Debiasing Random Oblique Projections for Subsampled OLS and Fast CUR in High Dimensions

Chengmei Niu[1] **Sachin Garg**[2] **Michał Dereziński**[2] **Zhenyu Liao**[1*]
[1]Huazhong University of Science and Technology
[2]University of Michigan, Ann Arbor
{chengmeiniu,zhenyu_liao}@hust.edu.cn {sachg,derezin}@umich.edu

Abstract

Random sampling is a fundamental tool in modern machine learning and numerical linear algebra for reducing the computational cost of large-scale matrix problems. Existing analyses, however, rely primarily on subspace embedding guarantees, which do *not* precisely characterize the statistical bias of *nonlinear* random oblique projections induced by sampling, which arises ubiquitously in subsampled least squares and fast low-rank approximation methods. Because (pseudo)inversion is *nonlinear*, these random oblique projections can be systematically *biased* even when the underlying sketch is unbiased, thereby introducing hidden bias into downstream least squares and low-rank approximation solutions. In this work, we develop a unified non-asymptotic theory for random oblique projections in high dimensions. We show that standard random sampling schemes generally induce a systematic statistical *bias* overlooked by classical subspace embedding-style analyses, and we propose a principled debiasing framework to correct it. We illustrate the power of the theory through two canonical applications. For subsampled least squares, we obtain sharp bias–variance characterizations, reveal previously unrecognized statistical *suboptimality* in widely used sampling schemes, and identify when debiasing yields provable improvements. For fast CUR decomposition, we develop a debiased approach with improved approximation accuracy. Numerical experiments further validate our theoretical findings.

1 Introduction

Many problems in modern machine learning (ML) and scientific computing involve data matrices that are too large to be stored or processed efficiently. Randomized numerical linear algebra (RandNLA) addresses this challenge through random sketching techniques [25, 26, 24, 3, 50, 6, 38, 19, 34, 45]. Given a tall matrix $\mathbf{X} \in \mathbb{R}^{n\times p}$ with $n \gg p$, one constructs a sketch $\tilde{\mathbf{X}} = \mathbf{S}\mathbf{X} \in \mathbb{R}^{m\times p}$, with $m \ll n$, to serve as a computationally efficient proxy for $\mathbf{X}$ in downstream tasks such as linear least squares, low-rank approximation, and iterative second-order optimization. The sketching matrix $\mathbf{S} \in \mathbb{R}^{m\times n}$ is typically generated either through *random sampling*, which randomly selects rows of $\mathbf{X}$, or through *random projection*, which forms random linear combinations of its rows.

Here, we focus on the random *oblique projection* $\tilde{\mathbf{P}}$ [10, 9] induced by random sampling, and its residual projection $\tilde{\mathbf{P}}_\perp$: $\tilde{\mathbf{P}} = \mathbf{X}(\mathbf{S}\mathbf{X})^\dagger\mathbf{S}$, $\tilde{\mathbf{P}}_\perp = \mathbf{I}_n - \tilde{\mathbf{P}}$. These operators are designed to approximate, and should thus be contrasted with, the orthogonal projection $\mathbf{P} = \mathbf{X}\mathbf{X}^\dagger$ onto the column space of $\mathbf{X}$, and the residual projection $\mathbf{P}_\perp = \mathbf{I}_n - \mathbf{P}$. Random oblique projections arise in a broad range of RandNLA methods, including subsampled least squares [25, 26], fast low-rank approximation such as CUR decomposition [57], and randomized optimization [4].

*Corresponding author: zhenyu_liao@hust.edu.cn.

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

A common design principle is that the subsampled Gram matrix $\tilde{\mathbf{X}}^\top\tilde{\mathbf{X}}$ should be an unbiased estimator of $\mathbf{X}^\top\mathbf{X}$, i.e., $\mathbb{E}[\tilde{\mathbf{X}}^\top\tilde{\mathbf{X}}] = \mathbf{X}^\top\mathbf{X}$. This, however, does *not* generally imply unbiasedness of the induced oblique projection $\tilde{\mathbf{P}} = \mathbf{X}(\mathbf{S}\mathbf{X})^\dagger\mathbf{S} = \mathbf{X}(\mathbf{X}^\top\mathbf{S}^\top\mathbf{S}\mathbf{X})^\dagger\mathbf{X}^\top\mathbf{S}^\top\mathbf{S}$. The pseudoinverse is nonlinear, and its *dependence* on the sampled cross term $\mathbf{X}^\top\mathbf{S}^\top\mathbf{S}$ persists because both involve the *same* random sampling matrix $\mathbf{S}$; hence, their expectation does not factorize, and one could have $\mathbb{E}[\tilde{\mathbf{P}}] \neq \mathbf{P}$ even when the subsampled Gram is unbiased. Existing analyses of random oblique projections (and consequently of subsampled least squares and fast low-rank approximation methods) rely primarily on Johnson–Lindenstrauss (JL)-style analyses [35]. Such results aim to establish that $\tilde{\mathbf{X}}^\top\tilde{\mathbf{X}} \approx \mathbf{X}^\top\mathbf{X}$ with high probability and therefore that $\tilde{\mathbf{P}} \approx \mathbf{P}$ in some (e.g., matrix norm) sense. While powerful and easy to use, such guarantees are often too coarse to capture finer statistical behaviors central to the practical performance of subsampled least squares and fast low-rank approximation methods, as simple and fundamental as *bias and variance*.

In this paper, we address this gap by developing a refined non-asymptotic theory for oblique projections induced by random sampling, and illustrate its implication through the two canonical applications of subsampled least squares and fast CUR decomposition.

1.1 Our contributions

Our main contributions are summarized as follows.

1. We provide in Theorem 3.1 a non-asymptotic characterization of the bias induced by random oblique projections under general random sampling, along with a principled *debiasing* approach.
2. Building on this result, we first establish in Section 4 sharp bias–variance characterizations for subsampled ordinary least squares (OLS), including a bias lower bound for classical subsampled OLS (Theorem 4.2) and an improved upper bound on the bias for the proposed *debiased* subsampled OLS, whose variance matches that of the classical solution (Theorem 4.4 versus Theorem 4.3). We then extend the analysis to randomized CUR with fixed selected columns and rows (Theorem 4.8), deriving a debiased fast CUR method with provably improved approximation accuracy.

1.2 Related work

Below, we briefly review related previous efforts. A more detailed discussion appears in Appendix A.

Subsampled OLS. Subsampled OLS reduces the computational cost of least squares, and its statistical properties have been extensively studied [25, 26, 56, 4, 41, 49, 21, 48, 22, 15, 52]. However, precise characterizations of its bias relative to full-data OLS remain limited beyond determinantal sampling schemes [20].

CUR decomposition. CUR decomposition approximates a matrix from sampled rows and columns, with accuracy depending critically on the middle factor $\mathbf{U}$. Two major choices are: $\mathbf{U} = \mathbf{C}^\dagger\mathbf{X}\mathbf{R}^\dagger$ that is robust but expensive; and $\mathbf{U} = \mathbf{X}^\dagger_{R,C}$ that reduces the cost but at the expense of stability. Pivoting can improve stability but may be costly at scale [30, 54, 32]. Recent sketching-based methods improve scalability [23, 46, 57, 59], but mostly rely on subspace embedding guarantees, without precise characterizations of, e.g., the bias and variance.

Random matrix theory. Random matrix theory (RMT) has motivated fine-grained analyses of optimization and generalization in large-scale ML [12, 47, 28, 44, 31, 39]. In particular, carefully designed leave-one-out arguments yield sharp non-asymptotic results for inverse Gram matrices beyond classical asymptotic regimes [16, 18, 17, 29, 45]. We extend this approach to the oblique projector $\tilde{\mathbf{P}}$, where the sampled inverse must be analyzed *jointly* with its cross term.

Notation. Scalars, vectors, and matrices are denoted by lowercase, bold lowercase, and bold uppercase letters, respectively. For $\mathbf{X} \in \mathbb{R}^{n\times p}$, we write $\mathbf{X}^\top$, $\mathbf{X}^\dagger$, $\mathbf{x}_i^\top$, $\|\mathbf{X}\|$, and $\|\mathbf{X}\|_F$ for its transpose, Moore–Penrose pseudoinverse, i^{th} row (with $\mathbf{x}_i \in \mathbb{R}^p$), spectral norm, and Frobenius norm, respectively. We write $\mathbf{X} \preceq \mathbf{Y}$ when $\mathbf{Y} - \mathbf{X}$ is positive semi-definite, and use $\mathbf{I}_p$ for the $p \times p$ identity matrix. For a random variable x, $\mathbb{E}[x]$ denotes its expectation, and $\mathbb{E}_\zeta[x] \equiv \mathbb{E}[x \mid \zeta]$ denotes

its expectation conditioned on an event ζ. We use the standard Big-O and Big-Ω notation, with $\tilde{O}(\cdot)$ suppressing logarithmic dependence.

2 Preliminaries

In this section, we introduce several definitions that will be used throughout the paper.

Definition 2.1 (**Random sampling**). For a matrix $\mathbf{X} \in \mathbb{R}^{n\times p}$ with $n \geq p$, a sketch $\tilde{\mathbf{X}} \in \mathbb{R}^{m\times p}$ of $\mathbf{X}$ can be constructed by sampling *with replacement* m of the n rows of $\mathbf{X}$ with a sampling distribution, $\{\pi_i\}_{i=1}^n$, $\sum_{i=1}^n \pi_i = 1$, and then rescaling by $1/\sqrt{m\pi_i}$. This procedure writes $\tilde{\mathbf{X}} = \mathbf{S}\mathbf{X}$, with *sampling matrix* $\mathbf{S} \in \mathbb{R}^{m\times n}$ having only one nonzero entry per row. One has $\mathbb{E}[\tilde{\mathbf{X}}^\top\tilde{\mathbf{X}}] = \mathbf{X}^\top\mathbf{X}$.

Definition 2.1 includes commonly used random sampling schemes such as uniform ($\pi_i = 1/n$), row-norm-based ($\pi_i = \|\mathbf{x}_i\|^2/\|\mathbf{X}\|_F^2$), exact/approximate leverage score sampling, as well as hybrid schemes that interpolate between them, e.g., shrinkage leverage score sampling [41].

Definition 2.2 (**Leverage score sampling**, [42]). For $\mathbf{X} \in \mathbb{R}^{n\times p}$ of rank p with $n \geq p$, the ith *leverage score* $\ell_i(\mathbf{X})$ of $\mathbf{X}$ is defined as $\ell_i(\mathbf{X}) = \mathbf{x}_i^\top(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{x}_i, i \in \{1,\dots,n\}$. The exact leverage score sampling refers to the random sampling approach in Definition 2.1 with $\pi_i = \ell_i(\mathbf{X})/p$.

Definition 2.3 (**Importance sampling approximation factor**, [45]). For $\mathbf{X} \in \mathbb{R}^{n\times p}$ with $n \geq p$ and a random sampling matrix $\mathbf{S} \in \mathbb{R}^{m\times n}$ in Definition 2.1 with sampling distribution $\{\pi_i\}_{i=1}^n$, the associated minimum and maximum *importance sampling approximation factors* are defined as $\theta_{\min}(\mathbf{X}) \equiv \min_{1\leq i\leq n} \ell_i(\mathbf{X})/(\pi_i p)$ and $\theta_{\max}(\mathbf{X}) \equiv \max_{1\leq i\leq n} \ell_i(\mathbf{X})/(\pi_i p)$.

