, for the normalized Gram matrices associated with A_n, $\tilde{A}_n$ and $\tilde{A}_n^*$, we define

$$G_n := \tfrac{1}{n} A_n^\top A_n, \tag{A.1}$$

$$\tilde{G}_n := \tfrac{1}{n} \tilde{A}_n^\top \tilde{A}_n, \tag{A.2}$$

$$\tilde{G}_n^* := \tfrac{1}{n} (\tilde{A}_n^*)^\top (\tilde{A}_n^*). \tag{A.3}$$

Lastly, recall that under Assumption RP, the matrix G_n converges to a positive definite matrix $\mathsf{G}_\infty \in \mathbb{R}^{d\times d}$ as $n \to \infty$. Accordingly, sans-serif font will be reserved for other limiting objects, such as the leading eigenvectors $\mathsf{v}_1, \mathsf{v}_2 \in \mathbb{R}^d$ of G_∞, as described in Assumptions RS.

Items related to probability. If Y is a random matrix, we write $\mathcal{L}(Y)$ to refer to its distribution, and if Z is another random matrix, we write $\mathcal{L}(Y|Z)$ to refer to the conditional distribution of Y given Z. If $\{Y_n\}$ is a sequence of random matrices converging in probability to another random matrix Y_∞ as $n \to \infty$, we write either $Y_n = Y_\infty + o_{\mathbb{P}}(1)$, or $Y_n \xrightarrow{\mathbb{P}} Y_\infty$. Next, if Y_n converges to Y_∞ in distribution, we write $\mathcal{L}(Y_n) \xrightarrow{d} \mathcal{L}(Y_\infty)$. In addition, it is important to define a notion of convergence for conditional distributions. Specifically, if $\{Z_n\}$ is a sequence of random matrices, we will need to define the convergence of the conditional distributions $\mathcal{L}(Y_n|Z_n)$. To do this, first note that ordinary convergence in distribution can be equivalently expressed in terms of various metrics on the space of probability measures. That is, the limit $\mathcal{L}(Y_n) \xrightarrow{d} \mathcal{L}(Y_\infty)$ is equivalent to $\varrho(\mathcal{L}(Y_n), \mathcal{L}(Y_\infty)) \to 0$, where ϱ is a metric such as the Lévy-Prohorov metric, or the bounded Lipschitz metric (cf. [Dudley, 2002, Sec. 11.3]). Likewise, for conditional distributions we write '$\mathcal{L}(Y_n|Z_n) \xrightarrow{d} \mathcal{L}(Y_\infty)$ in probability' if the sequence of scalar random variables $\{\varrho(\mathcal{L}(Y_n|Z_n), \mathcal{L}(Y_\infty))\}$ converges to 0 in probability.

A.2 Proof of the limit (0.3)

We begin with a reduction that is often used in the literature on bootstrap methods. Specifically, it is known that (0.3) can be reduced to showing that

$$\mathcal{L}\Big(\sqrt{t_n} \max_{j\in\mathcal{J}} \rho_{\sin}(\tilde{v}_j, v_j)\Big) \xrightarrow{d} \mathcal{L}(\xi_{V}), \tag{A.4}$$

and

$$\mathcal{L}\Big(\sqrt{t_n} \max_{j\in\mathcal{J}} \rho_{\sin}(\tilde{v}_j^*, \tilde{v}_j)\Big| S_n\Big) \xrightarrow{d} \mathcal{L}(\xi_{V}) \quad \text{in probability}, \tag{A.5}$$

for some random variable ξ_{V} whose distribution function is continuous. (For further details, please see Theorem 1.2.1, as well as Remark 1.2.1, and the discussion on p.5 of the book Politis et al. [1999].) Next, as a step towards showing the limits (A.4) and (A.5), we will use some algebraic identities involving the projection matrices associated with v_j, $\tilde{v}_j$, and $\tilde{v}_j^*$, which we denote as

$$P_j := v_j v_j^\top, \qquad \tilde{P}_j := \tilde{v}_j \tilde{v}_j^\top, \qquad \tilde{P}_j^* := (\tilde{v}_j^*)(\tilde{v}_j^*)^\top, \tag{A.6}$$

where the fact that these matrices depend on n has been suppressed. The relevant identities are

$$\sqrt{t_n}\rho_{\sin}(\tilde{v}_j, v_j) \;=\; \tfrac{1}{\sqrt{2}}\big\|\sqrt{t_n}(\tilde{P}_j - P_j)\big\|_F, \tag{A.7}$$

$$\sqrt{t_n}\rho_{\sin}(\tilde{v}_j^*, \tilde{v}_j) \;=\; \tfrac{1}{\sqrt{2}}\big\|\sqrt{t_n}(\tilde{P}_j^* - \tilde{P}_j)\big\|_F. \tag{A.8}$$

As a consequence of these identities, we may write

$$\sqrt{t_n}\max_{j\in\mathcal{J}}\rho_{\sin}(\tilde{v}_j, v_j) \;=\; f\Big(\sqrt{t_n}(\tilde{P}_1 - P_1), \dots, \sqrt{t_n}(\tilde{P}_k - P_k)\Big), \tag{A.9}$$

$$\sqrt{t_n}\max_{j\in\mathcal{J}}\rho_{\sin}(\tilde{v}_j^*, \tilde{v}_j) \;=\; f\Big(\sqrt{t_n}(\tilde{P}_1^* - \tilde{P}_1), \dots, \sqrt{t_n}(\tilde{P}_k^* - \tilde{P}_k)\Big) \tag{A.10}$$

where $f : (\mathbb{R}^{d\times d})^k \to \mathbb{R}$ is defined by $f(C_1, \dots, C_k) = \max_{j\in\mathcal{J}} \frac{1}{\sqrt{2}}\|C_j\|_F$. In turn, by the continuous mapping theorem and the Cramér-Wold theorem [Kallenberg, 2006, Theorem 3.27 and Corollary 4.5]), the limits (A.4) and (A.5) will hold if we can show that for any fixed matrices $M_1, \dots, M_k \in \mathbb{R}^{d\times d}$, there is an associated Gaussian random vector, say $(Z_1(M_1), \dots, Z_k(M_k)) \in \mathbb{R}^k$, such that

$$\mathcal{L}\Big(\langle\!\langle\sqrt{t_n}(\tilde{P}_1 - P_1), M_1\rangle\!\rangle, \dots, \langle\!\langle\sqrt{t_n}(\tilde{P}_k - P_k), M_k\rangle\!\rangle\Big) \;\xrightarrow{d}\; \mathcal{L}(Z_1(M_1), \dots, Z_k(M_k)), \quad \text{and} \tag{A.11}$$

$$\mathcal{L}\Big(\langle\!\langle\sqrt{t_n}(\tilde{P}_1^* - \tilde{P}_1), M_1\rangle\!\rangle, \dots, \langle\!\langle\sqrt{t_n}(\tilde{P}_k^* - \tilde{P}_k), M_k\rangle\!\rangle\Big|S_n\Big) \;\xrightarrow{d}\; \mathcal{L}(Z_1(M_1), \dots, Z_k(M_k)) \quad \text{in probability.} \tag{A.12}$$

These limits are established in Lemmas 4 and 5 below, where we handle certain key technical challenges. In addition, we must verify the condition that the limiting random variable ξ_V in (A.4) and (A.5) has a continuous distribution function. To do this, first note that the limit (A.11) allows us to view the tuple of matrices $\big(\sqrt{t_n}(\tilde{P}_j - P_j)\big)_{j\in\mathcal{J}}$ as converging in distribution to a Gaussian vector in the space $(\mathbb{R}^{d\times d})^{|\mathcal{J}|}$. Also, it is a basic fact that the norm of a Gaussian vector with a non-zero covariance matrix yields a random variable whose distribution function is continuous. So, given that the function f restricts to a norm on $(\mathbb{R}^{d\times d})^{|\mathcal{J}|}$, it suffices to show that the mentioned Gaussian vector in $(\mathbb{R}^{d\times d})^{|\mathcal{J}|}$ has positive variance when projected into at least one direction. In other words, to show that ξ_V has a continuous distribution function, it is enough to show that there is at least one index $j \in \mathcal{J}$ and matrix $M_j \in \mathbb{R}^{d\times d}$ such that $\langle\!\langle\sqrt{t_n}(\tilde{P}_j - P_j), M_j\rangle\!\rangle$ has a limiting Gaussian distribution with positive variance — and this is handled in Lemma 4. Altogether, this completes the proof of the first limit (0.3) in Theorem 1.

Remark. The proofs of the second and third limits (0.1) and (0.2) will require different versions of the Lemmas 4 and 5, and these are given later on in Lemmas 7 and 8. □

A.3 Proof of the limit (0.1)

By the reduction argument used at the beginning of Section A.2, it suffices to show that

$$\mathcal{L}\Big(\sqrt{t_n}\max_{j\in\mathcal{J}}\rho_{\sin}(\tilde{u}_j, u_j)\Big) \xrightarrow{d} \mathcal{L}(\xi_U), \tag{A.13}$$

and

$$\mathcal{L}\Big(\sqrt{t_n}\max_{j\in\mathcal{J}}\rho_{\sin}(\tilde{A}_n\tilde{v}_j^*, \tilde{A}_n\tilde{v}_j)\Big|S_n\Big) \xrightarrow{d} \mathcal{L}(\xi_U) \quad \text{in probability,} \tag{A.14}$$

for some random variable ξ_U whose distribution function is continuous. Next, to develop a counterparts of the relation (A.7), define the projection matrices

$$\Pi_j := u_j u_j^\top = \tfrac{A_n P_j A_n^\top}{\operatorname{tr}(P_j A_n^\top A_n)} \qquad \text{and} \qquad \tilde{\Pi}_j := \tilde{u}_j\tilde{u}_j^\top = \tfrac{A_n \tilde{P}_j A_n^\top}{\operatorname{tr}(\tilde{P}_j A_n^\top A_n)}.$$

Similarly, to develop a counterpart of (A.8), define the vectors $\breve{u}_j := \frac{\tilde{A}\tilde{v}_j}{\|\tilde{A}\tilde{v}_j\|_2}$ and $\breve{u}_j^* := \frac{\tilde{A}\tilde{v}_j^*}{\|\tilde{A}\tilde{v}_j^*\|_2}$, and their associated projections

$$\breve{\Pi}_j := \breve{u}_j\breve{u}_j^\top = \frac{\tilde{A}_n\tilde{P}_j^\top\tilde{A}_n^\top}{\operatorname{tr}(\tilde{P}_j\tilde{A}_n^\top\tilde{A}_n)} \qquad \text{and} \qquad \breve{\Pi}_j^* := (\breve{u}_j^*)(\breve{u}_j^*)^\top = \frac{\tilde{A}\tilde{P}_j^*\tilde{A}_n^\top}{\operatorname{tr}(\tilde{P}_j^*\tilde{A}_n^\top\tilde{A}_n)}.$$

Remark. For a finite n, it is possible that the denominators $\operatorname{tr}(\tilde{P}_jA_n^\top A_n)$, $\operatorname{tr}(\tilde{P}_j\tilde{A}^\top\tilde{A})$, or $\operatorname{tr}(\tilde{P}_j^*\tilde{A}_n^\top\tilde{A}_n)$ may be zero, and if this occurs, we instead define $\tilde{\Pi}_j$, $\breve{\Pi}_j$, or $\breve{\Pi}_j^*$ to be the zero matrix. However, the probability of such events will turn out to go to zero asymptotically, and hence, such events will be unimportant. This same type of consideration will occur at other points in the proofs, and so in order to avoid repetition, we will not make further mention of zero denominators that occur with vanishing probability as $n \to \infty$.

In the above notation, it is straightforward to check the identities

$$\rho_{\sin}(\tilde{u}_j, u_j) = \tfrac{1}{\sqrt{2}}\|\tilde{\Pi}_j - \Pi_j\|_F \tag{A.15}$$

$$\rho_{\sin}(\breve{u}_j^*, \breve{u}_j) = \tfrac{1}{\sqrt{2}}\|\breve{\Pi}_j^* - \breve{\Pi}_j\|_F. \tag{A.16}$$

Since the Frobenius norm of a symmetric matrix only depends on the non-zero eigenvalues, we may replace the matrices $(\tilde{\Pi}_j - \Pi_j)$ and $(\tilde{\Pi}_j^* - \tilde{\Pi}_j)$ above with different matrices whose non-zero eigenvalues are the same. In particular, the matrix $A_nMA_n^\top$ has the same non-zero eigenvalues as $(A_n^\top A_n)^{1/2}M(A_n^\top A_n)^{1/2}$ for any $M \in \mathbb{R}^{d\times d}$. So, if we recall the definition $G_n = \frac{1}{n}A_n^\top A_n$ from (A.1), it follows that the matrix

$$\tilde{\Delta}_j := G_n^{1/2}\Big(\frac{\tilde{P}_j}{\operatorname{tr}(\tilde{P}_jG_n)} - \frac{P_j}{\operatorname{tr}(P_jG_n)}\Big)G_n^{1/2} \tag{A.17}$$

has the same non-zero eigenvalues as $(\tilde{\Pi}_j - \Pi_j)$, and therefore

$$\rho_{\sin}(\tilde{u}_j, u_j) = \tfrac{1}{\sqrt{2}}\|\tilde{\Delta}_j\|_F. \tag{A.18}$$

Similarly, if we recall the definition $\tilde{G}_n = \frac{1}{n}\tilde{A}_n^\top\tilde{A}_n$ and define

$$\tilde{\Delta}_j^* := \tilde{G}_n^{1/2}\Big(\frac{\tilde{P}_j^*}{\operatorname{tr}(\tilde{P}_j^*\tilde{G}_n)} - \frac{\tilde{P}_j}{\operatorname{tr}(\tilde{P}_j\tilde{G}_n)}\Big)\tilde{G}_n^{1/2}, \tag{A.19}$$

then we have

$$\rho_{\sin}(\breve{u}_j^*, \breve{u}_j) = \tfrac{1}{\sqrt{2}}\|\tilde{\Delta}_j^*\|_F. \tag{A.20}$$

The key significance of working with the $d \times d$ matrices $\tilde{\Delta}_j$ and $\tilde{\Delta}_j^*$ is that they remain of a fixed size asymptotically, whereas the $n \times n$ matrices $(\tilde{\Pi}_j - \Pi_j)$ and $(\breve{\Pi}_j^* - \breve{\Pi}_j)$ expand as $n \to \infty$.

At this stage, the identities (A.18) and (A.20) will play the role that (A.7) and (A.8) did earlier. In turn, we may apply the previous reasoning based on the continuous mapping theorem and the Cramér-Wold theorem. In this way, the limits (A.13) and (A.14) will hold if we can show that for any fixed matrices $M_1, \dots, M_k \in \mathbb{R}^{d\times d}$, there is an associated Gaussian vector, say $(\zeta_1(M_1), \dots, \zeta_k(M_k)) \in \mathbb{R}^k$, such that

$$\mathcal{L}\Big(\langle\!\langle\sqrt{t_n}\tilde{\Delta}_1, M_1\rangle\!\rangle, \dots, \langle\!\langle\sqrt{t_n}\tilde{\Delta}_k, M_k\rangle\!\rangle\Big) \xrightarrow{d} \mathcal{L}(\zeta_1(M_1), \dots, \zeta_k(M_k)), \quad \text{and} \tag{A.21}$$

$$\mathcal{L}\Big(\sqrt{t_n}\tilde{\Delta}_1^*, M_1\rangle\!\rangle, \dots, \langle\!\langle\sqrt{t_n}\tilde{\Delta}_k^*, M_k\rangle\!\rangle\Big|S_n\Big) \xrightarrow{d} \mathcal{L}(\zeta_1(M_1), \dots, \zeta_k(M_k)) \quad \text{in probability.} \tag{A.22}$$

These limits are established in Lemma 7 below. Lastly, to ensure that the limiting random variable ξ_U in (A.13) and (A.14) has a continuous distribution function, the reasoning in Section A.2 shows that it is sufficient to exhibit at least one index $j \in \mathcal{J}$ and matrix $M_j \in \mathbb{R}^{d\times d}$ such that $\operatorname{var}(\zeta_j(M_j)) > 0$. This is also done in Lemma 7. $\square$

A.4 Proof of the limit (0.2)

As in the previous two subsections, the proof can be reduced to showing that

$$\mathcal{L}\Big(\tfrac{\sqrt{t_n}}{\sqrt{n}}\max_{j\in\mathcal{J}}|\tilde{\sigma}_j-\sigma_j|\Big)\ \xrightarrow{d}\ \mathcal{L}(\xi_{\Sigma}),\quad\text{and}\tag{A.23}$$

$$\mathcal{L}\Big(\tfrac{\sqrt{t_n}}{\sqrt{n}}\max_{j\in\mathcal{J}}|\tilde{\sigma}_j^*-\tilde{\sigma}_j|\,\Big|S\Big)\ \xrightarrow{d}\ \mathcal{L}(\xi_{\Sigma}),\quad\text{in probability}\tag{A.24}$$

for some random variable ξ_{Σ} whose distribution function is continuous. (Here, we use the normalizing factor $\frac{\sqrt{t_n}}{\sqrt{n}}$, rather than the $\sqrt{t_n}$ used in (A.4) and (A.5), because the singular values depend on the scaling of the matrix $\frac{1}{n}A_n^\top A_n$ — whereas the singular vectors do not.) Proceeding as before, the continuous mapping theorem and the Cramér-Wold theorem imply that (A.23) and (A.24) will hold if we can show that the following limits hold for any constants $c_1,\dots,c_k\in\mathbb{R}$,

$$\mathcal{L}\Big(\sum_{j=1}^{k}\tfrac{\sqrt{t_n}}{\sqrt{n}}c_j(\tilde{\sigma}_j-\sigma_j)\Big)\ \xrightarrow{d}\mathcal{L}(\zeta(c_1,\dots,c_k)),\tag{A.25}$$

and

$$\mathcal{L}\Big(\textstyle\sum_{j=1}^{k}\tfrac{\sqrt{t_n}}{\sqrt{n}}c_j(\sigma_j(\tilde{A}_n^*)-\sigma_j(\tilde{A}_n))\,\Big|\,S_n\Big)\ \xrightarrow{d}\ \mathcal{L}(\zeta(c_1,\dots,c_k))\quad\text{in probability,}\tag{A.26}$$

where $\zeta(c_1,\dots,c_k)$ is a Gaussian scalar random variable. These limits are established in Lemma 8. In addition, we can show that ξ_{Σ} has a continuous distribution function in the same way as was done for ξ_V and ξ_U, which amounts to showing that $\zeta(1,0,\dots,0)$ has positive variance — and this is shown in Lemma 8 as well. This completes the proof. $\square$

B Intermediate results

The results in this section are relevant to the proofs of all three limits in Theorem 1. The first is a well known result, usually called Slutsky's lemma [van der Vaart, 2000, Lemma 2.8], whereas the second is a conditional version of it that is tailored to the current paper. Hence, we only provide a proof of the conditional version. Lastly, in Lemmas 2 and 3, we provide a CLT for $\sqrt{t_n}(\tilde{G}_n-G_n)$, as well as its bootstrap counterpart $\sqrt{t_n}(\tilde{G}_n^*-\tilde{G}_n)$.

Fact 1 (Slutsky's lemma). *For each $n\geq 1$, let $T_n\in\mathbb{R}^{d_1\times d_2}$ and $R_n\in\mathbb{R}^{d_1'\times d_2'}$ be random matrices whose dimensions remain fixed as $n\to\infty$. In addition, suppose there is a random matrix $T_\infty\in\mathbb{R}^{d_1\times d_2}$ and a constant matrix $R_\infty\in\mathbb{R}^{d_1'\times d_2'}$ such that $\mathcal{L}(T_n)\xrightarrow{d}\mathcal{L}(T_\infty)$ and $R_n\to R_\infty$ in probability. Then, for any continuous function $g:\mathbb{R}^{d_1\times d_2}\times\mathbb{R}^{d_1'\times d_2'}\to\mathbb{R}$, the following limit holds*

$$\mathcal{L}\big(g(T_n,R_n)\big)\ \xrightarrow{d}\ \mathcal{L}\big(g(T_\infty,R_\infty)\big).\tag{B.1}$$

Lemma 1 (Conditional Slutsky's lemma). *For each $n\geq 1$, let $\mathcal{D}_n=\{X_{1,n},\dots,X_{t_n,n}\}$ be a set of random variables, and let $\mathcal{D}_n^*=\{X_{1,n}^*,\dots,X_{t_n,n}^*\}$ be sampled with replacement from $\mathcal{D}_n$. Also, for each $n\geq 1$, let $T_n^*=T_n(\mathcal{D}_n^*)$ be a real random matrix of size $d_1\times d_2$ computed from $\mathcal{D}_n^*$. In addition, let $R_n\in\mathbb{R}^{d_1'\times d_2'}$ be a random matrix that may depend on both $\mathcal{D}_n$ and $\mathcal{D}_n^*$. Lastly, suppose that there is a random matrix $T_\infty\in\mathbb{R}^{d_1\times d_2}$ and a constant matrix $R_\infty\in\mathbb{R}^{d_1'\times d_2'}$ such that*

$$\mathcal{L}(T_n^*|\mathcal{D}_n)\ \xrightarrow{d}\ \mathcal{L}(T_\infty)\quad\textit{in probability,}\tag{B.2}$$

and for any $\epsilon>0$,

$$\mathbb{P}\big(\|R_n-R_\infty\|_F>\epsilon\,\big|\mathcal{D}_n\big)\ \to\ 0\quad\textit{in probability.}\tag{B.3}$$

Then, for any continuous function $g:\mathbb{R}^{d_1\times d_2}\times\mathbb{R}^{d_1'\times d_2'}\to\mathbb{R}$, the following limit holds

$$\mathcal{L}\big(g(T_n^*,R_n)\big|\mathcal{D}_n\big)\ \xrightarrow{d}\ \mathcal{L}\big(g(T_\infty,R_\infty)\big)\quad\textit{in probability.}\tag{B.4}$$

Proof. By the continuous mapping theorem, it suffices to show that the bounded-Lipschitz metric between $\mathcal{L}\big(T_n^*, R_n\big|\mathcal{D}_n\big)$ and $\mathcal{L}\big(T_\infty, R_\infty\big)$ converges to 0 in probability. (Please see the comments in Section A.1 for additional background.) Let $\mathcal{F}$ denote the class of functions from $\mathbb{R}^{d_1\times d_2}\times\mathbb{R}^{d_1'\times d_2'}$ to $\mathbb{R}$ that are bounded in magnitude by 1 and are 1-Lipschitz with respect to the Frobenius norm. Then,

$$\begin{aligned}\sup_{f\in\mathcal{F}}\Big|\mathbb{E}[f(T_n^*, R_n)|\mathcal{D}_n)] - \mathbb{E}[f(T_\infty, R_\infty)]\Big| \ &\leq \ \sup_{f\in\mathcal{F}}\Big|\mathbb{E}\big[f(T_n^*, R_n) - f(T_n^*, R_\infty)\big|\mathcal{D}_n\big]\Big| \\ &\quad + \ \sup_{f\in\mathcal{F}}\Big|\mathbb{E}[f(T_n^*, R_\infty)|\mathcal{D}_n)] - \mathbb{E}[f(T_\infty, R_\infty)]\Big|.\end{aligned} \tag{B.5}$$

Regarding the second term on the right side, note that for each $f\in\mathcal{F}$, the associated function $h(\cdot) := f(\cdot, R_\infty)$ on $\mathbb{R}^{d_1\times d_2}$ is bounded in magnitude by 1 and is 1-Lipschiz with respect to the Frobenius norm. Hence, the assumption (B.2) implies that the second term in the bound (B.5) converges to 0 in probability.

Regarding the first term of the bound (B.5), we can decompose the expectation by writing the constant 1 as a sum of two indicators, $1 = 1_{E_n} + 1_{E_n^c}$, where we define the event $E_n := \{\|(T_n^*, R_n) - (T_n^*, R_\infty)\|_F > \epsilon\}$. Likewise, by noting that E_n is the same as $\{\|R_n - R_\infty\|_F > \epsilon\}$, we have

$$\begin{aligned}\sup_{f\in\mathcal{F}}\Big|\mathbb{E}[f(T_n^*, R_n)|\mathcal{D}_n)] - \mathbb{E}[f(T_n^*, R_\infty)|\mathcal{D}_n]\Big| \ &\leq \ \epsilon + 2\,\mathbb{P}\big(\|R_n - R_\infty\|_F > \epsilon\big|\mathcal{D}_n\big) \\ &= \ \epsilon + o_{\mathbb{P}}(1),\end{aligned} \tag{B.6}$$

where the second step follows from Assumption (B.3). Finally, since $\epsilon > 0$ can be taken arbitrarily small, this completes the proof. ☐

Remark. The proofs of the next two results are similar to the proofs of Lemmas 3 and 4 in Lopes et al. [2018], but there is an important distinction insofar as the current proofs handle row-sampling matrices — which were not addressed in that prior work.