The importance sampling approximation factors in Definition 2.3, introduced in [41, 45], measure the departure from the *exact* leverage score sampling scheme in Definition 2.2. In particular, one has $\theta_{\min}(\mathbf{X}) \leq 1 \leq \theta_{\max}(\mathbf{X})$ with $\theta_{\min}(\mathbf{X}) = \theta_{\max}(\mathbf{X}) = 1$ for exact leverage score sampling. As we shall see below, these parameters will determine how the sampling distribution enters the bias and variance bounds.

Definition 2.4 (**Subspace embedding**, [25, 42]). For $\mathbf{X} \in \mathbb{R}^{n\times p}$ with $n \geq p$ and a random sampling matrix $\mathbf{S} \in \mathbb{R}^{m\times n}$ with $m \leq n$, we say that a sketch $\tilde{\mathbf{X}} = \mathbf{S}\mathbf{X} \in \mathbb{R}^{m\times p}$ is an (ϵ, δ)-*subspace embedding*, or an (ϵ, δ)-approximation for $\mathbf{X}$ if

$$(1+\epsilon)^{-1}\mathbf{X}^\top\mathbf{X} \preceq \tilde{\mathbf{X}}^\top\tilde{\mathbf{X}} \preceq (1+\epsilon)\mathbf{X}^\top\mathbf{X}$$

holds with probability at least $1-\delta$.

The subspace embedding property in Definition 2.4 ensures that the approximation of $\mathbf{X}^\top\mathbf{X}$ by $\tilde{\mathbf{X}}^\top\tilde{\mathbf{X}}$ holds in a matrix partial order sense, with controlled probability. It underlines statistical guarantees for many choices of sampling schemes $\mathbf{S}$; see [42, 58].

Nonetheless, this property does not by itself control the expectation of a *nonlinear* function of the sketch. In particular, the unbiasedness $\mathbb{E}[\tilde{\mathbf{X}}^\top\tilde{\mathbf{X}}] = \mathbf{X}^\top\mathbf{X}$ in Definition 2.1 does *not* extend to the inverse $(\tilde{\mathbf{X}}^\top\tilde{\mathbf{X}})^{-1}$: $(\tilde{\mathbf{X}}^\top\tilde{\mathbf{X}})^{-1}$ can exhibit *significant* inverse bias, with $\|\mathbb{E}[(\tilde{\mathbf{X}}^\top\tilde{\mathbf{X}})^{-1}] - (\mathbf{X}^\top\mathbf{X})^{-1}\| \gg 0$. We recall two results that make this distinction explicit.

Remark 2.5 (**Inversion bias for Gaussian random projection**). For $\mathbf{S}$ having i.i.d. Gaussian entries of zero mean and variance $1/m$, and $m > p+1$, the inverse-Wishart distribution gives $\mathbb{E}[(\tilde{\mathbf{X}}^\top\tilde{\mathbf{X}})^{-1}] = \frac{m}{m-p-1}(\mathbf{X}^\top\mathbf{X})^{-1}$ [33]. Thus, a simple rescaling by $(m-p-1)/m$ removes its inversion bias exactly. Analogous scalar corrections approximately remove inversion bias for sub-Gaussian and LESS sketches [18].

The *inversion bias* for random sampling has been characterized in the following result.

Theorem 2.6 (**Inversion bias for random sampling**, [45, Theorem 3.1]). *For $\mathbf{X} \in \mathbb{R}^{n\times p}$ of rank p with $n \geq p$, let $\mathbf{S} \in \mathbb{R}^{m\times n}$ be a random sampling matrix with sampling distribution $\{\pi_i\}_{i=1}^n$ as in Definition 2.1 and $\theta_{\min}(\mathbf{X}), \theta_{\max}(\mathbf{X})$ as in Definition 2.3. Then, for diagonal matrix $\mathbf{D} = \operatorname{diag}\{D_{ii}\}_{i=1}^n$ the solution to $D_{ii} = \left(1 + \mathbf{x}_i^\top(\mathbf{X}^\top\mathbf{D}\mathbf{X})^{-1}\mathbf{x}_i/(m\pi_i)\right)^{-1}$, there exists $C > 0$ independent of n, p, so that for $m \geq C\theta_{\max}(\mathbf{X})p(\log(p/\delta) + (\log n/\log\log n)^{2/3}/\epsilon^{2/3})$, $\delta \leq m^{-3}$, when conditioned on an event ζ that holds with probability at least $1-\delta$,*

$$(1+\epsilon)^{-1}(\mathbf{X}^\top\mathbf{D}\mathbf{X})^{-1} \preceq \mathbb{E}_\zeta[(\tilde{\mathbf{X}}^\top\tilde{\mathbf{X}})^{-1}] \preceq (1+\epsilon)(\mathbf{X}^\top\mathbf{D}\mathbf{X})^{-1}.$$

Under the stated conditions, Theorem 2.6 characterizes the inversion bias through the diagonal matrix $\mathbf{D}$: the (conditional) expectation of $(\tilde{\mathbf{X}}^\top\tilde{\mathbf{X}})^{-1}$ is close to $(\mathbf{X}^\top\mathbf{D}\mathbf{X})^{-1}$, rather than the full inverse $(\mathbf{X}^\top\mathbf{X})^{-1}$. Thus, matching the expected Gram matrix does *not* lead to the expected inverse (as one may expect from classical subspace embedding guarantees). The *matrix-level* debiasing approach in [45, Proposition 3.2] addresses this discrepancy by replacing the sampling matrix with $\check{\mathbf{S}}$ of the same size. For $\check{\mathbf{X}} = \check{\mathbf{S}}\mathbf{X}$, one has $\mathbb{E}[(\check{\mathbf{X}}^\top\check{\mathbf{X}})^{-1}] \simeq (\mathbf{X}^\top\mathbf{X})^{-1}$ in a spectral norm sense.

The inversion bias result in Theorem 2.6 suggests that the random oblique projection $\tilde{\mathbf{P}} = \mathbf{X}(\mathbf{S}\mathbf{X})^\dagger\mathbf{S} = \mathbf{X}(\mathbf{X}^\top\mathbf{S}^\top\mathbf{S}\mathbf{X})^\dagger\mathbf{X}^\top\mathbf{S}^\top\mathbf{S}$ may likewise be a *biased* estimator of $\mathbf{P} = \mathbf{X}\mathbf{X}^\dagger$. However, the inverse debiasing approach in [45] is only part of the argument. The bias of $\tilde{\mathbf{P}}$ relative to $\mathbf{P}$ also depends on the cross term $\mathbf{X}^\top\mathbf{S}^\top\mathbf{S}$ formed from the same sketch $\mathbf{S}$. We now examine how the debiasing acts on this *joint dependence*.

3 Debiased oblique projection for random sampling

In the following result, we show that an effective debiasing exists for random oblique projections.

Theorem 3.1 (Precise characterizations of debiased oblique projection). *For $\mathbf{X} \in \mathbb{R}^{n\times p}$ of rank p with $n \geq p$, assume that $\min_{i,\|\mathbf{x}_i\|>0} \|\mathbf{x}_i\|^2/\|\mathbf{X}\|_F^2 \geq n^{-\alpha}$ for some constant $\alpha > 0$. Let $\mathbf{S}$ be a random sampling matrix with sampling distribution $\{\pi_i\}_{i=1}^n$ as in Definition 2.1, define the debiased sampling matrix $\check{\mathbf{S}} \in \mathbb{R}^{m\times n}$ as*[2]

$$\check{\mathbf{S}} = \operatorname{diag}\left\{1/\sqrt{1-\ell_{i_s}(\mathbf{X})/(m\pi_{i_s})}\right\}_{s=1}^m \cdot \mathbf{S}. \tag{1}$$

Then, there exists $C > 0$ independent of n, p so that for $m \geq C\theta_{\max}(\mathbf{X})p\log(p/\delta)$, $\delta \leq n^{-(3+\alpha)}$, and $\theta_{\max}(\mathbf{X})$ in Definition 2.3, when conditioned on an event ζ that holds with probability at least $1-\delta$, the debiased oblique projection $\check{\mathbf{P}} \equiv \mathbf{X}(\check{\mathbf{S}}\mathbf{X})^\dagger\check{\mathbf{S}}$ satisfies[3]

$$\left\|\mathbb{E}_\zeta[\check{\mathbf{P}}] - \mathbf{P}\right\|_F^2 = \epsilon^2 \cdot \|\mathbf{P}_\perp\|_F^2, \tag{2}$$

$$\mathbb{E}_\zeta\left[\|\check{\mathbf{P}} - \mathbf{P}\|_F^2\right] = \operatorname{tr}\left(\mathbf{P}_\perp \operatorname{diag}\left\{\ell_i(\mathbf{X})/(m\pi_i)\right\}_{i=1}^n\right) + \epsilon \cdot \|\mathbf{P}_\perp\|_F^2, \tag{3}$$

with $\epsilon = O(\sqrt{(\log n/\operatorname{loglog} n)^2 \cdot \theta_{\max}^3(\mathbf{X})p^3/m^3})$, for orthogonal projection $\mathbf{P} = \mathbf{X}\mathbf{X}^\dagger$ and residual projection $\mathbf{P}_\perp = \mathbf{I}_n - \mathbf{X}\mathbf{X}^\dagger$.

Heuristic derivation of Theorem 3.1. The main technical challenges in proving Theorem 3.1 is the dependence between the Gram inverse and the external cross-term in $\check{\mathbf{P}} = \mathbf{X}(\mathbf{X}^\top\check{\mathbf{S}}^\top\check{\mathbf{S}}\mathbf{X})^{-1}\mathbf{X}^\top\check{\mathbf{S}}^\top\check{\mathbf{S}}$, both of which involve the *same* sketch $\check{\mathbf{S}}$. Consider $\check{\mathbf{S}} = \operatorname{diag}\{\sqrt{F_{i_s i_s}}\}_{s=1}^m\mathbf{S}$, where F_{ii} are positive deterministic weights to be determined, we aim to establish that

$$\|\mathbb{E}[\check{\mathbf{P}}] - \mathbf{P}\|_F^2 \simeq 0.$$

To this end, let $\mathbf{x}_i^\top = \mathbf{e}_i^\top\mathbf{X}$, $\mathbf{z}_s = \mathbf{x}_{i_s}/\sqrt{\pi_{i_s}}$, and $\mathbf{w}_s = \mathbf{e}_{i_s}/\sqrt{\pi_{i_s}}$, we have

$$\mathbf{Q} = (\mathbf{X}^\top\check{\mathbf{S}}^\top\check{\mathbf{S}}\mathbf{X})^{-1} = \left(\frac{1}{m}\sum_{s=1}^m F_{i_s i_s}\mathbf{z}_s\mathbf{z}_s^\top\right)^{-1}.$$

Define the leave-one-out inverse $\mathbf{Q}_{-s} \equiv \left(\frac{1}{m}\sum_{k\neq s} F_{i_k i_k}\mathbf{z}_k\mathbf{z}_k^\top\right)^{-1}$, which is independent of $\mathbf{z}_s, \mathbf{w}_s$ by construction. It follows from the Sherman–Morrison formula in Lemma B.1 that

$$\mathbb{E}[\check{\mathbf{P}}] = \mathbb{E}\left[\frac{F_{i_s i_s}\mathbf{X}\mathbf{Q}_{-s}\mathbf{z}_s\mathbf{w}_s^\top}{1 + F_{i_s i_s}\mathbf{z}_s^\top\mathbf{Q}_{-s}\mathbf{z}_s/m}\right] = \sum_{i=1}^n \mathbb{E}\left[\frac{F_{ii}\mathbf{X}\mathbf{Q}_{-s}\mathbf{x}_i\mathbf{e}_i^\top}{1 + F_{ii}\mathbf{x}_i^\top\mathbf{Q}_{-s}\mathbf{x}_i/(m\pi_i)}\right].$$

[2] For rank-deficient $\mathbf{X}$ (i.e., $\operatorname{rank}(\mathbf{X}) < p$), it suffices to restrict to the column space of $\mathbf{X}$ (that is of dimension smaller than p) and Theorem 3.1 naturally extends to this setting.

[3] The event ζ captures the high-probability guarantee that for $m \geq C\theta_{\max}(\mathbf{X})p\log(p/\delta)$, $\mathbf{S}\mathbf{X}$ is an (ϵ,δ)-subspace embedding of $\mathbf{X}$ (see Lemma B.3 of Appendix B), ensuring a well-behaved sketch throughout the analysis. Here, we use "debiased" for the removal of the leading systematic bias, not exact unbiasedness.

This representation isolates the statistical dependence between the Gram inverse and the cross-term contributed by the omitted sample.

To further identify a candidate correction F_{ii}, we use the approximations that $\mathbf{x}_i^\top \mathbf{Q}_{-s}\mathbf{x}_i \simeq \ell_i(\mathbf{X})$ and $\|\mathbb{E}[\mathbf{Q}_{-s}] - \mathbb{E}[\mathbf{Q}]\| \simeq 0$, where $\ell_i(\mathbf{X}) = \mathbf{x}_i^\top(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{x}_i$ is the ith leverage score of $\mathbf{X}$. These approximations yield

$$\mathbb{E}[\check{\mathbf{P}}] \simeq \mathbf{X}\mathbb{E}[\mathbf{Q}] \sum_{i=1}^{n} \frac{F_{ii}\mathbf{x}_i\mathbf{e}_i^\top}{1 + F_{ii}\ell_i(\mathbf{X})/(m\pi_i)}.$$

Choosing

$$F_{ii} = \frac{m\pi_i}{m\pi_i - \ell_i(\mathbf{X})}, \tag{4}$$

gives $\frac{F_{ii}}{1+F_{ii}\ell_i(\mathbf{X})/(m\pi_i)} = 1$. Combining this cancellation with the inverse-debiasing result of [45, Proposition 3.2]: $\mathbb{E}[\mathbf{Q}] \simeq (\mathbf{X}^\top\mathbf{X})^{-1}$, yields the first-order approximation in (2):

$$\mathbb{E}[\check{\mathbf{P}}] \simeq \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top = \mathbf{P}.$$

For the second-moment result in (3), the identity $\check{\mathbf{P}} - \mathbf{P} = \check{\mathbf{P}}\mathbf{P}_\perp$ gives

$$\mathbb{E}\big[\|\check{\mathbf{P}} - \mathbf{P}\|_F^2\big] = \mathbb{E}\big[\mathrm{tr}(\mathbf{P}_\perp\check{\mathbf{P}}^\top\check{\mathbf{P}}\mathbf{P}_\perp)\big].$$

An analogous analysis yields the leading term $\mathrm{tr}\left(\mathbf{P}_\perp \,\mathrm{diag}\left\{\ell_i(\mathbf{X})/(m\pi_i)\right\}_{i=1}^n\right)$ in (3). The more intricate dependence in this second-moment calculation requires both leave-one-out and leave-two-out arguments. See Appendices E and F.1 for detailed exposition of the proof. □

The resulting weights in $\check{\mathbf{S}}$ coincide with those used for inverse debiasing in [45], but the inverse guarantee alone does not control the coupling between $\mathbf{Q}$ and the cross-term $\mathbf{X}^\top\check{\mathbf{S}}^\top\check{\mathbf{S}}$. Our analysis identifies how these weights, combined with residual orthogonality $\mathbf{X}^\top\mathbf{P}_\perp = \mathbf{0}$ in the leave-one-out expansion, cancel the leading contribution of this dependence to the projection bias, and leads to bounds in Theorem 3.1; see Appendices E and F.1.

Computing the exact scores ℓ_i in $\check{\mathbf{S}}$ may be expensive. They can be replaced in (1) by efficient approximations [24, 11], with an additional but marginal error determined by their relative accuracy. Corollary F.1 in Appendix F.2 quantifies this dependence.

Theorem 3.1 characterizes both first- and second-order moments of the debiased oblique projection. Precisely, we have that the debiased oblique projection $\check{\mathbf{P}}$ is close, in expectation, to the true projection $\mathbf{P}$, up to some controlled error for *all* sampling schemes. Moreover, the second-order moment $\mathbb{E}_\zeta\left[\|\check{\mathbf{P}} - \mathbf{P}\|_F^2\right]$ is also precisely characterized, as the sum of the leading-order term $\mathrm{tr}(\mathbf{P}_\perp \,\mathrm{diag}\{\ell_i(\mathbf{X})/(m\pi_i)\}_{i=1}^n)$ and some explicitly controlled error.

By Definition 2.3, $\theta_{\min}(\mathbf{X})p/m\cdot\mathbf{I}_n \preceq \mathrm{diag}\{\ell_i(\mathbf{X})/(m\pi_i)\}_{i=1}^n \preceq \theta_{\max}(\mathbf{X})p/m\cdot\mathbf{I}_n$. Thus, (3) gives $(\theta_{\min}(\mathbf{X})p/m - \epsilon)\|\mathbf{P}_\perp\|_F^2 \le \mathbb{E}_\zeta[\|\check{\mathbf{P}} - \mathbf{P}\|_F^2] \le (\theta_{\max}(\mathbf{X})p/m + \epsilon)\|\mathbf{P}_\perp\|_F^2$. For *exact* leverage score sampling, $\theta_{\min}(\mathbf{X}) = \theta_{\max}(\mathbf{X}) = 1$, yielding $\mathbb{E}_\zeta[\|\check{\mathbf{P}} - \mathbf{P}\|_F^2] = (p/m + \epsilon)\cdot\|\mathbf{P}_\perp\|_F^2$.

Remark 3.2 (**Scalar debiasing for exact leverage score sampling**). Note that under exact leverage score sampling, where $\pi_i = \ell_i(\mathbf{X})/p$ for all i (see again Definition 2.2), the debiased sampling matrix in (1) reduces to $\check{\mathbf{S}} = \sqrt{m/(m-p)}\cdot\mathbf{S}$, which coincides with the *scalar* debiasing factor appearing for Gaussian, sub-Gaussian, and LESS projections in Remark 2.5. In this setting, for the debiased oblique projection $\check{\mathbf{P}} = \mathbf{X}(\check{\mathbf{S}}\mathbf{X})^\dagger\check{\mathbf{S}}$, the scalar factor $m/(m-p)$ inside and outside the pseudoinverse cancels out, yielding $\check{\mathbf{P}} = \tilde{\mathbf{P}}$. Consequently, *no* debiasing is needed for exact leverage score sampling. One may thus expect that a similar phenomenon persists, at least to some extent, for close-to-exact leverage score sampling schemes. This statement will be made precise below in Section 4.2.

4 Application to subsampled OLS and fast CUR

We now apply our precise characterizations of random oblique projection in Theorem 3.1 to subsampled OLS and fast CUR. Subspace-embedding arguments provide statistical guarantees for these methods, and generally fall short of providing precise characterizations of, say its first and second

moments under random sampling [56, 41, 49, 22, 15]; see Definition 4.1 for their formal definitions in the context of subsampled OLS.

Given a data matrix $\mathbf{X} \in \mathbb{R}^{n\times p}$ and a response vector $\mathbf{y} \in \mathbb{R}^n$, the OLS solution is given by

$$\boldsymbol{\beta}_{\mathrm{OLS}} = \underset{\boldsymbol{\beta}\in\mathbb{R}^p}{\arg\min}\, L(\boldsymbol{\beta}) = \mathbf{X}^\dagger \mathbf{y}, \;\; \text{for} \;\; L(\boldsymbol{\beta}) = \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|^2, \tag{5}$$

which can be computationally expensive for n large. We consider the subsampled OLS [25, 56, 4]:

$$\tilde{\boldsymbol{\beta}} = (\mathbf{S}\mathbf{X})^\dagger \mathbf{S}\mathbf{y}, \tag{6}$$

where $\mathbf{S} \in \mathbb{R}^{m\times n}$ is a random sampling matrix. Consider also its *debiased* counterpart:

$$\check{\boldsymbol{\beta}} = (\check{\mathbf{S}}\mathbf{X})^\dagger \check{\mathbf{S}}\mathbf{y}, \tag{7}$$

where $\check{\mathbf{S}}$ is the debiased sampling matrix in (1) of Theorem 3.1.

Observe that $L(\tilde{\boldsymbol{\beta}}) = \|\mathbf{y} - \mathbf{X}(\mathbf{S}\mathbf{X})^\dagger \mathbf{S}\mathbf{y}\|^2$ and $L(\check{\boldsymbol{\beta}}) = \|\mathbf{y} - \mathbf{X}(\check{\mathbf{S}}\mathbf{X})^\dagger \check{\mathbf{S}}\mathbf{y}\|^2$. Thus, subsampling replaces the orthogonal projection $\mathbf{X}\mathbf{X}^\dagger$ with the oblique projections $\mathbf{X}(\mathbf{S}\mathbf{X})^\dagger \mathbf{S}$ and $\mathbf{X}(\check{\mathbf{S}}\mathbf{X})^\dagger \check{\mathbf{S}}$. The deviation of these oblique projections from orthogonal projections governs the statistical behavior of the corresponding estimators relative to full OLS.

We compare the classical subsampled OLS $\tilde{\boldsymbol{\beta}}$ in (6) and its debiased counterpart $\check{\boldsymbol{\beta}}$ in (7) with the full OLS $\boldsymbol{\beta}_{\mathrm{OLS}}$ in (5), using the following two metrics, consistent with prior work [4, 15, 20, 29, 49].

Definition 4.1 (Bias and variance of subsampled OLS). For a data matrix $\mathbf{X} \in \mathbb{R}^{n\times p}$ of rank p, a response vector $\mathbf{y} \in \mathbb{R}^n$, and the OLS solution $\boldsymbol{\beta}_{\mathrm{OLS}}$ defined in (5), let $\boldsymbol{\beta} \in \mathbb{R}^p$ be a random vector (such as the subsampled OLS solution $\tilde{\boldsymbol{\beta}}$ in (6) or its debiased counterpart $\check{\boldsymbol{\beta}}$). Define

1. *Bias* $\mathrm{Bias}_\zeta(\boldsymbol{\beta}) = L(\mathbb{E}_\zeta[\boldsymbol{\beta}]) - L(\boldsymbol{\beta}_{\mathrm{OLS}})$ as the deviation of $\boldsymbol{\beta}$ from $\boldsymbol{\beta}_{\mathrm{OLS}}$ in expectation; and
2. *Variance* $\mathrm{Var}_\zeta(\boldsymbol{\beta}) = \mathbb{E}_\zeta[L(\boldsymbol{\beta})] - L(\boldsymbol{\beta}_{\mathrm{OLS}})$ that quantifies the fluctuation due to random sampling;

both measured by the square loss $L(\cdot)$ in (5) and conditioned on a high probability event ζ (which is necessary for many random sampling schemes, see Footnote 3).