Lemma 2. *Suppose that the conditions of Theorem 1 hold. Then, for any fixed matrix $M\in\mathbb{R}^{d\times d}$, there is a Gaussian random variable $Z(M)$ such that*

$$\mathcal{L}\Big(\langle\!\langle M, \sqrt{t_n}(\tilde{G}_n - G_n)\rangle\!\rangle\Big) \ \xrightarrow{d} \ \mathcal{L}(Z(M)).$$

Proof. Let $s_{1,n},\dots,s_{t_n,n}\in\mathbb{R}^n$ denote the rows of $\sqrt{t_n}S_n$, and observe the algebraic relation

$$\langle\!\langle M, \sqrt{t_n}(\tilde{G}_n - G_n)\rangle\!\rangle = \tfrac{1}{\sqrt{t_n}\,n}\sum_{i=1}^{t_n}\Big(\langle\!\langle M, A_n^\top s_i s_i^\top A_n\rangle\!\rangle - \langle\!\langle M, A_n^\top A_n\rangle\!\rangle\Big). \tag{B.7}$$

Hence, if we define $X_{i,n} = \frac{1}{n}\Big(s_{i,n}^\top A_n M A_n^\top s_{i,n} - \langle\!\langle M, A_n^\top A_n\rangle\!\rangle\Big)$ for each $i\in\{1,\dots,t_n\}$, then it is straightforward to check that these random variables have mean zero and satisfy

$$\langle\!\langle M, \sqrt{t_n}(\tilde{G}_n - G_n)\rangle\!\rangle \ = \ \tfrac{1}{\sqrt{t_n}}\sum_{i=1}^{t_n} X_{i,n}. \tag{B.8}$$

Since the variables $X_{1,n},\dots,X_{t_n,n}$ are i.i.d. for each n, but have distributions that may vary with n, we now apply the Lindeberg CLT for triangular arrays [van der Vaart, 2000, Prop. 2.27]. This result requires us to verify two conditions as $n\to\infty$. The first is that $\operatorname{var}(X_{1,n})$ converges to a finite limit, and the second is that

$$\mathbb{E}\Big[X_{1,n}^2 1\big\{|X_{1,n}| > \epsilon\sqrt{t_n}\big\}\Big] \ \to \ 0 \qquad \text{for every fixed } \epsilon > 0. \tag{B.9}$$

We will now show that $\operatorname{var}(X_{1,n})$ converges to a limit separately in the cases where S_n is a row-sampling matrix or a Gaussian random projection. In the row-sampling case, this follows directly from Assumption RS and the fact that

$$\operatorname{var}(X_{1,n}) = \operatorname{var}\Big(\tfrac{1}{n}s_{1,n}^\top A_n M A_n^\top s_{1,n}\Big) = \operatorname{var}(\tilde{r}_n^\top M \tilde{r}_n),$$

where $\tilde{r}_n$ is the first row of $\frac{1}{\sqrt{n}}\tilde{A}_n$. Meanwhile, in the case of Gaussian random projections, it is possible to explicitly calculate $\operatorname{var}(X_{1,n})$. Specifically, if $z \sim N(0, I_n)$ is a standard Gaussian vector and $Q \in \mathbb{R}^{n\times n}$ is fixed, then

$$\operatorname{var}(z^\top Q z) = 2\|Q\|_F^2, \tag{B.10}$$

which can be found in [Bai and Silverstein, 2010, eqn. 9.8.6]. Consequently, we have

$$\begin{aligned} \operatorname{var}(X_{1,n}) \;&=\; 2\big\|\tfrac{1}{n}A_n M A_n^\top\big\|_F^2 \\ &=\; 2\operatorname{tr}(M^\top G_n M G_n), \end{aligned} \tag{B.11}$$

and since Assumption RP ensures $G_n \to \mathsf{G}_\infty$, it follows that $\operatorname{var}(X_{1,n}) \to \operatorname{tr}(M^\top \mathsf{G}_\infty M \mathsf{G}_\infty)$, as needed.

To complete the proof, it remains to check the Lindeberg condition (B.9) in the cases of the two types of sketching matrices. First we handle the case of row sampling. Using the Cauchy-Schwarz inquality, followed by a Chernoff bound, we have

$$\mathbb{E}\Big[X_{1,n}^2 1\big\{|X_{1,n}| > \epsilon\sqrt{t_n}\big\}\Big] \;\le\; \sqrt{\mathbb{E}\big[X_{1,n}^4\big]\mathbb{E}\big[e^{|X_{1,n}|}\big]\, e^{-\epsilon\sqrt{t_n}}}. \tag{B.12}$$

Using the general inequality $(a+b)^4 \le 8(a^4+b^4)$, we may bound the fourth moment as

$$\begin{aligned} \mathbb{E}[X_{1,n}^4] \;&\le\; 8\sum_{l=1}^n p_l\big(\tfrac{1}{np_l}a_l^\top M a_l\big)^4 \;+\; 8\langle\!\langle M, G_n\rangle\!\rangle^4 \\ &\le\; 8\|M\|_{\text{op}}^4 \max_{1\le l\le n}\|\tfrac{1}{\sqrt{np_l}}a_l\|_2^8 \;+\; c_0, \end{aligned} \tag{B.13}$$

where we have used the fact that $\langle\!\langle M, G_n\rangle\!\rangle \le c_0$ for some positive constant $c_0 > 0$. Similarly, we have

$$\begin{aligned} \mathbb{E}\big[e^{|X_{1,n}|}\big]e^{-\epsilon\sqrt{t_n}} \;&\le\; \exp\Big(|\langle\!\langle M, G_n\rangle\!\rangle| - \epsilon\sqrt{t_n}\Big)\textstyle\sum_{l=1}^n p_l \exp\big(\big|\tfrac{1}{np_l}a_l^\top M a_l\big|\big) \\ &\le\; \exp\Big(c_0 - \epsilon\sqrt{t_n} + \|M\|_{\text{op}}\max_{1\le l\le n}\|\tfrac{1}{\sqrt{np_l}}a_l\|_2^2\Big). \end{aligned} \tag{B.14}$$

Hence, under Assumption RS, there is a constant $c(\epsilon) > 0$ such that the bound

$$\mathbb{E}\big[e^{|X_{1,n}|}\big]e^{-\epsilon\sqrt{t_n}} \;\le\; e^{c_0}e^{-c(\epsilon)\sqrt{t_n}}$$

holds for all large t. Combining with (B.13) and noting that the limit $\max_{1\le l\le n}\|\frac{1}{\sqrt{np_l}}a_l\|_2^8\, e^{-c(\epsilon)\sqrt{t_n}} \to 0$ holds under Assumption RS, it follows that the Lindeberg condition (B.9) indeed holds in the case of row sampling.

Lastly, to handle the case of Gaussian random projections, it follows from the Cauchy-Schwarz and Chebyshev inequalities that

$$\mathbb{E}\Big[X_{1,n}^2 1\big\{|X_{1,n}| > \epsilon\sqrt{t_n}\big\}\Big] \;\le\; \sqrt{\mathbb{E}\big[X_{1,n}^4\big]\tfrac{1}{\epsilon^2 t_n}\operatorname{var}(X_{1,n})}. \tag{B.15}$$

Next, it is known from [Bai and Silverstein, 2010, Lemma B.26] that if $s_{1,n}$ is a standard Gaussian vector in $\mathbb{R}^n$, then the following bound holds in terms of the matrix $K_n := \frac{1}{n}A_n M A_n^\top$ and an absolute constant $c > 0$,

$$\begin{aligned} \mathbb{E}[X_{1,n}^4] \;&\le\; c\Big(\operatorname{tr}(K_nK_n^\top)^2 \;+\; \operatorname{tr}\big((K_nK_n^\top)^2\big)\Big) \\ &=\; c\Big(\operatorname{tr}(MG_nM^\top G_n) \;+\; \operatorname{tr}\big((MG_nM^\top G_n)^2\big)\Big). \end{aligned} \tag{B.16}$$

In turn, since this bound converges to $\operatorname{tr}(M\mathsf{G}_\infty M^\top\mathsf{G}_\infty) + \operatorname{tr}\big((M\mathsf{G}_\infty M^\top\mathsf{G}_\infty)^2\big)$, we conclude that $\mathbb{E}[X_{1,n}^4]$ is bounded, and so (B.15) implies that the Lindeberg condition (B.9) holds in the case of Gaussian random projections. □

Lemma 3. *Suppose that the conditions of Theorem 1 hold, and for each fixed $M \in \mathbb{R}^{d\times d}$, let $Z(M)$ be the Gaussian random variable in the statement of Lemma 2. Then, the following limit holds as $n \to \infty$,*

$$\mathcal{L}\Big(\langle\!\langle M, \sqrt{t_n}(\tilde{G}_n^* - \tilde{G}_n)\rangle\!\rangle \Big| S_n\Big) \xrightarrow{d} \mathcal{L}(Z(M)), \qquad \textit{in probability.}$$

Proof. Recall that $s_{1,n}, \dots, s_{t_n,n} \in \mathbb{R}^n$ denote the rows of $\sqrt{t_n}S_n$, and let $s^*_{1,n}, \dots, s^*_{t_n,n}$ be t_n i.i.d. samples drawn with replacement from $s_{1,n}, \dots, s_{t_n,n}$. By analogy with the proof of Lemma 2, consider the algebraic relation

$$\langle\!\langle M, \sqrt{t_n}(\tilde{G}_n^* - \tilde{G}_n)\rangle\!\rangle = \tfrac{1}{\sqrt{t_n}\, n} \sum_{i=1}^{t_n} \Big(\langle\!\langle M, A_n^\top (s^*_{i,n})(s^*_{i,n})^\top A_n\rangle\!\rangle - \langle\!\langle M, \tilde{A}_n^\top \tilde{A}_n\rangle\!\rangle\Big), \tag{B.17}$$

and define the random variable $X^*_{i,n} = \frac{1}{n}\Big((s^*_{i,n})^\top A_n M A_n^\top (s^*_{i,n}) - \langle\!\langle M, \tilde{A}_n^\top \tilde{A}_n\rangle\!\rangle\Big)$ for each $i \in \{1, \dots, t_n\}$ so that

$$\langle\!\langle M, \sqrt{t_n}(\tilde{G}_n^* - \tilde{G}_n)\rangle\!\rangle \;=\; \tfrac{1}{\sqrt{t_n}} \sum_{i=1}^{t_n} X^*_{i,n}. \tag{B.18}$$

Also observe that $\mathbb{E}[X^*_{i,n}|S_n] = 0$ for every $i \in \{1, \dots, t_n\}$. To complete the proof, it suffices to show that the conditions of the Lindeberg CLT for triangular arrays hold in a conditional sense (cf. [van der Vaart, 2000, p.330-331]). More precisely, it suffices to show that the conditional variance $\mathrm{var}(\langle\!\langle M, \sqrt{t_n}(\tilde{G}_n^* - \tilde{G}_n)\rangle\!\rangle|S_n)$ converges to a limit in probability, and that the following limit holds for each fixed $\epsilon > 0$,

$$\mathbb{E}\Big[(X^*_{1,n})^2 1\big\{|X^*_{1,n}| > \epsilon\sqrt{t_n}\big\}\Big| S_n\Big] \;\to\; 0 \qquad \text{in probability.} \tag{B.19}$$

To show the latter condition in the cases of either the row-sampling or Gaussian random projections, let L_n denote the left side of (B.19) and note that the definition of sampling with replacement implies

$$L_n = \tfrac{1}{t_n}\textstyle\sum_{i=1}^{t_n} X_{i,n}^2 1\{X_{i,n} > \epsilon\sqrt{t_n}\}.$$

Consequently, for any $\epsilon > 0$, Markov's inequality gives

$$\mathbb{P}(L_n > \epsilon) \;\leq\; \tfrac{1}{\epsilon}\mathbb{E}[L_n] \;=\; \tfrac{1}{\epsilon}\mathbb{E}\big[X_{1,n}^2 1\{X_{1,n} > \epsilon\sqrt{t_n}\}\big], \tag{B.20}$$

and so the condition (B.19) must hold because the limit $\mathbb{E}\big[X_{1,n}^2 1\{X_{1,n} > \epsilon\sqrt{t_n}\}\big] \to 0$ was established in the proof of Lemma 2.

Finally, we show that $\mathrm{var}(\langle\!\langle M, \sqrt{t_n}(\tilde{G}_n^* - \tilde{G}_n)\rangle\!\rangle|S_n)$ has a limit in probability, and the argument will apply in the same manner to the cases of row sampling and Gaussian random projections. Define the random variable $\widehat{\varsigma}^2_{t_n} := \frac{1}{t_n}\sum_{i=1}^t X_{i,n}^2 - (\frac{1}{t_n}\sum_{i=1}^{t_n} X_{i,n})^2$, and observe that

$$\begin{aligned} \mathrm{var}\Big(\langle\!\langle M, \sqrt{t_n}(\tilde{G}_n^* - \tilde{G}_n)\rangle\!\rangle\Big| S_n\Big) \;&=\; \mathrm{var}\left(X^*_{1,n}|S_n\right) \\ &=\; \widehat{\varsigma}^2_{t_n}, \end{aligned} \tag{B.21}$$

which follows from the relation (B.18) and the fact that, conditionally on S_n, the random variable $X^*_{1,n}$ is a sample from the discrete uniform distribution on $\{X_{1,n}, \dots, X_{t_n,n}\}$. Thus, it remains to show that $\widehat{\varsigma}^2_{t_n}$ converges to a limit in probability. Due to basic facts about the sample variance $\widehat{\varsigma}^2_{t_n}$, it is known that $\mathbb{E}[\widehat{\varsigma}^2_{t_n}] = \frac{t_n - 1}{t_n}\mathrm{var}(X_{1,n})$ and $\mathrm{var}(\widehat{\varsigma}^2_{t_n}) = \mathcal{O}(\frac{1}{t_n}\mathbb{E}[X_{1,n}^4])$ [Kenney and Keeping, 1951, p. 164]. Furthermore, the proof of Lemma 2 shows that $\frac{1}{t_n}\mathbb{E}[X_{1,n}^4] \to 0$ under either Assumption RP or RS, and so it follows that $\widehat{\varsigma}^2_{t_n}$ must converge in probability to the same limit as $\mathrm{var}(X_{1,n})$, which completes the proof. □

C Lemmas for the right singular vectors

In Lemmas 4 and 5 we provide a joint CLT for the projection matrices $\sqrt{t_n}(\tilde{P}_j - P_j)$, as well as their bootstrap counterparts $\sqrt{t_n}(\tilde{P}_j^* - \tilde{P}_j)$ with $j = 1, \dots, k$. (Recall that the definitions of definitions of P_j, $\tilde{P}_j$, and $\tilde{P}_j^*$ are given in (A.6).)