4.1 Bias and variance characterizations of subsampled OLS

We first establish in Theorem 4.2 that the classical subsampled OLS $\tilde{\boldsymbol{\beta}}$ in (6), *without debiasing*, can exhibit a *substantial bias* (in the sense of Definition 4.1), at least under *certain* sampling schemes. We then derive in Theorem 4.3 an upper bound on the variance of $\tilde{\boldsymbol{\beta}}$. Finally, in Theorem 4.4, we present precise bias–variance characterizations for the proposed *debiased* subsampled OLS $\check{\boldsymbol{\beta}}$, demonstrating that debiasing not only mitigates the bias but also preserves, *if not improves*, the variance behavior.

Theorem 4.2 (Bias lower bound for subsampled OLS). *Let $p = 4k$ for any $k \geq 1$ and let $n \geq 2p$. Then, there exists a data matrix $\mathbf{X} \in \mathbb{R}^{n\times p}$, a response vector $\mathbf{y} \in \mathbb{R}^n$, an approximate leverage score sampling matrix $\mathbf{S} \in \mathbb{R}^{m\times n}$ with $\{\pi_i\}_{i=1}^n$ satisfying $\ell_i(\mathbf{X})/(2p) \leq \pi_i \leq 2\ell_i(\mathbf{X})/p$ for all i, such that for all $m \geq p$ and* any *real constant γ,*

$$\mathrm{Bias}(\gamma \cdot \tilde{\boldsymbol{\beta}}) = \Omega(p^2/m^2) \cdot \|\mathbf{r}\|^2,$$

with $\mathrm{Bias}(\cdot)$ *in Definition 4.1*[4]*, $\tilde{\boldsymbol{\beta}} = (\mathbf{S}\mathbf{X})^\dagger \mathbf{S}\mathbf{y}$ in* (6)*, and residual vector $\mathbf{r} = \mathbf{y} - \mathbf{X}\beta_{\mathrm{OLS}}$.*

Theorem 4.2 asserts that, under a certain random sampling scheme (that is $1/2$-approximate leverage score sampling), there exist scenarios for which the bias of classical subsampled OLS $\tilde{\boldsymbol{\beta}}$ is *at least* of order $\Omega(p^2/m^2)$. This lower bound holds for *all* rescaled solutions of the form $\gamma \cdot \tilde{\boldsymbol{\beta}}$, with $\gamma \in \mathbb{R}$. This result stands in sharp contrast to the Gaussian inverse result in Remark 2.5, where an appropriate scalar suffices to eliminate the inversion bias.

The proof of Theorem 4.2 in Appendix C is based on the construction of a "hard" example that admits a non-uniform bias across coordinates of the subsampled OLS, so that it *cannot* be corrected via a single scalar rescaling.

In the following result, we derive a variance upper bound for standard subsampled OLS.

[4]The expectation is taken without conditioning on the high-probability event ζ.

Theorem 4.3 (Variance upper bound for subsampled OLS). *For a data matrix $\mathbf{X} \in \mathbb{R}^{n\times p}$ of rank p, a response vector $\mathbf{y} \in \mathbb{R}^n$, and random sampling matrix $\mathbf{S} \in \mathbb{R}^{m\times n}$ with distribution $\{\pi_i\}_{i=1}^n$ as in Definition 2.1, there exists $C > 0$ independent of n, p so that for $m \geq C\theta_{\max}(\mathbf{X})p\log(p/\delta)$, any $\delta \in (0,1)$, when conditioned on an event ζ that holds with probability at least $1-\delta$, the subsampled OLS $\tilde{\boldsymbol{\beta}} = (\mathbf{SX})^\dagger \mathbf{Sy}$ in (6) satisfies*

$$\mathrm{Var}_\zeta(\tilde{\boldsymbol{\beta}}) = (1-\delta)^{-1}(\Delta(\mathbf{X}) + \epsilon \cdot \|\mathbf{r}\|^2),$$

with $\epsilon = O\left(\sqrt{\log(p/\delta)\theta^3_{\max}(\mathbf{X})p^3/m^3}\right)$, for $\mathrm{Var}_\zeta(\cdot)$ in Definition 4.1, and

$$\Delta(\mathbf{X}) = \mathbf{r}^\top \mathrm{diag}\left\{\ell_i(\mathbf{X})/(m\pi_i)\right\}_{i=1}^n \mathbf{r}, \quad \mathbf{r} \equiv \mathbf{y} - \mathbf{X}\boldsymbol{\beta}_{\mathrm{OLS}}. \tag{8}$$

Theorem 4.3 provides an upper bound for the variance of classical subsampled OLS $\tilde{\boldsymbol{\beta}}$ *without debiasing*. Following the discussion after Theorem 3.1, the variance (upper bound) in Theorem 4.3 can be similarly decomposed as the sum of the dominant term $\Delta(\mathbf{X}) = O(\|\mathbf{r}\|^2 \cdot \theta_{\max}(\mathbf{X})p/m)$ and higher-order term of order $\tilde{O}(\|\mathbf{r}\|^2 \cdot \sqrt{\theta^3_{\max}(\mathbf{X})p^3/m^3})$ akin to Theorem 3.1. The proof of Theorem 4.3 is given in Appendix D.

We next give bounds on bias and variance of the *debiased* solution; the detailed proof is in Appendix E.

Theorem 4.4 (Precise bias–variance characterizations for debiased subsampled OLS). *For a data matrix $\mathbf{X} \in \mathbb{R}^{n\times p}$ of rank p with $n \geq p$ and response vector $\mathbf{y} \in \mathbb{R}^n$, assume that $\min_{i,\|\mathbf{x}_i\|>0} \|\mathbf{x}_i\|^2/\|\mathbf{X}\|_F^2 \geq n^{-\alpha}$ for some constant $\alpha > 0$. Let $\check{\mathbf{S}} \in \mathbb{R}^{m\times n}$ be the debiased sampling matrix in (1). Then, there exists $C > 0$ independent of n, p so that for $m \geq C\theta_{\max}(\mathbf{X})p\log(p/\delta)$, $\delta \leq n^{-(3+\alpha)}$, when conditioned on an event ζ that holds with probability at least $1-\delta$, the debiased subsampled OLS $\check{\boldsymbol{\beta}}$ in (7) satisfies*

$$\mathrm{Bias}_\zeta(\check{\boldsymbol{\beta}}) = \epsilon^2\|\mathbf{r}\|^2, \ \mathrm{Var}_\zeta(\check{\boldsymbol{\beta}}) = \Delta(\mathbf{X}) + \epsilon\|\mathbf{r}\|^2, \ \epsilon = O\Big(\sqrt{(\log n/\log\log n)^2 \cdot \theta^3_{\max}(\mathbf{X})p^3/m^3}\Big),$$

with $\mathrm{Bias}_\zeta(\cdot)$ and $\mathrm{Var}_\zeta(\cdot)$ in Definition 4.1, and $\Delta(\mathbf{X})$ defined in (8).

Comparing the bias-variance characterizations of the *debiased* solution $\check{\boldsymbol{\beta}}$ in Theorem 4.4 with those of the classical subsampled OLS $\tilde{\boldsymbol{\beta}}$ (without debiasing) in Theorems 4.2 and 4.3, we find that the *debiased* solution $\check{\boldsymbol{\beta}}$ yields a clear improvement. Specifically, it achieves a *smaller bias* (of order $\tilde{O}(\theta^3_{\max}(\mathbf{X})p^3/m^3)$ compared with the lower bound $\Omega(p^2/m^2)$ in Theorem 4.2 without debiasing), while incurring *no larger variance*, up to logarithmic factors.

For sketched OLS, [14] show that LESS embeddings nearly match sub-Gaussian sketches in variance, achieving $\mathrm{Var}_\zeta(\tilde{\boldsymbol{\beta}}) = p/(m-p-1) \cdot \|\mathbf{r}\|^2 + \tilde{O}(\sqrt{p}/m)\|\mathbf{r}\|^2$. However, such a characterization remains absent for subsampled OLS under popular random sampling schemes such as leverage score sampling and SRHT, for which existing bounds generally take the form $\mathrm{Var}_\zeta(\tilde{\boldsymbol{\beta}}) = \tilde{O}(p/m)\|\mathbf{r}\|^2$; see [8, 58]. The results in Theorems 4.3 and 4.4 close this gap by providing a finer-grained analysis for subsampled OLS, for both classical and debiased sampling schemes. In particular, we show that the previously established coarse bound $\tilde{O}(p/m)\|\mathbf{r}\|^2$ can be further refined as $\tilde{O}(p/m)\|\mathbf{r}\|^2 = \Delta(\mathbf{X}) + \tilde{O}(\sqrt{\theta^3_{\max}(\mathbf{X})p^3/m^3})\|\mathbf{r}\|^2$. See also Table 1 for a summary of these results.

4.2 Bias–variance characterizations for leverage score sampling and SRHT

In this subsection, we focus on the two widely-used special cases of leverage score sampling and subsampled randomized Walsh–Hadamard transform (SRHT). We show that the matrix-level debiasing in (1) is indeed *not* necessary (as opposed to general sampling schemes considered in Section 4.1), at least *to some extent*, for these two special cases of random sampling schemes.

Recall from our discussion in Remark 3.2 that in the case of *exact* and close-to-exact leverage score sampling, our proposed debiasing approach in (1) also becomes a simple *scalar* rescaling. As a consequence, one may expect that classical subsampled OLS solution $\tilde{\boldsymbol{\beta}}$ yields *small bias* in a similar spirit. This is given in the following result, the proof of which is given in Appendix E.4.

Corollary 4.5 (Bias–variance characterizations for subsampled OLS under leverage score sampling). *Under the settings and notations of Theorem 4.4, for $\mathbf{S} \in \mathbb{R}^{m\times n}$ the sampling matrix with sampling probabilities $\pi_i \in [\ell_i(\mathbf{X})/(p\theta_{\max}(\mathbf{X})), \ell_i(\mathbf{X})/(p\theta_{\min}(\mathbf{X}))]$ with $\theta_{\min}(\mathbf{X}) \in$*

$[1/2, 1]$ *as defined in Definition 2.3, there exists* $C > 0$, $\alpha > 0$, $\delta \le n^{-(3+\alpha)}$ *such that for* $m \ge C\theta_{\max}(\mathbf{X})p\log(p/\delta)$, *when conditioned on the event* ζ *that holds with the probability at least* $1-\delta$, *the standard subsampled OLS* $\tilde{\boldsymbol{\beta}}$ *in* (6) *satisfies*

$$\mathrm{Bias}_\zeta(\tilde{\boldsymbol{\beta}}) = \epsilon^2 \cdot \|\mathbf{r}\|^2, \ \textit{and} \ \ \mathrm{Var}_\zeta(\tilde{\boldsymbol{\beta}}) = \Delta(\mathbf{X}) + \epsilon \cdot \|\mathbf{r}\|^2,$$

with $\epsilon = O(\sqrt{(\log n/\operatorname{loglog} n)^2 \cdot \theta^3_{\max}(\mathbf{X})p^3/m^3} + \epsilon_\theta \cdot \theta_{\max}(\mathbf{X})p/m)$, $\epsilon_\theta = \max\{\theta^{-1}_{\min}(\mathbf{X}) - 1, 1 - \theta^{-1}_{\max}(\mathbf{X})\}$, $\mathrm{Bias}_\zeta(\cdot), \mathrm{Var}_\zeta(\cdot)$ *in Definition 4.1, and* $\Delta(\mathbf{X})$ *in* (8).

SRHT [1] provides a related construction: a randomized orthogonal transform spreads the leverage scores before uniform sampling. We first recall its definition as follows.

Definition 4.6 (**Subsampled randomized Walsh–Hadamard transform**, **SRHT**, [1])**.** For a data matrix $\mathbf{X} \in \mathbb{R}^{n\times p}$ of rank p with $n \ge p$ and a response vector $\mathbf{y} \in \mathbb{R}^n$, assume without loss of generality that $n = 2^q$ for some integer q. Then, the SRHT of $\mathbf{X}$ and $\mathbf{y}$ are given by

$$\tilde{\mathbf{X}}_{\mathrm{SRHT}} = \mathbf{S}\mathbf{H}_n\mathbf{D}_n\mathbf{X}/\sqrt{n} \in \mathbb{R}^{m\times p}, \quad \tilde{\mathbf{y}}_{\mathrm{SRHT}} = \mathbf{S}\mathbf{H}_n\mathbf{D}_n\mathbf{y}/\sqrt{n} \in \mathbb{R}^m,$$

respectively, for uniform random sampling matrix $\mathbf{S} \in \mathbf{R}^{m\times n}$ (with $\pi_i = 1/n$ in Definition 2.1), $\mathbf{H}_n \in \mathbb{R}^{n\times n}$ the Walsh–Hadamard matrix of size n, and diagonal $\mathbf{D}_n \in \mathbb{R}^{n\times n}$ having i.i.d. Rademacher random variables on its diagonal. And we define, similarly to the subsampled OLS solution $\tilde{\boldsymbol{\beta}}$ in (6), the subsampled OLS solution using SRHT as

$$\tilde{\boldsymbol{\beta}}_{\mathrm{SRHT}} = \tilde{\mathbf{X}}^\dagger_{\mathrm{SRHT}}\tilde{\mathbf{y}}_{\mathrm{SRHT}}. \tag{9}$$

The identities $\mathbf{H}_n^\top\mathbf{H}_n/n = \mathbf{I}_n$ and $\mathbf{D}_n^2 = \mathbf{I}_n$ imply $\mathbf{X}^\top\mathbf{D}_n\mathbf{H}_n^\top\mathbf{H}_n\mathbf{D}_n\mathbf{X}/n = \mathbf{X}^\top\mathbf{X}$, with a corresponding identity for $\mathbf{X}^\top\mathbf{y}$. Moreover, the transform is known to effectively spread the leverage scores of (the rows of) the data matrix $\mathbf{X}$, in the sense that, with high probability, the leverage scores of $\mathbf{H}_n\mathbf{D}_n\mathbf{X}/\sqrt{n}$ satisfy $\ell_i(\mathbf{H}_n\mathbf{D}_n\mathbf{X}/\sqrt{n}) = p/n \pm \tilde{O}(\sqrt{p}/n)$, for all i, see, for example, [26], [55, Theorems 3.1 and 3.2], and [45, Lemma E.4]. As a consequence, SRHT can be interpreted as an instance of approximate leverage score sampling applied to the transformed matrix $\mathbf{H}_n\mathbf{D}_n\mathbf{X}/\sqrt{n}$, and is therefore expected to be nearly unbiased per Corollary 4.5. This intuition is formalized in the following result; see Appendix E.5 for the proof.

Corollary 4.7 (**Bias-variance characterizations for subsampled OLS under SRHT**)**.** *Under the setting and notations of Theorem 4.4, let* $\mathbf{X}_{\mathrm{HD}} = \mathbf{H}_n\mathbf{D}_n\mathbf{X}/\sqrt{n}$, *and assume that* $\min_{i,\|\mathbf{x}_{\mathrm{HD},i}\|>0} \|\mathbf{x}_{\mathrm{HD},i}\|^2/\|\mathbf{X}_{\mathrm{HD}}\|_F^2 \ge n^{-\alpha}$ *for some constant* $\alpha > 0$. *For* $\tilde{\mathbf{X}}_{\mathrm{SRHT}} \in \mathbb{R}^{m\times p}$ *the SRHT of* $\mathbf{X}$ *and* $\tilde{\mathbf{y}}_{\mathrm{SRHT}} \in \mathbb{R}^m$ *the SRHT of* $\mathbf{y}$ *as in Definition 4.6, there exists* $C > 0$, $n\exp(-p) < \delta \le n^{-(3+\alpha)}$ *such that for* $m \ge Cp\log(p/\delta)$, *when conditioned on an event* ζ *that holds with probability at least* $1-\delta$, *the subsampled OLS with SRHT* $\tilde{\boldsymbol{\beta}}_{\mathrm{SRHT}}$ *in* (9) *satisfies*

$$\mathrm{Bias}_\zeta(\tilde{\boldsymbol{\beta}}_{\mathrm{SRHT}}) = \epsilon^2 \cdot \|\mathbf{r}\|^2, \ \textit{and} \ \ \mathrm{Var}_\zeta(\tilde{\boldsymbol{\beta}}_{\mathrm{SRHT}}) = \Delta(\mathbf{X}_{\mathrm{HD}}) + \epsilon \cdot \|\mathbf{r}\|^2,$$

with $\epsilon = O(\sqrt{(\log n/\operatorname{loglog} n)^2 \cdot p^3/m^3} + \sqrt{p\log(n/\delta)}/m)$.

Table 1 summarizes existing upper bounds on the bias of subsampled OLS estimator established in prior work under uniform sampling, leverage score sampling, and SRHT, and contrasts them with our results on debiased subsampled OLS. We see that the proposed debiased approach *consistently* achieves a substantially smaller bias across all three sampling schemes.

4.3 Numerical results for subsampled OLS

Figure 1 provides empirical evidence supporting our theoretical findings on classical and debiased subsampled OLS, by evaluating the effect of sketch size m on the (relative) subsampled OLS bias, $(L(\hat{\mathbb{E}}[\boldsymbol{\beta}]) - L(\boldsymbol{\beta}_{\mathrm{OLS}}))/L(\boldsymbol{\beta}_{\mathrm{OLS}})$ and variance, $(\hat{\mathbb{E}}[L(\boldsymbol{\beta})] - L(\boldsymbol{\beta}_{\mathrm{OLS}}))/L(\boldsymbol{\beta}_{\mathrm{OLS}})$, with $L(\cdot)$ as in (5). We use a matrix $\mathbf{X} \in \mathbb{R}^{n\times p}$ from the MSD dataset [5], and $\mathbf{y} \in \mathbb{R}^n$ is the corresponding response vector. See Appendix H for implementation details and further experiments.

We see from the left panel of Figure 1 that approximate leverage score sampling (**Lev**) and **SRHT** (see Definition 4.6) have *nearly identical* bias with and without debiasing, consistent with our Corollaries 4.5 and 4.7, respectively. Uniform sampling (**UNI**) exhibits the largest bias, while its debiased counterpart using exact leverage scores (**DUNI**) substantially reduces the bias, bringing

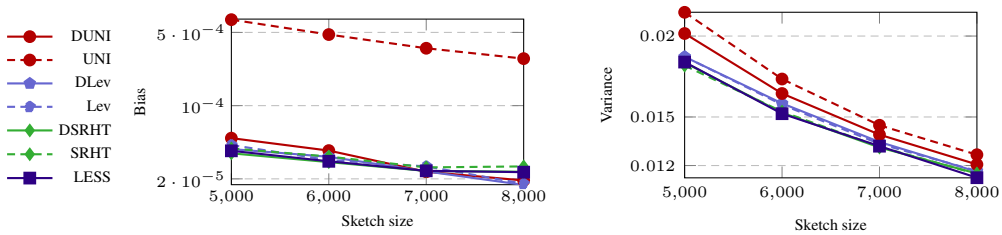

Figure 1: Bias and variance as functions of the sketch size m, comparing debiased sampling (in solid lines) and standard sampling (in dashed lines) on MSD dataset. **DUNI**, **DLev**, **DSRHT** are the corresponding debiased versions. Expectations are estimated from 500 independent runs.

its performance close to that of Lev, SRHT, their debiased variants (**DLev** and **DSRHT**), and the random projection approach LESS [16, 29]. The right panel of Figure 1 shows that Lev, SRHT, DLev, DSRHT, and LESS have *comparable* variances, whereas those of UNI and DUNI are slightly larger, with DUNI marginally lower than UNI.

Overall, we see from Figure 1 that debiasing empirically reduces bias without increasing variance. Similar trends are observed on the Flight Delay dataset; see Appendix H.3.

4.4 Extension to fast CUR decomposition

CUR decompositions construct approximations from selected rows and columns of the data matrix, and preserve structural properties such as sparsity and non-negativity. Given a data matrix $\mathbf{X} \in \mathbb{R}^{n\times p}$, a CUR decomposition selects c columns and r rows of $\mathbf{X}$ to get $\mathbf{C} \in \mathbb{R}^{n\times c}$ and $\mathbf{R} \in \mathbb{R}^{r\times p}$, and then computes the middle matrix $\mathbf{U} \in \mathbb{R}^{c\times r}$ by minimizing the reconstruction error $\|\mathbf{CUR} - \mathbf{X}\|_F^2$ [43, 53]:

$$\mathbf{U}_{\mathrm{CUR}} = \underset{\mathbf{U}\in\mathbb{R}^{c\times r}}{\arg\min} \|\mathbf{CUR} - \mathbf{X}\|_F^2 = \mathbf{C}^\dagger \mathbf{X} \mathbf{R}^\dagger. \tag{10}$$

While the solution $\mathbf{U}_{\mathrm{CUR}}$ in (10) provides the optimal reconstruction for the selected factors, its pseudoinverses and matrix multiplications can be costly for large data.

The OLS estimator in (5) is the one-sided counterpart of $\mathbf{U}_{\mathrm{CUR}}$ in (10), under the specialization $\mathbf{X} = \mathbf{y} \in \mathbb{R}^n$ and $\mathbf{R} = 1$. In this sense, subsampled OLS in (6) improves the computational efficiency by sketching on the left. Fast CUR decomposition [57, 59] extends this idea to both sides:

$$\tilde{\mathbf{U}} = \underset{\mathbf{U}\in\mathbb{R}^{c\times r}}{\arg\min} \|\mathbf{S_C CUR S_R}^\top - \mathbf{S_C X S_R}^\top\|_F^2 = (\mathbf{S_C C})^\dagger \mathbf{S_C X S_R}^\top (\mathbf{R S_R}^\top)^\dagger, \tag{11}$$

where $\mathbf{S_C} \in \mathbb{R}^{m_c\times n}$ and $\mathbf{S_R} \in \mathbb{R}^{m_r\times p}$ are independent sampling matrices; see Algorithm 1 in Appendix G for details. As discussed in Section 3, the sketches in the random oblique projections defining $\tilde{\mathbf{U}}$ may introduce systematic bias and could degrade the performance of fast CUR. Motivated by this connection, we extend our single sketch-induced oblique-projection analysis and debiasing strategy from subsampled OLS in (7) to CUR, and propose the following *debiased* CUR solution:

$$\check{\mathbf{U}} = (\check{\mathbf{S}}_\mathbf{C}\mathbf{C})^\dagger (\check{\mathbf{S}}_\mathbf{C}\mathbf{X}\check{\mathbf{S}}_\mathbf{R}^\top)(\mathbf{R}\check{\mathbf{S}}_\mathbf{R}^\top)^\dagger. \tag{12}$$

Here, $\check{\mathbf{S}}_\mathbf{C} \in \mathbb{R}^{m_c\times n}$ and $\check{\mathbf{S}}_\mathbf{R} \in \mathbb{R}^{m_r\times p}$ are the debiased sampling matrices as in Theorem 3.1.