Lemma 4. *Suppose that the conditions of Theorem 1 hold. Then, for any fixed matrices $M_1, \ldots, M_k \in \mathbb{R}^{d\times d}$, there is a Gaussian vector $(Z_1(M_1), \ldots, Z_k(M_k))$ in $\mathbb{R}^k$ such that*

$$\mathcal{L}\Big(\sqrt{t_n}\langle\!\langle \tilde{P}_1 - P_1, M_1\rangle\!\rangle, \ldots, \sqrt{t_n}\langle\!\langle \tilde{P}_k - P_k, M_k\rangle\!\rangle\Big) \xrightarrow{d} \mathcal{L}\big(Z_1(M_1), \ldots, Z_k(M_k)\big). \tag{C.1}$$

Furthermore, there is a choice of the matrix $M_1 \in \mathbb{R}^{d\times d}$ such that the Gaussian variable $Z_1(M_1)$ has positive variance.

Proof. By the Cramér-Wold theorem, the limit (C.1) can be established by showing that for any constants $c_1, \ldots, c_k \in \mathbb{R}$, the sum $\sum_{j=1}^k c_j\sqrt{t_n}\langle\!\langle \tilde{P}_j - P_j, M_j\rangle\!\rangle$ converges in distribution to a Gaussian random variable. We will show this first, and then at the end of the proof, we will exhibit a choice of M_1 for which $\text{var}(Z_1(M_1)) > 0$.

Recall that $\mathcal{S}^{d\times d} \subset \mathbb{R}^{d\times d}$ denotes the subspace of symmetric matrices, and for each $j \in \{1, \ldots, k\}$, let $\psi_j : \mathcal{S}^{d\times d} \to \mathbb{R}$ denote the function that satisfies $\psi_j(G_n) = \langle\!\langle P_j, M_j\rangle\!\rangle$ and $\psi_j(\tilde{G}_n) = \langle\!\langle \tilde{P}_j, M_j\rangle\!\rangle$, so that

$$\sqrt{t}\langle\!\langle \tilde{P}_j - P_j, M_j\rangle\!\rangle \;=\; \sqrt{t_n}(\psi_j(\tilde{G}_n) - \psi_j(G_n)).$$

To apply the mean-value theorem to the difference on the right, we may rely on the fact from matrix calculus that ψ_j is continuously differentiable in an open neighborhood of any symmetric matrix whose jth eigenvalue is isolated [Magnus and Neudecker, 2019, Theorem 8.9]. Since all the eigenvalues of G_∞ are isolated, we may let $\mathcal{U} \subset \mathcal{S}^{d\times d}$ denote an open neighborhood on which all the functions $\psi_1, \ldots, \psi_k$ are continuously differentiable. Also, we need to define a random variable $\tilde{R}_{n,j}$ and a random matrix $\tilde{D}_{j,n} \in \mathcal{S}^{d\times d}$ in the following two cases: either (1) both of the matrices G_n and $\tilde{G}_n$ lie in $\mathcal{U}$, or (2) at least one of the matrices G_n or $\tilde{G}_n$ falls outside of $\mathcal{U}$. In the first case, let $\tilde{R}_{n,j} = 0$, and let $\tilde{D}_{j,n}$ denote the differential (gradient) $\psi_j'(\breve{G}_{j,n}) \in \mathcal{S}^{d\times d}$ evaluated at a random matrix $\breve{G}_{j,n}$ that is a convex combination of G_n and $\tilde{G}_n$. In the second case, let $\tilde{R}_{n,j} = \sqrt{t_n}(\psi_j(\tilde{G}_n) - \psi_j(G_n))$ and let $\tilde{D}_{j,n} = 0$. Based on these definitions, the mean-value theorem ensures that the following relation always holds

$$\sqrt{t_n}\Big(\psi_j(\tilde{G}_n) - \psi_j(G_n)\Big) \;=\; \Big\langle\!\!\Big\langle \tilde{D}_{j,n}, \sqrt{t_n}(\tilde{G}_n - G_n)\Big\rangle\!\!\Big\rangle \;+\; \tilde{R}_{n,j}. \tag{C.2}$$

Hence, if we let $\tilde{D}_n = \sum_{j=1}^k c_j\tilde{D}_{j,n}$ and $\tilde{R}_n = \sum_{j=1}^k c_j\tilde{R}_{n,j}$, then

$$\sum_{j=1}^k c_j\sqrt{t_n}\langle\!\langle \tilde{P}_j - P_j, M_j\rangle\!\rangle \;=\; \Big\langle\!\!\Big\langle \tilde{D}_n, \sqrt{t_n}(\tilde{G}_n - G_n)\Big\rangle\!\!\Big\rangle + \tilde{R}_n. \tag{C.3}$$

Based on this relation, as well as Lemma 2 and Slutsky's lemma (Fact 1), the proof of the limit (C.1) reduces to establishing the following limits

$$\tilde{R}_n \to 0 \qquad \text{in probability,} \tag{C.4}$$

and

$$\tilde{D}_n \to \mathsf{D}_\infty \qquad \text{in probability,} \tag{C.5}$$

for some constant matrix $\mathsf{D}_\infty \in \mathcal{S}^{d\times d}$. These limits are established below.

Let $\mathcal{E}_n$ denote the event that both $\tilde{G}_n$ and G_n lie in the neighborhood $\mathcal{U}$. Given that $\tilde{R}_n$ can only be non-zero when $\mathcal{E}_n^c$ occurs, we have

$$\mathbb{P}(|\tilde{R}_n| > \epsilon) \;\leq\; \mathbb{P}(\mathcal{E}_n^c).$$

Furthermore, since we assume that G_n converges to G_∞, and since it is shown in Lemma 6 below that $\tilde{G}_n$ converges in probability to G_∞, it follows that $\mathbb{P}(\mathcal{E}_n^c) \to 0$. This establishes the limit (C.4). With regard to the limit (C.5), observe that our definitions give the relation

$$\tilde{D}_n = \sum_{j=1}^k c_j 1_{\mathcal{E}_n}\psi_j'(\breve{G}_{j,n}),$$

where $1_{\mathcal{E}_n}$ is the indicator of the event $\mathcal{E}_n$. Based on the mentioned limits of $\tilde{G}_n$ and G_n, it follows that the convex combination $\breve{G}_{j,n}$ must converge in probability to G_∞. In addition, since the differential ψ_j' is continuous on $\mathcal{U}$, it follows that $\psi_j'(\breve{G}_{j,n})$ converges in probability to the constant matrix $\psi_j'(\mathsf{G}_\infty)$. Hence, if we put

$$\mathsf{D}_\infty := \sum_{j=1}^{k} c_j \psi_j'(\mathsf{G}_\infty), \tag{C.6}$$

then the limit (C.5) holds, and the proof of (C.1) is complete.

Now, we turn to showing that $\operatorname{var}(Z_1(M_1)) > 0$. Let v_1 and v_2 denote the pair of eigenvectors of G_∞ mentioned in Assumption RS, and consider the particular choice of the matrix

$$M_1 = \mathsf{v}_2\mathsf{v}_1^\top.$$

Then, our previous argument leads to

$$\sqrt{t_n}\langle\!\langle \tilde{P}_1 - P_1, M_1\rangle\!\rangle \;=\; \langle\!\langle \psi_1'(\mathsf{G}_\infty), \sqrt{t_n}(\tilde{G}_n - G_n)\rangle\!\rangle \;+\; o_{\mathbb{P}}(1),$$

where we recall that the definition of ψ_1 depends on M_1. Based on an analytical formula for the matrix differential of an eigenprojection [Magnus and Neudecker, 2019, Theorem 8.9], the differential $\psi_1'(\mathsf{G}_\infty)$ can be obtained explicitly. In particular, the inner product on the right side above may be calculated as follows, where λ_1 and λ_2 denote the eigenvalues of G_∞ corresponding to v_1 and v_2, and the symbol $\dagger$ refers to the Moore-Penrose inverse,

$$\begin{aligned}\langle\!\langle \psi_1'(\mathsf{G}_\infty), \sqrt{t_n}(\tilde{G}_n - G_n)\rangle\!\rangle \;&=\; \Big\langle\!\!\Big\langle \mathsf{v}_1\mathsf{v}_1^\top M_1\big(\lambda_1 I_d - \mathsf{G}_\infty\big)^\dagger + \big(\lambda_1 I_d - \mathsf{G}_\infty\big)^\dagger M_1\mathsf{v}_1\mathsf{v}_1^\top \;,\; \sqrt{t_n}(\tilde{G}_n - G_n)\Big\rangle\!\!\Big\rangle, \\ &=\; \tfrac{1}{\lambda_1-\lambda_2}\langle\!\langle M_1^\top, \sqrt{t_n}(\tilde{G}_n - G_n)\rangle\!\rangle.\end{aligned} \tag{C.7}$$

where the second step is obtained by noting ${}^\top\mathsf{v}_1\mathsf{v}_1^\top M_1 = 0$, as well as $(\lambda_1 I_d - \mathsf{G}_\infty)^\dagger \mathsf{v}_2 = \frac{1}{\lambda_1-\lambda_2}\mathsf{v}_2$. In turn, by rearranging the last expression in (C.7), we have

$$\langle\!\langle \psi_1'(\mathsf{G}_\infty), \sqrt{t_n}(\tilde{G}_n - G_n)\rangle\!\rangle \;=\; \tfrac{1}{\lambda_1-\lambda_2}\,\tfrac{1}{\sqrt{t_n}}\sum_{i=1}^{t_n}\Big(s_{i,n}^\top\big(\tfrac{1}{n}A_n M_1^\top A_n^\top\big)s_{i,n} - \operatorname{tr}\big(\tfrac{1}{n}A_n M_1^\top A_n^\top\big)\Big) \tag{C.8}$$

and since the terms of the sum are mean-zero and i.i.d., we have

$$\operatorname{var}\Big(\langle\!\langle \psi_1'(\mathsf{G}_\infty), \sqrt{t_n}(\tilde{G}_n - G_n)\rangle\!\rangle\Big) \;=\; \tfrac{1}{(\lambda_1-\lambda_2)^2}\operatorname{var}\Big(s_{1,n}^\top\big(\tfrac{1}{n}A_n M_1^\top A_n^\top\big)s_{1,n}\Big).$$

Altogether, it remains to separately check that the right side of this display has a positive limit in the cases of row sampling and Gaussian random projections. In the row sampling case, this follows directly from Assumption RS since $\operatorname{var}(s_{1,n}^\top(\frac{1}{n}A_n M_1^\top A_n^\top)s_{1,n}) = \operatorname{var}(\tilde{r}_n^\top(\mathsf{v}_1\mathsf{v}_2^\top)\tilde{r}_n) \to \ell(\mathsf{v}_1\mathsf{v}_2^\top)$. Alternatively, in the case when S_n is a Gaussian random projection, we may use the formula (B.10) which leads to

$$\begin{aligned}\operatorname{var}\Big(s_{1,n}^\top\big(\tfrac{1}{n}A_n M_1^\top A_n^\top\big)s_{1,n}\Big) \;&=\; 2\Big\|\tfrac{1}{n}A_n M_1^\top A_n^\top\Big\|_F^2 \\ &=\; 2\operatorname{tr}\Big(M_1 G_n M_1^\top G_n\Big) \\ &=\; 2\operatorname{tr}\Big(M_1 \mathsf{G}_\infty M_1^\top \mathsf{G}_\infty\Big) + o(1) \\ &=\; 2\Big(\mathsf{v}_1^\top\mathsf{G}_\infty\mathsf{v}_1\Big)\Big(\mathsf{v}_2^\top\mathsf{G}_\infty\mathsf{v}_2\Big) + o(1) \\ &=\; 2\lambda_1\lambda_2 + o(1).\end{aligned} \tag{C.9}$$

This clearly leads to a positive limit for the variance of $\langle\!\langle \psi_1'(\mathsf{G}_\infty), \sqrt{t_n}(\tilde{G}_n - G_n)\rangle\!\rangle$, which completes the proof. $\square$

Lemma 5. *Suppose the conditions of Theorem 1 hold, and for any fixed matrices $M_1, \ldots, M_k \in \mathbb{R}^{d\times d}$, let $(Z_1(M_1), \ldots, Z_k(M_k))$ be the Gaussian vector in the statement of Lemma 4. Then, as $n \to \infty$,*

$$\mathcal{L}\Big(\sqrt{t_n}\langle\!\langle \tilde{P}_1^* - \tilde{P}_1, M_1\rangle\!\rangle, \ldots, \sqrt{t_n}\langle\!\langle \tilde{P}_k^* - \tilde{P}_k, M_k\rangle\!\rangle \,\Big|\, S_n\Big) \xrightarrow{d} \mathcal{L}\big(Z_1(M_1), \ldots, Z_k(M_k)\big) \quad \textit{in probability.} \tag{C.10}$$

Proof. The argument is similar to the proof of Lemma 4, but with some differences that we explain here. By the Cramér-Wold theorem, it suffices to show that the following limit holds for any fixed numbers $c_1, \ldots, c_k \in \mathbb{R}$,

$$\mathcal{L}\Big(\textstyle\sum_{j=1}^k c_j \sqrt{t_n}\langle\!\langle \tilde{P}_j^* - \tilde{P}_j, M_j\rangle\!\rangle \Big| S_n\Big) \xrightarrow{d} \mathcal{L}\Big(\textstyle\sum_{j=1}^k c_j Z_j(M_j)\Big) \qquad \text{in probability.}$$

To show this, we will combine Lemmas, 1, 2, and 4.

Let the fixed matrix $\mathsf{D}_\infty \in \mathbb{R}^{d\times d}$ denote the limit in (C.6) that depends on $c_1, \ldots, c_k$, and observe that the proof of Lemma 4 shows that

$$\mathcal{L}\Big(\big\langle\!\big\langle \mathsf{D}_\infty, \sqrt{t_n}(\tilde{G}_n - G_n)\big\rangle\!\big\rangle\Big) \xrightarrow{d} \mathcal{L}\Big(\textstyle\sum_{j=1}^k c_j Z_j(M_j)\Big).$$

Next, we claim that the following expansion holds

$$\textstyle\sum_{j=1}^k c_j \sqrt{t_n}\langle\!\langle \tilde{P}_j^* - \tilde{P}_j, M_j\rangle\!\rangle \;=\; \big\langle\!\big\langle \tilde{D}_n^*, \sqrt{t_n}(\tilde{G}_n^* - \tilde{G}_n)\big\rangle\!\big\rangle + \tilde{R}_n^* \tag{C.11}$$

where $\tilde{D}_n^* \in \mathbb{R}^{d\times d}$ is a random matrix and $\tilde{R}_n^*$ is a random scalar that satisfy the following limits for any fixed $\epsilon > 0$,

$$\mathbb{P}\Big(\|\tilde{D}_n^* - \mathsf{D}_\infty\|_F > \epsilon \,\Big|\, S_n\Big) \;\to\; 0 \quad \text{ in probability,} \tag{C.12}$$

$$\mathbb{P}\Big(|R_n^* - 0| > \epsilon \,\Big|\, S_n\Big) \;\to\; 0 \quad \text{ in probability.} \tag{C.13}$$

As a consequence of this claim, and the fact that $\mathcal{L}(\langle\!\langle \mathsf{D}_\infty, \sqrt{t_n}(\tilde{G}_n^* - \tilde{G}_n)\rangle\!\rangle | S_n)$ conditionally converges in distribution to the same limit as $\mathcal{L}(\langle\!\langle \mathsf{D}_\infty, \sqrt{t_n}(\tilde{G}_n - G_n)\rangle\!\rangle)$ (by Lemma 2), the proof will be completed by the conditional version of Slutsky's lemma (Lemma 1). Thus, it remains to verify the three parts (C.11), (C.12), and (C.13) of the claim.

For each $j \in \{1, \ldots, k\}$, let $\psi_j : \mathcal{S}^{d\times d} \to \mathbb{R}$ be as defined in the proof of Lemma 4, and let $\mathcal{U} \subset \mathcal{S}^{d\times d}$ again denote the open neighborhood of G_∞ on which all the functions $\psi_1, \ldots, \psi_k$ are continuously differentiable. Also, let $\mathcal{E}_n'$ denote the event that $\tilde{G}_n^*$ lies in $\mathcal{U}$, and recall that $\mathcal{E}_n$ denotes the event that both $\tilde{G}_n$ and G_n lie in $\mathcal{U}$. To establish the expansion (C.11), we now define $\tilde{D}_n^*$ and $\tilde{R}_n^*$ as follows. When $\mathcal{E}_n' \cap \mathcal{E}_n$ holds, the mean-value theorem ensures that for each $j \in \{1, \ldots, k\}$ we have

$$\begin{aligned} \sqrt{t_n}\langle\!\langle \tilde{P}_j^* - \tilde{P}_j, M_j\rangle\!\rangle \;&=\; \sqrt{t_n}\big(\psi_j(\tilde{G}_n^*) - \psi_j(\tilde{G}_n)\big) \\ &=\; \langle\!\langle \tilde{D}_{j,n}^*, \sqrt{t_n}(\tilde{G}_n^* - \tilde{G}_n)\rangle\!\rangle, \end{aligned} \tag{C.14}$$

where $\tilde{D}_{j,n}^*$ is a shorthand for the differential $\psi_j'(\breve{G}_{j,n}^*) \in \mathcal{S}^{d\times d}$ evaluated at a point $\breve{G}_{j,n}^*$ that is a convex combination of $\tilde{G}_{j,n}^*$ and $\tilde{G}_{j,n}$. Accordingly, when $\mathcal{E}_n' \cap \mathcal{E}_n$ holds, we define $\tilde{D}_n^* = \sum_{j=1}^k c_j \tilde{D}_{j,n}^*$ and $\tilde{R}_n^* = 0$. Oppositely, when the event $(\mathcal{E}_n' \cap \mathcal{E}_n)^c$ holds, we put $\tilde{D}_n^* = 0$ and $\tilde{R}_n^* = \sum_{j=1}^k c_j \sqrt{t_n}\big(\psi_j(\tilde{G}_n^*) - \psi_j(\tilde{G}_n)\big)$. Based on these definitions, it follows that the expansion (C.11) always holds.

Turning to the limit (C.13), the event $\{|\tilde{R}_n^* - 0| > \epsilon\}$ can only occur when $(\mathcal{E}_n' \cap \mathcal{E}_n)^c$ occurs. Consequently, a union bound gives

$$\mathbb{E}\Big[\mathbb{P}\Big(|R_n^* - 0| > \epsilon \,\Big|\, S_n\Big)\Big] \;\le\; \mathbb{P}((\mathcal{E}_n')^c) + \mathbb{P}(\mathcal{E}_n^c). \tag{C.15}$$

Furthermore, we know from the proof of Lemma 4 that $\mathbb{P}(\mathcal{E}_n^c) \to 0$, and also, it is straightforward to check that $\mathbb{P}((\mathcal{E}_n')^c) \to 0$. Thus, (C.15) implies (C.13) via Markov's inequality.

Lastly, handling the limit (C.12) can be reduced to showing $\tilde{G}_n^* \xrightarrow{\mathbb{P}} \mathsf{G}_\infty$ (unconditionally), which is done in Lemma 6 below. This is sufficient because the limit $\tilde{G}_n \xrightarrow{\mathbb{P}} \mathsf{G}_\infty$ (cf. Lemma 6) and the continuity of $\psi_j'(\cdot)$ on the neighborhood $\mathcal{U}$ imply $\psi_j'(\breve{G}_{j,n}^*) \xrightarrow{\mathbb{P}} \psi_j'(\mathsf{G}_\infty)$ for all $j \in \{1, \ldots, k\}$, which leads to $\mathbb{P}(\|\tilde{D}_n^* - \mathsf{D}_\infty\|_F > \epsilon) \to 0$. In turn, this implies (C.12) via Markov's inequality, and the proof is complete. $\square$

Lemma 6. *Suppose that the conditions of Theorem 1 hold. Then, the following limits hold as $n \to \infty$,*

$$\tilde{G}_n \xrightarrow{\mathbb{P}} \mathsf{G}_\infty \tag{C.16}$$

and

$$\tilde{G}_n^* \xrightarrow{\mathbb{P}} \mathsf{G}_\infty. \tag{C.17}$$

Proof. The first limit (C.16) is a direct consequence of Lemma 2. To handle the second limit (C.17), observe that Chebyshev's inequality (conditional on S_n) gives

$$\begin{aligned}
\mathbb{P}\big(\|\tilde{G}_n^* - \tilde{G}_n\|_F > \epsilon \big| S_n\big) &\leq \tfrac{1}{\epsilon^2}\mathbb{E}\big[\|\tilde{G}_n^* - \tilde{G}_n\|_F^2 \big| S_n\big] \\
&= \tfrac{1}{\epsilon^2}\mathbb{E}\Big[\big\|\tfrac{1}{t_n}\textstyle\sum_{i=1}^{t} \tfrac{1}{n}A_n^\top(s_{i,n}^*)(s_{i,n}^*)^\top A_n - \tfrac{1}{n}A_n^\top S_n^\top S_n A_n\big\|_F^2 \Big| S_n\Big] \\
&= \tfrac{1}{\epsilon^2 t_n}\mathbb{E}\Big[\big\|A_n^\top(s_{1,n}^*)(s_{1,n}^*)^\top A_n - \tfrac{1}{n}A_n^\top S_n^\top S_n A_n\big\|_F^2 \Big| S_n\Big] \\
&= \tfrac{1}{\epsilon^2 t_n}\tfrac{1}{t_n}\textstyle\sum_{i=1}^{t_n} \big\|\tfrac{1}{n}A_n^\top s_{i,n}s_{i,n}^\top A_n - \tfrac{1}{n}A_n^\top S_n^\top S_n A_n\big\|_F^2,
\end{aligned} \tag{C.18}$$

where the third line follows from the general fact that if $Y_1, \dots, Y_t$ are independent mean-zero random matrices, then $\mathbb{E}[\|Y_1 + \cdots + Y_{t_n}\|_F^2] = \sum_{i=1}^{t_n}\mathbb{E}[\|Y_i\|_F^2]$, and the fourth line follows from the definition of sampling with replacement. So, taking an expectation over S_n on both sides of the previous display, we have

$$\begin{aligned}
\mathbb{P}\big(\|\tilde{G}_n^* - \tilde{G}_n\|_F > \epsilon\big) &\leq \tfrac{1}{\epsilon^2 t_n}\mathbb{E}\Big[\big\|\tfrac{1}{n}A_n^\top s_{1,n}s_{1,n}^\top A_n - \tfrac{1}{n}A_n^\top S_n^\top S_n A_n\big\|_F^2\Big] \\
&\leq \tfrac{2}{\epsilon^2 t_n}\mathbb{E}\Big[\big\|\tfrac{1}{n}A_n^\top s_{1,n}s_{1,n}^\top A_n - G_n\big\|_F^2\Big] + \tfrac{2}{\epsilon^2 t_n}\mathbb{E}\Big[\big\|G_n - \tfrac{1}{n}A_n^\top S_n^\top S_n A_n\big\|_F^2\Big] \\
&= \big(\tfrac{2}{\epsilon^2 t_n} + \tfrac{2}{\epsilon^2 t_n^2}\big)\Big(\mathbb{E}\big[(\tfrac{1}{n}s_{1,n}^\top A_n A_n^\top s_{1,n})^2\big] - \|G_n\|_F^2\Big) \\
&= \big(\tfrac{2}{\epsilon^2 t_n} + \tfrac{2}{\epsilon^2 t_n^2}\big)\Big(\operatorname{var}\big(\tfrac{1}{n}s_{1,n}^\top A_n A_n^\top s_{1,n}\big) + \big(\operatorname{tr}(G_n)^2 - \|G_n\|_F^2\big)\Big),
\end{aligned} \tag{C.19}$$

where the third line relies on expanding $A_n S_n^\top S_n A_n$ as a sum and using the identity $\mathbb{E}[\|Y_1 + \cdots + Y_{t_n}\|_F^2] = \sum_{i=1}^{t_n}\mathbb{E}[\|Y_i\|_F^2]$ that was mentioned just a moment ago. Finally, the proof of Lemma 2 shows that the quantity $\operatorname{var}(\frac{1}{n}s_{1,n}^\top A_n A_n^\top s_{1,n}) + (\operatorname{tr}(G_n)^2 - \|G_n\|_F^2)$ converges to a finite limit (under either Assumption RP or RS). Hence, the $\mathcal{O}(1/t_n)$ prefactor requires $\mathbb{P}\big(\|\tilde{G}_n^* - \tilde{G}_n\|_F > \epsilon\big)$ to converge to 0, as needed. □

D Lemma for the left singular vectors

The following lemma is needed for proving the limit (0.1) in Theorem 1.