Table 1: Bias characterizations for classical subsampled OLS established in previous efforts versus that for the proposed debiased subsampled OLS in Theorem 4.4, under uniform sampling (**UNI**), (exact/approximate) leverage score sampling (**Lev**) with $\theta_{\min}(\mathbf{X})/(2\theta_{\min}(\mathbf{X}) - 1) \le \theta_{\max}(\mathbf{X}) \le 1/(1 - \sqrt{\theta_{\max}(\mathbf{X})p/m})$, and **SRHT** with $p \ge \log(n/\delta)$.

Reference	**UNI**	**Lev**	**SRHT**
[4]	$\tilde{O}\left(\sqrt{\theta_{\max}^3(\mathbf{X})p^3/m^3}\right)$	$\tilde{O}\left(\sqrt{p^3/m^3}\right)$	$\tilde{O}\left(\sqrt{p^3/m^3}\right)$
[56]	$\tilde{O}\left(\theta_{\max}^2(\mathbf{X})p^2/m^2\right)$	$\tilde{O}\left(p^2/m^2\right)$	$\tilde{O}\left(p^2/m^2\right)$
This work	$\tilde{O}\left(\theta_{\max}^3(\mathbf{X})p^3/m^3\right)$	$\tilde{O}\left(p^3/m^3\right)$	$\tilde{O}\left(p^3/m^3 + p/m^2\right)$

The following result characterizes the bias of the proposed debiased solution $\check{\mathbf{U}}$. Full bias–variance characterizations and their proofs are deferred to Appendix G.

Theorem 4.8 (Precise bias characterization for debiased fast CUR decomposition). *For a data matrix $\mathbf{X} \in \mathbb{R}^{n\times p}$, let $\mathbf{C} \in \mathbb{R}^{n\times c}$ and $\mathbf{R} \in \mathbb{R}^{r\times p}$ be the pre-selected columns and rows of $\mathbf{X}$, and let $\mathbf{U}_{\mathrm{CUR}}$ be defined as in* (10)*. For standard random sampling matrices $\mathbf{S}_{\mathbf{C}}$ and $\mathbf{S}_{\mathbf{R}}$ used in Algorithm 1 in Appendix G, define their corresponding* debiased *sampling matrices $\check{\mathbf{S}}_{\mathbf{C}} \in \mathbb{R}^{m_c\times n}$ and $\check{\mathbf{S}}_{\mathbf{R}} \in \mathbb{R}^{m_r\times p}$ as in* (1)*. Then, there exists $C > 0$ independent of n, p, c, r so that for $m_c \geq C\theta_{\max}(\mathbf{C})c\log(c/\delta)$ and $m_r \geq C\theta_{\max}(\mathbf{R}^\top)r\log(r/\delta)$, when conditioned on an event ζ that holds with probability at least $1-\delta$, the debiased matrix $\check{\mathbf{U}}$ defined in* (12) *satisfies,*

$$\mathrm{Bias}_\zeta(\check{\mathbf{U}}) = \epsilon^2 \cdot L(\mathbf{U}_{\mathrm{CUR}})$$

with $\epsilon = \tilde{O}(\sqrt{\theta^3_{\max}(\mathbf{C})c^3/m_c^3}) + \tilde{O}(\sqrt{\theta^3_{\max}(\mathbf{R}^\top)r^3/m_r^3})$, for $L(\mathbf{U}) \equiv \|\mathbf{CUR}-\mathbf{X}\|_F^2$, the associated bias $\mathrm{Bias}_\zeta(\check{\mathbf{U}}) \equiv L(\mathbb{E}_\zeta[\check{\mathbf{U}}]) - L(\mathbf{U}_{\mathrm{CUR}})$, in a spirit similar to that of subsampled OLS in Definition 4.1. Here, $\theta_{\max}(\mathbf{C})$ and $\theta_{\max}(\mathbf{R}^\top)$ as in Definition 2.3 denote the maximum importance sampling approximation factors for $\mathbf{C}$ and $\mathbf{R}^\top$, respectively.

In a spirit similar to Corollaries 4.5 and 4.7, Corollaries G.2 and G.3 in Appendix G.2 also show that debiasing is unnecessary for fast CUR under leverage score sampling and SRHT. For completeness, we report the numerical results for fast CUR in Appendix H.4.

5 Conclusion and perspectives

We develop a non-asymptotic framework for random oblique projections and reveal a systematic bias arising from the nonlinear interaction between random sampling and the Moore–Penrose pseudoinverse, a phenomenon beyond the reach of classical subspace embedding analyses. To address this issue, we introduce a unified debiasing strategy that corrects this structural bias while preserving computational efficiency. When applied to subsampled least squares, our theory yields sharp bias and variance characterizations and shows that widely used schemes, including leverage score sampling and SRHT, are statistically suboptimal in general. For fast CUR decomposition, we construct a debiased estimator with provable improvements in approximation accuracy.

More broadly, our results highlight the need for bias-aware analysis in sketching-based algorithms. Extending this perspective to other methods, including sketch-and-solve Newton algorithms, randomized preconditioning, and streaming low-rank approximation, may further advance the statistical understanding of randomized numerical linear algebra.

Acknowledgments and Disclosure of Funding

Z. Liao and C. Niu were supported by the National Key Research and Development Program of China (No. 2025YFA1018600), the National Natural Science Foundation of China (via fund NSFC-12571561), and the Fundamental Research Support Program of HUST (2025BRSXB0004). MD and SG were supported in part by NSF CAREER Grant CCF-233865 and a Google ML and Systems Junior Faculty Award.

References

[1] 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.

[2] Greg W. Anderson, Alice Guionnet, and Ofer Zeitouni. *An Introduction to Random Matrices*, volume 118 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, 2010.

[3] Haim Avron, Kenneth L. Clarkson, and David P. Woodruff. Faster kernel ridge regression using sketching and preconditioning. *SIAM Journal on Matrix Analysis and Applications*, 38(4):1116–1138, 2017.

[4] Burak Bartan and Mert Pilanci. Distributed sketching for randomized optimization: Exact characterization, concentration and lower bounds. *arXiv preprint arXiv:2203.09755*, 2022.

[5] Thierry Bertin-Mahieux, Daniel PW Ellis, Brian Whitman, and Paul Lamere. The million song dataset. 2011.

[6] Raghu Bollapragada, Richard H Byrd, and Jorge Nocedal. Exact and inexact subsampled Newton methods for optimization. *IMA Journal of Numerical Analysis*, 39(2):545–578, 2019.

[7] Emmanuel J. Candès and Michael B. Wakin. An introduction to compressive sampling. *IEEE Signal Processing Magazine*, 25(2):21–30, March 2008.

[8] Xue Chen and Eric Price. Active regression via linear-sample sparsification. In *Proceedings of the Thirty-Second Conference on Learning Theory*, volume 99 of *Proceedings of Machine Learning Research*, pages 663–695. PMLR, 25–28 Jun 2019.

[9] Jocelyn T Chi and Ilse C F Ipsen. A projector-based approach to quantifying total and excess uncertainties for sketched linear regression. *Information and Inference: A Journal of the IMA*, 11(3):1055–1077, 08 2021.

[10] Jocelyn T. Chi and Ilse C.F. Ipsen. Multiplicative perturbation bounds for multivariate multiple linear regression in schatten p-norms. *Linear Algebra and its Applications*, 624:87–102, 2021.

[11] Kenneth L Clarkson and David P Woodruff. Low-rank approximation and regression in input sparsity time. *Journal of the ACM (JACM)*, 63(6):1–45, 2017.

[12] Romain Couillet and Zhenyu Liao. *Random Matrix Methods for Machine Learning*. Cambridge University Press, 2022.

[13] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. ImageNet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 248–255, 2009.

[14] Michał Dereziński. Algorithmic Gaussianization through sketching: Converting data into sub-Gaussian random designs. In *Proceedings of Thirty Sixth Conference on Learning Theory*, pages 3137–3172. PMLR, July 2023.

[15] Michał Dereziński, Kenneth L Clarkson, Michael W Mahoney, and Manfred K Warmuth. Minimax experimental design: Bridging the gap between statistical and worst-case approaches to least squares regression. In *Conference on Learning Theory*, pages 1050–1069. PMLR, 2019.

[16] Michał Dereziński, Jonathan Lacotte, Mert Pilanci, and Michael W Mahoney. Newton-LESS: Sparsification without trade-offs for the sketched Newton update. In *Advances in Neural Information Processing Systems*, volume 34, pages 2835–2847. Curran Associates, Inc., 2021.

[17] Michal Derezinski, Feynman T Liang, Zhenyu Liao, and Michael W. Mahoney. Precise expressions for random projections: Low-rank approximation and randomized Newton. In *Advances in Neural Information Processing Systems*, volume 33, pages 18272—18283. Curran Associates, Inc., 2020.

[18] Michał Dereziński, Zhenyu Liao, Edgar Dobriban, and Michael Mahoney. Sparse sketches with small inversion bias. In *Proceedings of Thirty Fourth Conference on Learning Theory*, volume 134, pages 1467–1510. PMLR, 2021.

[19] Michał Dereziński and Michael W Mahoney. Recent and upcoming developments in randomized numerical linear algebra for machine learning. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 6470–6479, 2024.

[20] Michał Dereziński, Manfred K Warmuth, and Daniel Hsu. Unbiased estimators for random design regression. *Journal of Machine Learning Research*, 23(167):1–46, 2022.

[21] Michal Derezinski and Manfred KK Warmuth. Unbiased estimates for linear regression via volume sampling. *Advances in Neural Information Processing Systems*, 30, 2017.

[22] Edgar Dobriban and Sifan Liu. Asymptotics for sketching in least squares regression. In *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019.

[23] Yijun Dong and Per-Gunnar Martinsson. Simpler is better: a comparative study of randomized pivoting algorithms for CUR and interpolative decompositions. *Advances in Computational Mathematics*, 49(4), 2023.

[24] Petros Drineas, Malik Magdon-Ismail, Michael W. Mahoney, and David P. Woodruff. Fast approximation of matrix coherence and statistical leverage. *The Journal of Machine Learning Research*, 13(1):3475–3506, December 2012.

[25] Petros Drineas, Michael W Mahoney, and Shan Muthukrishnan. Sampling algorithms for l_2 regression and applications. In *Proceedings of the Seventeenth Annual ACM-SIAM Symposium on Discrete Algorithm*, pages 1127–1136, 2006.

[26] Petros Drineas, Michael W Mahoney, Shan Muthukrishnan, and Tamás Sarlós. Faster least squares approximation. *Numerische mathematik*, 117(2):219–249, 2011.

[27] Michael Elad and Michal Aharon. Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Transactions on Image Processing*, 15(12):3736–3745, December 2006.