Lemma 7. *Suppose that the conditions of Theorem 1 hold, and for each $j \in \{1, \dots, k\}$, let $\tilde{\Delta}_j$ and $\tilde{\Delta}_j^*$ be as defined in* (A.17) *and* (A.19). *Then, for any fixed matrices $M_1, \dots, M_k \in \mathbb{R}^{d\times d}$, there is an associated Gaussian random vector $(\zeta_1(M_1), \dots, \zeta_k(M_k)) \in \mathbb{R}^k$ such that*

$$\mathcal{L}\Big(\langle\!\langle \sqrt{t_n}\tilde{\Delta}_1, M_1\rangle\!\rangle, \dots, \langle\!\langle \sqrt{t_n}\tilde{\Delta}_k, M_k\rangle\!\rangle\Big) \xrightarrow{d} \mathcal{L}(\zeta_1(M_1), \dots, \zeta_k(M_k)), \quad \textit{and} \tag{D.1}$$

$$\mathcal{L}\Big(\sqrt{t_n}\tilde{\Delta}_1^*, M_1\rangle\!\rangle, \dots, \langle\!\langle \sqrt{t_n}\tilde{\Delta}_k^*, M_k\rangle\!\rangle \Big| S_n\Big) \xrightarrow{d} \mathcal{L}(\zeta_1(M_1), \dots, \zeta_k(M_k)) \quad \textit{in probability.} \tag{D.2}$$

Furthermore, there is a choice of the matrix $M_1 \in \mathbb{R}^{d\times d}$ such that the Gaussian variable $\zeta_1(M_1)$ has positive variance.

Proof. It is straightforward to check that the algebraic relation

$$\langle\!\langle \sqrt{t_n}\tilde{\Delta}_j, M_j \rangle\!\rangle \;=\; \langle\!\langle \sqrt{t_n}(\tilde{P}_j - P_j), \tilde{K}_j \rangle\!\rangle \tag{D.3}$$

holds for every $j \in \{1, \dots, k\}$, where we let

$$\tilde{K}_j = \tfrac{G_n^{1/2} M_j G_n^{1/2}}{\operatorname{tr}(\tilde{P}_j G_n)} - \tfrac{\langle\!\langle P_j, G_n^{1/2} M_j G_n^{1/2} \rangle\!\rangle}{\operatorname{tr}(\tilde{P}_j G_n)\operatorname{tr}(P_j G_n)} G_n.$$

In addition, since G_∞ has isolated eigenvalues, the function that maps a symmetric matrix to its jth eigenprojection is continuous in an open neighborhood of G_∞ [Magnus and Neudecker, 2019, Theorem 8.9]. Next, if we let $(\lambda_1, \mathsf{v}_1), \dots, (\lambda_d, \mathsf{v}_d)$ denote the eigenvalue-eigenvectors pairs of G_∞ with corresponding eigenprojections $\mathsf{P}_j := \mathsf{v}_j \mathsf{v}_j^\top$, then the limits

$$P_j = \mathsf{P}_j + o(1) \qquad \text{and} \qquad \tilde{P}_j = \mathsf{P}_j + o_{\mathbb{P}}(1),$$

follow from $G_n = \mathsf{G}_\infty + o(1)$ and $\tilde{G}_n = \mathsf{G}_\infty + o_{\mathbb{P}}(1)$ (by Lemma 6). (Also recall that the dependence of P_j and $\tilde{P}_j$ on n is suppressed.) In turn, we have

$$\tilde{K}_j = \tfrac{1}{\lambda_j} \mathsf{G}_\infty^{1/2} M_j \mathsf{G}_\infty^{1/2} - \tfrac{1}{\lambda_j} \langle\!\langle \mathsf{P}_j, M_j \rangle\!\rangle \mathsf{G}_\infty + o_{\mathbb{P}}(1).$$

So, if we combine this expression for $\tilde{K}_j$ with (D.3), Lemma 4, and Slutsky's lemma, it follows that the limit (D.1) holds.

To show there is a choice of M_1 for which $\langle\!\langle \sqrt{t_n}\tilde{\Delta}_1, M_1 \rangle\!\rangle$ converges to a distribution with a positive variance, consider the choice $M_1 = \mathsf{v}_2 \mathsf{v}_1^\top$. In this case, it can be checked that

$$\tilde{K}_1 = \tfrac{\sqrt{\lambda_2}}{\sqrt{\lambda_1}} M_1 + o_{\mathbb{P}}(1),$$

and in the proof of Lemma 4 it is shown that the limiting distribution of $\langle\!\langle \sqrt{t_n}(\tilde{P}_j - P_j), M_1 \rangle\!\rangle$ has positive variance (under either Assumption RS or RP). Thus, by Slutsky's lemma, the random variable $\langle\!\langle \sqrt{t_n}\tilde{\Delta}_1, M_1 \rangle\!\rangle$ must also have a limiting distribution with positive variance.

Lastly, it remains to establish the limit (D.2). For each $j \in \{1, \dots, k\}$, define the random matrix

$$\tilde{K}_j^* = \tfrac{\tilde{G}_n^{1/2} M_j \tilde{G}_n^{1/2}}{\operatorname{tr}(\tilde{P}_j^* \tilde{G}_n)} - \tfrac{\langle\!\langle \tilde{P}_j, \tilde{G}_n^{1/2} M_j \tilde{G}_n^{1/2} \rangle\!\rangle}{\operatorname{tr}(\tilde{P}_j^* \tilde{G}_n)\operatorname{tr}(\tilde{P}_j \tilde{G}_n)} \tilde{G}_n, \tag{D.4}$$

which leads to the algebraic relation

$$\langle\!\langle \sqrt{t_n}\tilde{\Delta}_j^*, M_j \rangle\!\rangle \;=\; \langle\!\langle \sqrt{t_n}(\tilde{P}_j^* - \tilde{P}_j), \tilde{K}_j^* \rangle\!\rangle. \tag{D.5}$$

In addition, using the reasoning that led to the limit $\tilde{P}_j = \mathsf{P}_j + o_{\mathbb{P}}(1)$ and the fact that $\tilde{G}_n^* = \mathsf{G}_\infty + o_{\mathbb{P}}(1)$ (by Lemma 6), it can be checked that $\tilde{P}_j^* = \mathsf{P}_j + o_{\mathbb{P}}(1)$, which leads to

$$\tilde{K}_j^* \;=\; \tfrac{1}{\lambda_j} \mathsf{G}_\infty^{1/2} M_j \mathsf{G}_\infty^{1/2} - \tfrac{1}{\lambda_j} \langle\!\langle \mathsf{P}_j, M_j \rangle\!\rangle \mathsf{G}_\infty + o_{\mathbb{P}}(1). \tag{D.6}$$

Finally, by combining this with the relation (D.5), Lemma 5, and the conditional of version of Slutsky's lemma (Lemma 1), it follows that the limit (D.2) holds. ☐

E Lemma for the singular values

The following lemma gives a joint CLT for $\sqrt{t_n}(\sigma_j(\tilde{A}_n) - \sigma_j(A_n))$ with $j = 1, \dots, k$, as well as for the bootstrap counterparts $\sqrt{t_n}(\sigma_j(\tilde{A}_n^*) - \sigma_j(\tilde{A}_n))$.

Lemma 8. *Suppose that the conditions of Theorem 1 hold. Then, for any fixed real numbers $c_1, \dots, c_k$, there is a Gaussian random variable $\zeta(c_1, \dots, c_k)$ such that as $n \to \infty$,*

$$\mathcal{L}\Big(\textstyle\sum_{j=1}^{k} \frac{\sqrt{t_n}}{\sqrt{n}} c_j(\sigma_j(\tilde{A}_n) - \sigma_j(A_n))\Big) \ \xrightarrow{d} \ \mathcal{L}(\zeta(c_1, \dots, c_k)), \tag{E.1}$$

and

$$\mathcal{L}\Big(\textstyle\sum_{j=1}^{k} \frac{\sqrt{t_n}}{\sqrt{n}} c_j(\sigma_j(\tilde{A}_n^*) - \sigma_j(\tilde{A}_n))\Big| S_n\Big) \ \xrightarrow{d} \ \mathcal{L}(\zeta(c_1, \dots, c_k)) \quad \textit{in probability.} \tag{E.2}$$

Lastly, the random variable $\zeta(1, 0, \dots, 0)$ has positive variance.

Proof. First, we prove the limit (E.1). For any fixed positive semidefinite matrix $M \in \mathcal{S}^{d\times d}$ and index $j \in \{1, \dots, k\}$, let $\varphi_j(M) = \sqrt{\lambda_j(M)}$ so that we have the relation

$$\tfrac{\sqrt{t_n}}{\sqrt{n}}(\sigma_j(\tilde{A}_n) - \sigma_j(A_n)) \ = \ \sqrt{t_n}\big(\varphi_j(\tilde{G}_n) - \varphi_j(G_n)\big).$$

Since the limiting matrix G_∞ has isolated eigenvalues, it is a fact from matrix calculus that there is an open neighborhood $\mathcal{V} \subset \mathcal{S}^{d\times d}$ of G_∞ such that all of the functions $\varphi_1, \dots, \varphi_k$ are continuously differentiable on $\mathcal{V}$ [Magnus and Neudecker, 2019, Theorem 8.9]. Consequently, if we let $\varphi_j'(\mathsf{G}_\infty) \in \mathcal{S}^{d\times d}$ denote the differential of φ_j at G_∞, and define the matrix $\mathsf{J}_\infty = \sum_{j=1}^{k} c_j\, \varphi_j'(\mathsf{G}_\infty)$, then the argument in the proof of Lemma 4 can be re-used to show that the following expansion holds

$$\textstyle\sum_{j=1}^{k} \frac{\sqrt{t_n}}{\sqrt{n}} c_j(\sigma_j(\tilde{A}_n) - \sigma_j(A_n)) \ = \ \big\langle\!\big\langle \mathsf{J}_\infty, \sqrt{t_n}(\tilde{G}_n - G_n)\big\rangle\!\big\rangle + o_{\mathbb{P}}(1).$$

In turn, it follows directly from Lemma 2 that the limit (E.1) holds.

Next, to handle the limit (E.2), an argument that is analogous to the proof of Lemma 5 can be used. In particular, the same reasoning can be used to show that there is a random matrix $\tilde{J}_n^* \in \mathcal{S}^{d\times d}$ and a remainder variable $\tilde{W}_n^*$, such that the following equation holds

$$\textstyle\sum_{j=1}^{k} \frac{\sqrt{t_n}}{\sqrt{n}} c_j(\sigma_j(\tilde{A}_n^*) - \sigma_j(\tilde{A}_n)) \ = \ \big\langle\!\big\langle \tilde{J}_n^*, \sqrt{t_n}(\tilde{G}_n^* - \tilde{G}_n)\big\rangle\!\big\rangle + \tilde{W}_n^*. \tag{E.3}$$

In addition, for any $\epsilon > 0$, the limits $\mathbb{P}\big(\|\tilde{J}_n^* - \mathsf{J}_\infty\|_F > \epsilon \big| S_n\big) = o_{\mathbb{P}}(1)$ and $\mathbb{P}(|\tilde{W}_n^* - 0| > \epsilon|S_n) = o_{\mathbb{P}}(1)$ can be shown to hold as well. Thus, the conditional version of Slutsky's lemma (Lemma 1), combined with Lemma 5, lead to the the desired limit (E.2).

Lastly, to prove that $\zeta(1, 0, \dots, 0)$ has positive variance, it suffices to show that the random variable $\langle\!\langle \varphi_1'(\mathsf{G}_\infty), \sqrt{t_n}(\tilde{G}_n - G_n)\rangle\!\rangle$ converges to a Gaussian random variable with positive variance. Using an analytical formula for the differential $\varphi_1'(\mathsf{G}_\infty)$ available in [Magnus and Neudecker, 2019, Theorem 8.9], we have

$$\begin{aligned} \big\langle\!\big\langle \varphi_1'(\mathsf{G}_\infty), \sqrt{t_n}(\tilde{G}_n - G_n)\big\rangle\!\big\rangle \ &= \ \big\langle\!\big\langle \tfrac{1}{2\sqrt{\lambda_1}} \mathsf{v}_1\mathsf{v}_1^\top, \sqrt{t_n}(\tilde{G}_n - G_n)\big\rangle\!\big\rangle \\ &= \ \tfrac{1}{2\sqrt{\lambda_1}} \tfrac{1}{\sqrt{t_n}} \sum_{i=1}^{t_n} \Big(s_{i,n}^\top\big(\tfrac{1}{n}A_n\mathsf{v}_1\mathsf{v}_1^\top A_n^\top\big)s_{i,n} - \operatorname{tr}\big(\tfrac{1}{n}A_n\mathsf{v}_1\mathsf{v}_1^\top A_n^\top\big)\Big). \end{aligned} \tag{E.4}$$

Since the last sum consists of i.i.d. zero-mean random variables, the variance of the limiting Gaussian distribution will be positive if the sequence $\operatorname{var}(s_{1,n}^\top\big(\frac{1}{n}A_n\mathsf{v}_1\mathsf{v}_1^\top A_n^\top\big)s_{1,n})$ has a positive limit. In the case of a row-sampling sketching matrix, the positive limit follows from Assumption RS, and in the case of a Gaussian random projection, this can be verified by essentially repeating the calculation (C.9). □

F Examples satisfying Assumptions RP and RS

Example 1. For any positive definite matrix $\mathsf{G}_\circ \in \mathbb{R}^{d\times d}$, we may define an associated sequence of vectors $a_1, a_2, \dots$ in $\mathbb{R}^d$ as follows. If $1 \le l \le d$, define a_l to be the lth row of the matrix $\sqrt{d}\mathsf{G}_\circ^{1/2}$, and if $l > d$, define the successive vectors in a cyclical manner, $a_{d+1} = a_1, a_{d+2} = a_2, \dots, a_{2d} = a_d, a_{2d+1} = a_1$, and so on. In this notation, let the rows of $A_n \in \mathbb{R}^{n\times d}$ consist of the first n such vectors. When n is an exact multiple of d, we have

$$\tfrac{1}{n}A_n^\top A_n = \tfrac{1}{d}\textstyle\sum_{l=1}^{d} a_l a_l^\top \ = \ \mathsf{G}_\circ.$$

More generally, as $n \to \infty$ with d held fixed, it is simple to check that $\frac{1}{n}A_n^\top A_n \to \mathsf{G}_\circ$, and so the matrix $\mathsf{G}_\circ$ plays the role of G_∞ in this example. Likewise, any choice of $\mathsf{G}_\circ$ whose eigenvalues $\lambda_1(\mathsf{G}_\circ), \dots, \lambda_d(\mathsf{G}_\circ)$ all have multiplicity 1 will ensure that Assumption RP holds.

Next, we consider Assumption RS. It is straightforward to check that for any fixed n and fixed $C \in \mathbb{R}^{d\times d}$ we have

$$\operatorname{var}(\tilde{r}_n^\top C \tilde{r}_n) = \sum_{l=1}^n p_l \big(\tfrac{1}{np_l} a_l^\top C a_l\big)^2 - \Big(\tfrac{1}{n}\sum_{l=1}^n a_l^\top C a_l\Big)^2. \tag{F.1}$$

If we focus on the case of uniform sampling with $p_l = 1/n$ for all $l \in \{1, \dots, n\}$, then we have the following limit as $n \to \infty$,

$$\operatorname{var}(\tilde{r}_n^\top C \tilde{r}_n) \;\to\; \ell(C) = \tfrac{1}{d}\sum_{l=1}^d (a_l^\top C a_l)^2 - \Big(\tfrac{1}{d}\sum_{l=1}^d a_l^\top C a_l\Big)^2.$$

To consider the choices $C = \mathsf{v}_1\mathsf{v}_1^\top$ or $C = \mathsf{v}_1\mathsf{v}_2^\top$ in the case of uniform sampling, we may use the algebraic identity

$$a_l^\top \mathsf{v}_1 = \sqrt{d} e_l^\top \mathsf{G}_\circ^{1/2}\mathsf{v}_1 = \sqrt{d\lambda_1(\mathsf{G}_\circ)} e_l^\top \mathsf{v}_1 \tag{F.2}$$

to evaluate the limit function $\ell(\cdot)$ as

$$\ell(\mathsf{v}_1\mathsf{v}_1^\top) = \lambda_1^2(\mathsf{G}_\circ)\Big(d\sum_{l=1}^d (e_l^\top \mathsf{v}_1)^4 - 1\Big) \quad \text{and} \quad \ell(\mathsf{v}_1\mathsf{v}_2^\top) = \lambda_1(\mathsf{G}_\circ)\lambda_2(\mathsf{G}_\circ) d \sum_{l=1}^d (e_l^\top \mathsf{v}_1)^2 (e_l^\top \mathsf{v}_2)^2.$$

Based on these formulas, it follows that we have $\ell(\mathsf{v}_1\mathsf{v}_1^\top) > 0$ and $\ell(\mathsf{v}_1\mathsf{v}_2^\top) > 0$ under two rather generic conditions: First, we have $\ell(\mathsf{v}_1\mathsf{v}_1^\top) > 0$ as long as v_1 is not parallel to a vector of the form $(\frac{\pm 1}{\sqrt{d}}, \dots, \frac{\pm 1}{\sqrt{d}})$, which can be checked by noting that the Cauchy-Schwarz inequality

$$1 = \sum_{l=1}^d 1 \cdot (e_l^\top \mathsf{v}_1)^2 \leq \sqrt{d}\sqrt{\sum_{l=1}^d (e_l^\top \mathsf{v}_1)^4}$$

holds with equality precisely when v_1 is parallel to a vector of the stated form. Second, we have $\ell(\mathsf{v}_1\mathsf{v}_2^\top) > 0$ as long as there is at least one coordinate in $\{1, \dots, d\}$ where v_1 and v_2 are both non-zero. Lastly, since the set of values $\{\|a_1\|_2, \dots, \|a_d\|_2\}$ is fixed with respect to n, it follows that in the case of uniform sampling, the growth condition $\max_{1\leq l\leq n} \frac{1}{\sqrt{np_l}}\|a_l\|_2 = o(t_n^{1/8})$ is satisfied as well.

Example 2. Although Theorem 1 is based on a framework in which the matrix $A_n \in \mathbb{R}^{n\times d}$ is deterministic, it is of interest to know if Assumptions RP or RS are likely to hold for particular realizations of A_n generated at random by "nature". Likewise, for the purposes of this example only, the matrix A_n will be treated as being generated independently of all sources of algorithmic randomness. Furthermore, since the input matrix to an SVD algorithm is often viewed as having rows that represent data points in $\mathbb{R}^d$, we will consider the case where the rows of A_n are i.i.d. samples from a centered elliptical distribution (cf. Cambanis et al. [1981]).* In detail, this means that each vector a_i can be expressed in the form $a_i = \sqrt{d}\nu_i \mathsf{G}_\circ^{1/2} U_i$, where $\mathsf{G}_\circ \in \mathbb{R}^{d\times d}$ is a fixed positive definite matrix with isolated eigenvalues, and the pairs $(\nu_1, U_1), (\nu_2, U_2), \dots$ are i.i.d. elements in $\mathbb{R} \times \mathbb{R}^d$. In addition, each U_i is uniformly distributed on the unit ℓ_2-sphere, and each ν_i is a non-negative random variable (independent of U_i) with a finite moment generating function and $\mathbb{E}[\nu_i^2] = 1$.

To consider Assumption RP, we will use the relation $\mathbb{E}[a_l a_l^\top] = \mathsf{G}_\circ$, which follows from the fact that $\mathbb{E}[U_i U_i^\top] = \frac{1}{d} I_d$. Therefore, if we apply the law of large numbers to $\frac{1}{n}A_n^\top A_n = \frac{1}{n}\sum_{l=1}^n a_l a_l^\top$, then we have the limit $\frac{1}{n}A_n^\top A_n \to \mathsf{G}_\circ$ in probability. Thus, Assumption RP holds in probability.

Next, to consider Assumption RS in the case of uniform sampling, the formula (F.1) gives

$$\operatorname{var}(\tilde{r}_n^\top C \tilde{r}_n | A_n) = \tfrac{1}{n}\sum_{l=1}^n \big(a_l^\top C a_l\big)^2 - \Big(\tfrac{1}{n}\sum_{l=1}^n a_l^\top C a_l\Big)^2. \tag{F.3}$$

In order to verify Assumption RS in probability, we will first show that $\operatorname{var}(\tilde{r}_n^\top C\tilde{r}_n|A_n)$ converges in probability to a constant $\ell(C)$. Using a known formula for the variance of quadratic forms involving elliptical random vectors [Hu et al., 2019, Lemma A.1], as well as the law of large numbers, we have

$$\operatorname{var}(\tilde{r}_n^\top C \tilde{r}_n | A_n) \;\to\; \ell(C)$$

*Distributions of this type are commonly used in multivariate data analysis.

in probability, where the limit is given by

$$\ell(C) \;=\; \tfrac{\mathbb{E}[\nu_1^4]}{1+2/d}\Big(\operatorname{tr}(C\mathsf{G}_\circ)^2 + \operatorname{tr}(\mathsf{G}_\circ C\mathsf{G}_\circ C^\top) + \operatorname{tr}(\mathsf{G}_\circ C\mathsf{G}_\circ C)\Big) - \operatorname{tr}(C\mathsf{G}_\circ)^2.$$

In turn, this formula for $\ell(C)$ implies

$$\ell(\mathsf{v}_1\mathsf{v}_1^\top) = \lambda_1^2(\mathsf{G}_\circ)\Big(\tfrac{3\mathbb{E}[\nu_1^4]}{1+2/d} - 1\Big) \qquad \text{and} \qquad \ell(\mathsf{v}_1\mathsf{v}_2^\top) = \lambda_1\lambda_2\tfrac{\mathbb{E}[\nu_1^4]}{1+2/d}, \tag{F.4}$$

where the positivity of $\ell(\mathsf{v}_1\mathsf{v}_2^\top)$ is clear, and the positivity of $\ell(\mathsf{v}_1\mathsf{v}_1^\top)$ holds when $d \geq 2$, because $\mathbb{E}[\nu_1^4] \geq (\mathbb{E}[\nu_1^2])^2 = 1$. Lastly, to verify the growth condition involving $\max_{1\leq l\leq n} \frac{1}{\sqrt{np_l}}\|a_l\|_2$, recall that $p_l = 1/n$ for uniform sampling, and that ν_i is assumed to have a moment generating function. Under these conditions, it follows from [van der Vaart and Wellner, 1996, Lemma 2.2.2] that $\mathbb{E}[\max_{1\leq l\leq n}\|a_l\|_2] = \mathcal{O}(\log(n))$, which implies that $t_n^{-1/8}\max_{1\leq l\leq n}\|a_l\|_2 \to 0$ in probability under the mild condition $\log(n)t_n^{-1/8} \to 0$.

G Supplementary Experiments

In this section, we provide three sets of experiments that go beyond the settings considered in the main text. Subsequently, we provide additional results about running times.

G.1 Subsampled Randomized Hadamard Transform (SRHT)

Here, we revisit the synthetic example that we introduced in Section 5.1 to demonstrate that the bootstrap error estimation also works for other sampling schemes. Specifically, we consider the subsampled randomized hadamard transform (SRHT) Ailon and Chazelle [2006] to construct the sketch of A.

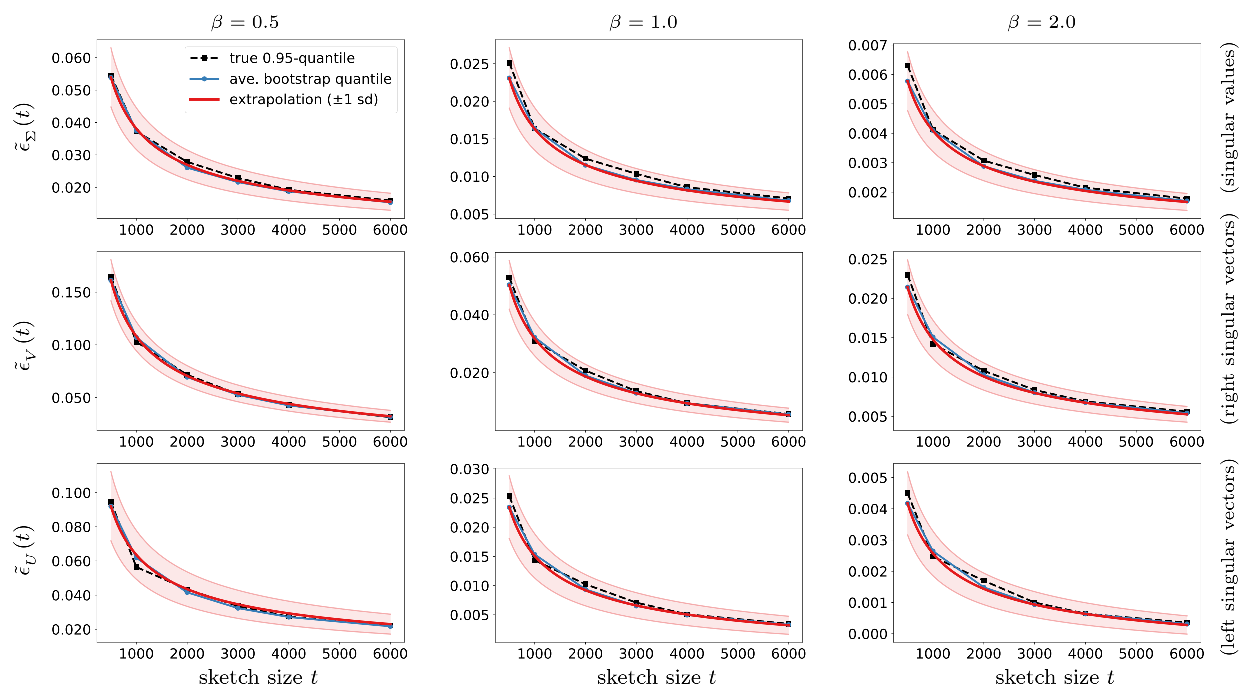

Figure 1: We consider matrices of dimension $(n, d) = (10^5, 3\times 10^3)$ that have singular value decay profiles of the form $\sigma_j = j^{-\beta}$ for $j \in \{1, \ldots, d\}$ with $\beta \in \{0.5, 1.0, 2.0\}$. The error variables correspond to the index set $\mathcal{J} = \{1\}$, and the simulations involve 500 trials and 30 bootstraps per trial. The rows correspond to the error quantiles for the singular values (top), right singular vectors (middle), and left singular vectors (bottom). The labelling scheme of the curves is the same as in the main text.