[28] Zhou Fan and Zhichao Wang. Spectra of the Conjugate Kernel and Neural Tangent Kernel for linear-width neural networks. In *Advances in Neural Information Processing Systems*, volume 33, pages 7710–7721. Curran Associates, Inc., 2020.

[29] Sachin Garg, Kevin Tan, and Michał Dereziński. Distributed least squares in small space via sketching and bias reduction. In A. Globerson, L. Mackey, D. Belgrave, A. Fan, U. Paquet, J. Tomczak, and C. Zhang, editors, *Advances in Neural Information Processing Systems*, volume 37, pages 73745–73782. Curran Associates, Inc., 2024.

[30] Gene H Golub and Charles F Van Loan. *Matrix Computations*. Johns Hopkins University press, 2013.

[31] Lingyu Gu, Yongqi Du, Yuan Zhang, Di Xie, Shiliang Pu, Robert Qiu, and Zhenyu Liao. "Lossless" Compression of Deep Neural Networks: A High-dimensional Neural Tangent Kernel Approach. In *Advances in Neural Information Processing Systems*, volume 35, pages 3774–3787. Curran Associates, Inc., 2022.

[32] Ming Gu and Stanley C. Eisenstat. Efficient algorithms for computing a strong rank-revealing QR factorization. *SIAM Journal on Scientific Computing*, 17(4):848–869, 1996.

[33] L. R Haff. An identity for the Wishart distribution with applications. *Journal of Multivariate Analysis*, 9(4):531–544, 1979.

[34] Nathan Halko, Per-Gunnar Martinsson, and Joel A Tropp. Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions. *SIAM review*, 53(2):217–288, 2011.

[35] William B. Johnson and Joram Lindenstrauss. Extensions of Lipschitz mappings into a Hilbert space. *Contemporary Mathematics*, pages 189–206, 1984.

[36] Yehuda Koren, Robert Bell, and Chris Volinsky. Matrix factorization techniques for recommender systems. *Computer*, 42(8):30–37, August 2009.

[37] Alex Krizhevsky. Learning Multiple Layers of Features from Tiny Images. Technical report, University of Toronto, 2009.

[38] Jonathan Lacotte and Mert Pilanci. Adaptive and oblivious randomized subspace methods for high-dimensional optimization: Sharp analysis and lower bounds. *IEEE Transactions on Information Theory*, 68(5):3281–3303, 2022.

[39] Zhenyu Liao and Michael W. Mahoney. Random Matrix Theory for Deep Learning: Beyond Eigenvalues of Linear Models. *IEEE Signal Processing Magazine*, 43(2):93–106, March 2026.

[40] Ping Ma, Yongkai Chen, Xinlian Zhang, Xin Xing, Jingyi Ma, and Michael W Mahoney. Asymptotic analysis of sampling estimators for randomized numerical linear algebra algorithms. *Journal of Machine Learning Research*, 23(177):1–45, 2022.

[41] Ping Ma, Michael W Mahoney, and Bin Yu. A statistical perspective on algorithmic leveraging. *Journal of Machine Learning Research*, 16:861–911, 2015.

[42] Michael W. Mahoney. Randomized algorithms for matrices and data. *Foundations and Trends® in Machine Learning*, 3(2):123–224, 2011.

[43] Michael W. Mahoney and Petros Drineas. CUR matrix decompositions for improved data analysis. *Proceedings of the National Academy of Sciences*, 106(3):697–702, 2009.

[44] Song Mei and Andrea Montanari. The generalization error of random features regression: Precise asymptotics and the double descent curve. *Communications on Pure and Applied Mathematics*, 75(4):667–766, 2022.

[45] Chengmei Niu, Zhenyu Liao, Zenan Ling, and Michael W Mahoney. Fundamental bias in inverting random sampling matrices with application to sub-sampled Newton. *arXiv preprint arXiv:2502.13583*, 2025.

[46] Taejun Park and Yuji Nakatsukasa. Accuracy and stability of CUR decompositions with oversampling. *SIAM Journal on Matrix Analysis and Applications*, 46(1):780–810, 2025.

[47] Jeffrey Pennington and Pratik Worah. Nonlinear random matrix theory for deep learning. In *Advances in Neural Information Processing Systems*, volume 30 of *NIPS'17*, pages 2637–2646. Curran Associates, Inc., 2017.

[48] Eric Price, Zhao Song, and David P Woodruff. Fast regression with an ℓ_∞ guarantee. *arXiv preprint arXiv:1705.10723*, 2017.

[49] Garvesh Raskutti and Michael W. Mahoney. A statistical perspective on randomized sketching for ordinary least-squares. *Journal of Machine Learning Research*, 17(213):1–31, 2016.

[50] Farbod Roosta-Khorasani and Michael W Mahoney. Sub-sampled Newton methods. *Mathematical Programming*, 174:293–326, 2019.

[51] Jack W. Silverstein and Zhidong Bai. On the empirical distribution of eigenvalues of a class of large dimensional random matrices. *Journal of Multivariate Analysis*, 54(2):175–192, 1995.

[52] Zhao Song, Mingquan Ye, Junze Yin, and Lichen Zhang. A nearly-optimal bound for fast regression with ℓ_∞ guarantee. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pages 32463–32482. PMLR, 23–29 Jul 2023.

[53] D. C. Sorensen and Mark Embree. A DEIM induced CUR factorization. *SIAM Journal on Scientific Computing*, 38(3):A1454–A1482, 2016.

[54] Lloyd N Trefethen and David Bau. *Numerical Linear Algebra*. SIAM, 2022.

[55] Joel A Tropp. Improved analysis of the subsampled randomized Hadamard transform. *Advances in Adaptive Data Analysis*, 3(01n02):115–126, 2011.

[56] Shusen Wang, Alex Gittens, and Michael W. Mahoney. Sketched ridge regression: Optimization perspective, statistical perspective, and model averaging. *Journal of Machine Learning Research*, 18(218):1–50, 2018.

[57] Shusen Wang, Zhihua Zhang, and Tong Zhang. Towards more efficient SPSD matrix approximation and CUR matrix decomposition. *Journal of Machine Learning Research*, 17(209):1–49, 2016.

[58] David P. Woodruff. Sketching as a tool for numerical linear algebra. *Foundations and Trends® in Theoretical Computer Science*, 10(1–2):1–157, 2014.

[59] Haishan Ye, Shusen Wang, Zhihua Zhang, and Tong Zhang. Fast generalized matrix regression with applications in machine learning. *arXiv preprint arXiv:1912.12008*, 2019.

Supplementary Material of
Debiasing Random Oblique Projections for Subsampled OLS and Fast CUR in High Dimensions

The technical appendices are organized as follows.

- Appendix A reviews the literature most relevant to our work.
- Appendix B collects the technical lemmas used throughout the paper.
- Appendix C presents the lower bound on the bias of the classical subsampled OLS estimator in Theorem 4.2.
- Appendix D proves the variance upper bound for the classical subsampled OLS estimator in Theorem 4.3.
- Appendix E derives the bias and variance results for the debiased subsampled OLS estimator in Theorem 4.4.
- Appendix F shows the proof of the statistical characterizations of the debiased oblique projection in Theorem 3.1.
- Appendix G contains the full bias-variance characterizations and the proof of the debiased CUR decomposition in Theorems 4.8 and G.1.
- Appendix H provides implementation details for the numerical experiments in Section 4.3, additional results for subsampled OLS, and numerical results for fast CUR decomposition.

A Related work

Our work connects to three closely related lines of research: subsampled OLS, sketching-based CUR decomposition, and recent analyses based on random matrix theory beyond classical asymptotics. We briefly review these directions below and position our contributions relative to prior work.

Classical subsampled OLS. Subsampled OLS estimators are a classical tool for accelerating least-squares problems by reducing the problem size while approximately preserving the information in the original design matrix [25, 26, 56, 4]. Beyond computational efficiency, a growing body of work has studied its statistical properties in the ℓ_2 norm, including bias and variance [41, 49, 56, 22, 15, 4, 20]. Despite this progress, existing analyses do not provide a precise characterization of the bias between the subsampled OLS estimator and the full OLS solution, except for when using specialized determinantal sampling schemes, which are computationally expensive [20]. A central difficulty is the lack of a fine-grained understanding of the oblique projection operator $\tilde{\mathbf{P}}$ induced by random sampling. As a result, prior work typically relies on coarse bounds or asymptotic arguments that do not fully capture the mechanism by which sampling introduces bias. Our work addresses this gap by leveraging precise characterizations of the oblique projector $\tilde{\mathbf{P}}$ to obtain a more refined analysis of the bias in subsampled OLS estimators.

A complementary line of work studies the coordinate-wise accuracy of sketched least-squares estimators. In particular, [48, 52] establish high-probability ℓ_∞ guarantees for sketches satisfying oblivious subspace embedding (OSE) and oblivious coordinate-wise embedding (OCE) properties, including SRHT-type sketches. These results complement the ℓ_2 bias and variance perspective considered here.

CUR decomposition. Traditionally, low-rank matrix approximations are obtained via SVD-based methods, such as truncated SVD or randomized SVD [34], which operate in the space of matrix singular vectors. In contrast, CUR decomposition constructs low-rank approximations directly from selected rows and columns of the data matrix [43], thereby preserving structural properties such as sparsity, non-negativity, and interpretability. These properties are often important in applications such as signal and image processing [7, 27], recommendation systems [36], and big data analysis [43]. Given a matrix $\mathbf{X} \in \mathbb{R}^{n\times p}$, a key design choice in CUR decomposition is the middle factor $\mathbf{U}$. Two common options are $\mathbf{U} = \mathbf{C}^\dagger \mathbf{X} \mathbf{R}^\dagger$, where $\mathbf{C}$ and $\mathbf{R}$ consist of sampled columns and rows of $\mathbf{X}$, and $\mathbf{U} = \mathbf{X}_{R,C}^\dagger$, where $\mathbf{X}_{R,C}$ denotes their intersection. The former is more robust but computationally

expensive, while the latter is more efficient but can be unstable when $\mathbf{X}_{R,C}$ is nearly singular. Classical approaches mitigate this instability via pivoting-based methods, such as column-pivoted QR and LU with complete pivoting [30, 54], which enjoy strong theoretical guarantees [32] but are often impractical for large-scale problems.

To improve scalability, recent work has proposed randomized sketching methods. For $\mathbf{U} = \mathbf{X}_{R,C}^{\dagger}$, pivoting is typically performed on a randomized sketch $\tilde{\mathbf{X}}$ instead of $\mathbf{X}$ [23, 46], with oversampling shown to improve stability and accuracy [46]. For another choice $\mathbf{U} = \mathbf{C}^{\dagger}\mathbf{X}\mathbf{R}^{\dagger}$, existing approaches rely on the sketches of $\mathbf{C}$, $\mathbf{R}$, and $\mathbf{X}$ to form estimators such as $\tilde{\mathbf{C}}^{\dagger}\tilde{\mathbf{X}}\tilde{\mathbf{R}}^{\dagger}$ [57, 59]. However, existing analyses rely largely on subspace embedding arguments and provide only coarse guarantees, without precisely characterizing how random sketching affects the induced projection operators $\mathbf{C}\tilde{\mathbf{C}}^{\dagger}$ and $\tilde{\mathbf{R}}^{\dagger}\mathbf{R}$. In this work, we develop a fine-grained analysis of these random-sampling-based estimators by precisely characterizing the associated projection operators, leading to sharper guarantees and improved approximations of $\mathbf{C}^{\dagger}\mathbf{X}\mathbf{R}^{\dagger}$.