Figure 1 shows that the bootstrap quantile estimates, as well as their extrapolated versions, are good approximations to the true quantiles over the entire range of t. This behavior is also consistent across the different decay parameters $\beta = \{0.5, 1.0, 2.0\}$.

G.2 Synthetic matrices with alternative decay profile

As an alternative to the singular value decay profile $\sigma_j = j^{-\beta}$ with parameter values $\beta \in \{0.5, 1, 2\}$ used for the experiments of the main text, we now look at a profile of the form $\sigma_j = 10^{-\gamma j}$, with parameter values $\gamma \in \{0.05, 0.1, 0.5\}$. In particular, this type of decay profile arises in many applications related to differential equations and dynamical systems. Apart from the change in the decay profile, the experiments here were organized in the same manner as those for the synthetic matrices in the main text, and the results are plotted in the same format.

Figure 2 shows how close the bootstrap quantile estimates are to the true quantiles $q_U(t)$, $q_\Sigma(t)$, and $q_V(t)$ for sketch sizes $t \in \{500, \ldots, 6000\}$. Overall, the plots show that the bootstrap estimates are quite accurate, and in essence, the results for the current setting are on par with those shown in the main text.

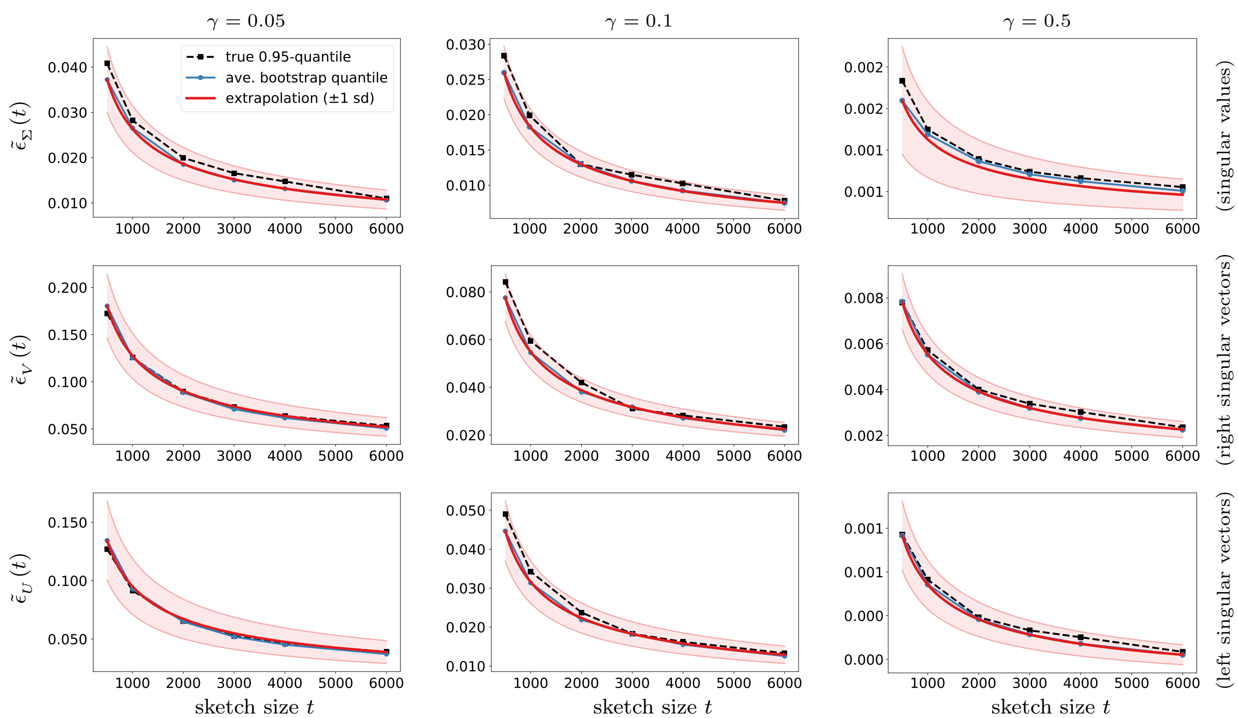

Figure 2: We consider artificial matrices of dimension $(n, d) = (10^5, 3 \times 10^3)$ that have singular value decay profiles of the form $\sigma_j = 10^{-\gamma j}$ for $j \in \{1, \ldots, d\}$ with $\gamma \in \{0.05, 0.1, 0.5\}$. The simulations involve 500 trials and 30 bootstraps per trial. The rows correspond to the error quantiles for the singular values (top), right singular vectors (middle), and left singular vectors (bottom). The labelling scheme of the curves is the same as in the main text.

G.3 Results for the index set $\mathcal{J} = \{1, 2, 3\}$

Recall that in our experiments, the sketching error variables are defined with respect to an index set $\mathcal{J} \subset \{1, \dots, k\}$ according to

$$\tilde{\epsilon}_U(t) = \max_{j \in \mathcal{J}} \rho_{\sin}(\tilde{u}_j, u_j)$$

$$\tilde{\epsilon}_V(t) = \max_{j \in \mathcal{J}} \rho_{\sin}(\tilde{v}_j, v_j),$$

$$\tilde{\epsilon}_\Sigma(t) = \max_{j \in \mathcal{J}} |\tilde{\sigma}_j - \sigma_j|.$$

Whereas the experiments in the main text considered the sketching errors for the leading triple (u_1, σ_1, v_1) corresponding to $\mathcal{J} = \{1\}$, we now look at the case when $\mathcal{J} = \{1, 2, 3\}$. In other words, the new experiments in this section correspond to a situation where the user would like to have *simultaneous* control over the sketching errors associated the top three singular vectors/values. Apart from this change in the choice of $\mathcal{J}$, all other aspects of the design and presentation of the experiments remain the same as in the main text. Given that a maximum is now being taken over a larger set of indices, the magnitudes of $\tilde{\epsilon}_\Sigma(t)$, $\tilde{\epsilon}_U(t)$, and $\tilde{\epsilon}_V(t)$ will necessarily be larger. Nevertheless, the important point to notice is that the quality of the bootstrap quantile estimates remains essentially as good as in the case when $\mathcal{J} = \{1\}$.

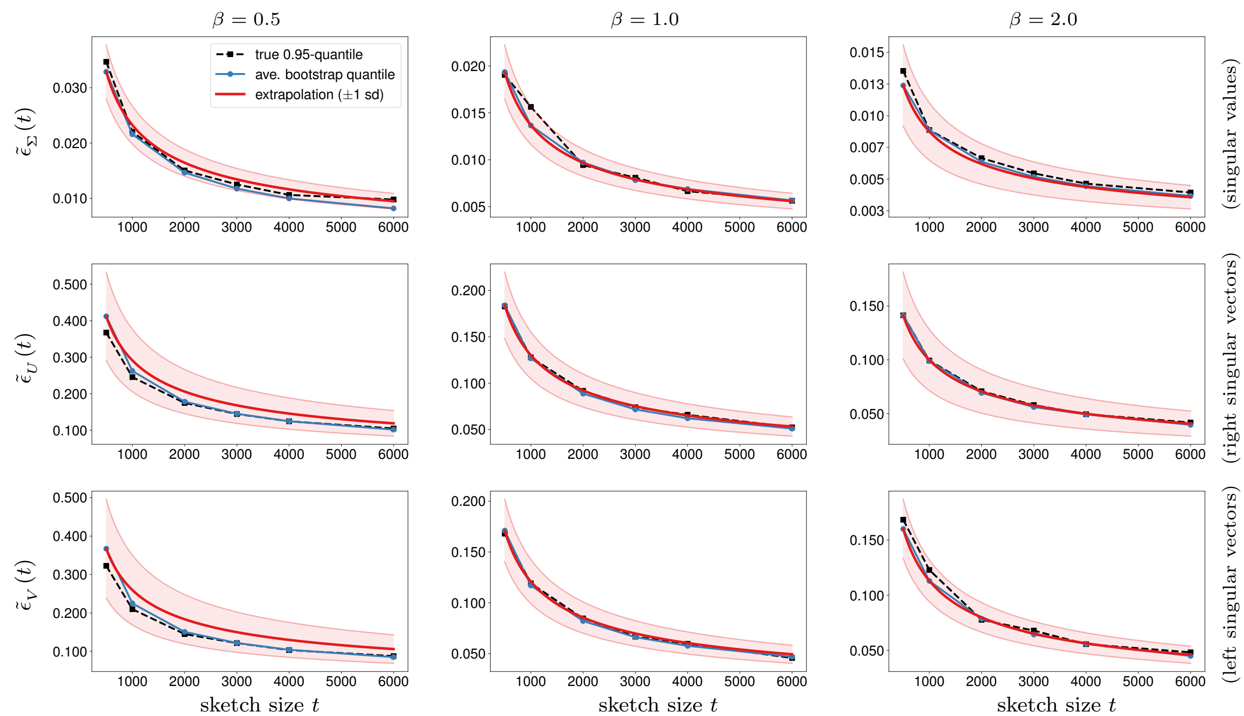

Figure 3: We consider artificial matrices of dimension $(n, d) = (10^5, 3 \times 10^3)$ that have singular value decay profiles of the form $\sigma_j = j^{-\beta}$ for $j \in \{1, \dots, d\}$ with parameter values $\beta \in \{0.5, 1, 2\}$. The error variables correspond to the index set $\mathcal{J} = \{1, 2, 3\}$, and the simulations involve 500 trials and 30 bootstraps per trial. The rows correspond to the error quantiles for the singular values (top), right singular vectors (middle), and left singular vectors (bottom). The labelling scheme of the curves is the same as in the main text.

G.4 Computational performance

The bootstrap method for error estimation can be executed in parallel, since the bootstrap replicates are independent of the others. Indeed, this a feature of the bootstrap method and allows one to take advantage of modern computational environments such as cloud computing. For instance, up to several thousands

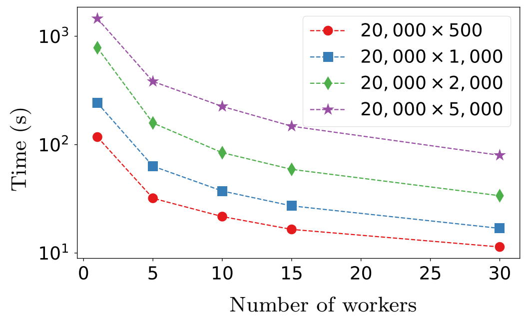

Figure 4: Computational time that is required to compute 30 bootstrap estimates using serverless computing. Here we compute the error estimates for the top 5 right singular vectors and values. In practice, one might even choose the number of workers larger than the number of required bootstrap samples B. In this case, the algorithm stops once B samples are received and one does not need to wait for the results of workers (so-called stragglers) which have a low response time.

workers can be allocated on AWS Lambda in less than ten seconds. In the following, we just use up to 30 low budget workers for our experiments. (The computational times can be reduced by using computational more powerful workers, however, these come with a higher price tag.)

Given a row index set, the individual workers can pull the row subset from a source matrix that is stored in external memory (e.g., cloud storage like AWS S3). Then, each worker computes an error estimate that is returned to the central node. The computational times for executing our algorithm for various numbers of workers are illustrated in Figure 4. Here, we consider a tall synthetic source matrix with $(n, d) = (10^5, 20 \times 10^3)$ that is characterized by low effective rank, i.e., the singular values of the source matrix were chosen as $\Sigma = \text{diag}(1^{-\beta}, 2^{-\beta}, \ldots, d^{-\beta})$ with $\beta \in \{2.0\}$.

References

Nir Ailon and Bernard Chazelle. Approximate nearest neighbors and the fast johnson-lindenstrauss transform. In *Proceedings of the thirty-eighth annual ACM symposium on Theory of computing*, pages 557–563, 2006.

Z. D. Bai and J. W. Silverstein. *Spectral Analysis of Large Dimensional Random Matrices*. Springer, New York, 2010.

S. Cambanis, S. Huang, and G. Simons. On the theory of elliptically contoured distributions. *Journal of Multivariate Analysis*, 11(3):368–385, 1981.

R. M. Dudley. *Real Analysis and Probability*. Cambridge, 2002.

J. Hu, W. Li, Z. Liu, and W. Zhou. High-dimensional covariance matrices in elliptical distributions with application to spherical test. *The Annals of Statistics*, 47(1):527–555, 2019.

O. Kallenberg. *Foundations of Modern Probability*. Springer, 2006.

F. Kenney and E. S. Keeping. *Mathematics of Statistics, part 2*. D. Van Nostrand Company, 1951.

M. E. Lopes, S. Wang, and M. W. Mahoney. Error estimation for randomized least-squares algorithms via the bootstrap. *arXiv:1803.08021*, 2018.

J. R. Magnus and H. Neudecker. *Matrix Differential Calculus with Applications in Statistics and Econometrics*. Wiley, 2019.

D. N. Politis, J. P. Romano, and M. Wolf. *Subsampling*. Springer, 1999.

A. W. van der Vaart. *Asymptotic Statistics*. Cambridge, 2000.

A. W. van der Vaart and J. A. Wellner. *Weak Convergence and Empirical Processes*. Springer, 1996.