Random matrix theory. Our analysis is closely related to recent advances inspired by asymptotic random matrix theory (RMT) [12], which studies the spectral behavior of large-dimensional random matrices [2]. RMT-based techniques have recently been adopted in large-scale machine learning to analyze optimization and generalization properties [47, 28, 44, 12]. In particular, building on tools such as Stieltjes transforms, several works have developed fine-grained characterizations of inverse Gram matrices. Using Sherman-Morrison rank-one updates, [16, 18, 29, 45] derive precise expressions for $\mathbb{E}[(\tilde{\mathbf{X}}^{\top}\tilde{\mathbf{X}})^{-1}]$ beyond classical asymptotic analyses. Our work is closely connected to this line of research and extends these techniques to obtain a fine-grained characterization of the oblique projection operator $\tilde{\mathbf{P}}$ under random sampling.

B Useful lemmas

In this section, we introduce a few technical lemmas to be used in subsequent sections.

Lemma B.1 (Sherman–Morrison formula)**.** *For an invertible matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$ and two vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$, $\mathbf{A} + \mathbf{u}\mathbf{v}^{\top}$ is invertible if and only if $1 + \mathbf{v}^{\top}\mathbf{A}^{-1}\mathbf{u} \neq 0$ and*

$$(\mathbf{A} + \mathbf{u}\mathbf{v}^{\top})^{-1} = \mathbf{A}^{-1} - \frac{\mathbf{A}^{-1}\mathbf{u}\mathbf{v}^{\top}\mathbf{A}^{-1}}{1 + \mathbf{v}^{\top}\mathbf{A}^{-1}\mathbf{u}}.$$

Besides, it also follows that

$$(\mathbf{A} + \mathbf{u}\mathbf{v}^{\top})^{-1}\mathbf{u} = \frac{\mathbf{A}^{-1}\mathbf{u}}{1 + \mathbf{v}^{\top}\mathbf{A}^{-1}\mathbf{u}}.$$

Lemma B.2 (Properties of the Moore–Penrose Pseudoinverse)**.** *Let $\mathbf{A} \in \mathbb{R}^{n\times p}$, $\mathbf{B} \in \mathbb{R}^{p\times d}$. Then,*

$$(\mathbf{A}\mathbf{B})^{\dagger} = \mathbf{B}^{\dagger}\mathbf{A}^{\dagger},$$

if any one of the following conditions holds: (1) $\mathbf{A}^{\top}\mathbf{A} = \mathbf{I}_p$; (2) $\mathbf{B}\mathbf{B}^{\top} = \mathbf{I}_p$; or (3) $\operatorname{rank}(\mathbf{A}) = \operatorname{rank}(\mathbf{B}) = p$.

Lemma B.3 (Subspace embedding for random sampling, [45, Lemma 2.7])**.** *Given $\mathbf{X} \in \mathbb{R}^{n\times p}$ of rank p with $n \geq p$, let $\mathbf{S} \in \mathbb{R}^{m\times n}$ be a random sampling matrix as in Definition 2.1. Then, there exists $C > 0$ independent of n, p such that for $m \geq C\theta_{\max}(\mathbf{X})p\log(p/\delta)/\epsilon^2$, failure probability $\delta \in (0, 1/2)$, $\epsilon > 0$, and $\theta_{\max}(\mathbf{X})$ in Definition 2.3, $\mathbf{X}^{\top}\mathbf{S}^{\top}\mathbf{S}\mathbf{X}$ is an (ϵ, δ)-approximation of $\mathbf{X}^{\top}\mathbf{X}$.*

Lemma B.3 corresponds to the special case $\mathbf{C} = 0$ of the regularized formulation in [45, Lemma 2.7]. It establishes explicit conditions on the sample size m under which $\mathbf{S}\mathbf{X}$ forms an (ϵ, δ)-subspace embedding of $\mathbf{X}$. This result serves as a technical foundation, ensuring that the sketch $\mathbf{S}\mathbf{X}$ is well behaved throughout the subsequent analysis.

C Proof of Theorem 4.2

We begin with the following result from [18], which is used in the proof of Theorem 4.2.

Lemma C.1 ([18, Lemma 36]). *There is a universal constant $C > 0$ such that for any positive integer b, if $x \sim$Binomial$(b, 0.5)$ then,*

$$\mathbb{E}\Big[\frac{1}{x + b/2}\Big] \geq \Big(1 + \frac{1}{Cb}\Big) \cdot \frac{1}{b}.$$

For notational convenience, in this proof we will denote $\boldsymbol{\beta}_{\mathrm{OLS}}$ by $\boldsymbol{\beta}^*$. Consider a matrix $\mathbf{X}$ whose first $2p$ rows are given as follows:

$$\mathbf{X}_{[1:2p]} = \begin{bmatrix} \frac{1}{2} & 0 & 0 & \cdots & 0 \\ \frac{\sqrt{3}}{2} & 0 & 0 & \cdots & 0 \\ 0 & \frac{1}{2} & 0 & \cdots & 0 \\ 0 & \frac{\sqrt{3}}{2} & 0 & \cdots & 0 \\ 0 & 0 & \frac{1}{2} & \cdots & 0 \\ 0 & 0 & \frac{\sqrt{3}}{2} & \cdots & 0 \\ 0 & 0 & 0 & \cdots & \frac{1}{2} \\ 0 & 0 & 0 & \cdots & \frac{\sqrt{3}}{2} \end{bmatrix}.$$

Also, if $n > 2p$, let rows subsequent to $(2p)^{th}$ row be all zeros. In addition, consider the response vector $\mathbf{y} \in \mathbb{R}^n$ as follows:

$$[\mathbf{y}]_i = \begin{cases} -1, & i \leq p \text{ and } i \text{ is odd}, \\ 1, & i \leq p \text{ and } i \text{ is even}, \\ 1, & p < i \leq 2p, \\ 0, & i > 2p. \end{cases}$$

Note that $\mathbf{X}^\top \mathbf{X} = \mathbf{I}_p$, the exact least squares estimator $\boldsymbol{\beta}^*$ can be found as $\boldsymbol{\beta}^* = \mathbf{X}^\top \mathbf{y}$. The entries of $\boldsymbol{\beta}^*$ are given as

$$\boldsymbol{\beta}^*_i = \begin{cases} \frac{\sqrt{3}}{2} - \frac{1}{2}, & i \leq \frac{p}{2}, \\ \frac{\sqrt{3}}{2} + \frac{1}{2}, & \frac{p}{2} < i \leq p. \end{cases}$$

Also, note that $\|\mathbf{y}\|^2 = 2p$, $\|\mathbf{X}\boldsymbol{\beta}^* - \mathbf{y}\|^2 = \Theta(p)$, $\|\mathbf{X}\boldsymbol{\beta}^*\|^2 = \Theta(p)$. The exact leverage scores of $\mathbf{X}$ lie in the set $\{\frac{1}{4}, \frac{3}{4}, 0\}$. Consider the probability distribution $\{\pi_i\}_{i=1}^n$, where $\pi_i = \frac{1}{2p}$ for $i \leq 2p$ and $\pi_i = 0$ for $i > 2p$. Note that $\{\pi_i\}_{i=1}^n$, provides a $\frac{1}{2}$-approximation to exact leverage score sampling. Let $m \geq p$, $\mathbf{S} \in \mathbb{R}^{m \times n}$ be sampling matrix sampled from the distribution $\{\pi_i\}_{i=1}^n$. Let $\tilde{\boldsymbol{\beta}} = (\mathbf{X}^\top \mathbf{S}^\top \mathbf{S} \mathbf{X})^\dagger \mathbf{X}^\top \mathbf{S}^\top \mathbf{S} \mathbf{y}$ be the sketched least squares estimator and $\bar{\boldsymbol{\beta}} = \mathbb{E}[\tilde{\boldsymbol{\beta}}]$. We can specify the distribution of every entry in $\tilde{\boldsymbol{\beta}}$. To see that, we start by noting that $\mathbf{X}^\top \mathbf{S}^\top \mathbf{S} \mathbf{X}$ is a diagonal matrix where

$$[\mathbf{X}^\top \mathbf{S}^\top \mathbf{S} \mathbf{X}]_{ii} = (\frac{3}{4} \cdot s_{2i} + \frac{1}{4} \cdot s_{2i-1}) \frac{2p}{m}, \text{ for } 1 \leq i \leq p.$$

where s_{2i} and s_{2i-1} denote the number of times $(2i)^{th}$ and $(2i-1)^{th}$ rows are selected while sampling $\mathbf{S}$ from $\{\pi_i\}_{i=1}^n$. Therefore,

$$[(\mathbf{X}^\top \mathbf{S}^\top \mathbf{S} \mathbf{X})^\dagger]_{ii} = \begin{cases} \frac{4}{3s_{2i} + s_{2i-1}} \cdot \frac{m}{2p}, & \text{if } s_{2i} + s_{2i-1} > 0, \\ 0, & \text{otherwise.} \end{cases}$$

The ith entry of $\tilde{\boldsymbol{\beta}}$ is given as

$$[\tilde{\boldsymbol{\beta}}]_i = \begin{cases} \frac{2(\sqrt{3}s_{2i} - s_{2i-1})}{3s_{2i} + s_{2i-1}}, & i \leq \frac{p}{2} \text{ and } s_{2i} + s_{2i-1} > 0, \\ 0, & i \leq \frac{p}{2} \text{ and } s_{2i} + s_{2i-1} = 0, \\ \frac{2(\sqrt{3}s_{2i} + s_{2i-1})}{3s_{2i} + s_{2i-1}}, & \frac{p}{2} < i \leq p \text{ and } s_{2i} + s_{2i-1} > 0, \\ 0, & \frac{p}{2} < i \leq p \text{ and } s_{2i} + s_{2i-1} = 0. \end{cases}$$

Note that, for $1 \le i \le \frac{p}{2}$, the distribution of $[\tilde{\boldsymbol{\beta}}]_i's$ are identical, and for $i > \frac{p}{2}$, the distribution of $[\tilde{\boldsymbol{\beta}}]_i's$ are identical. Due to this observation, $\mathbb{E}[\tilde{\boldsymbol{\beta}}]_i = \mathbb{E}[\tilde{\boldsymbol{\beta}}]_1$ for all $i \le \frac{p}{2}$ and $\mathbb{E}[\tilde{\boldsymbol{\beta}}]_i = \mathbb{E}[\tilde{\boldsymbol{\beta}}]_p$ for all $i > \frac{p}{2}$. We first provide an upper bound on $\mathbb{E}[\tilde{\boldsymbol{\beta}}]_1$. We have,

$$\mathbb{E}[\tilde{\boldsymbol{\beta}}]_1 = \mathbb{E}\Big[\mathbf{1}_{s_1+s_2>0} \cdot \frac{2\big((1+\sqrt{3})s_2 - (s_1+s_2)\big)}{2s_2 + (s_1+s_2)}\Big]$$

Let $b = s_1 + s_2$ and $x = s_2$

$$\begin{aligned}
\mathbb{E}[\tilde{\boldsymbol{\beta}}]_1 &= \mathbb{E}\left[\mathbf{1}_{\{b>0\}} \frac{2((1+\sqrt{3})x - b)}{2x+b}\right] \\
&= \mathbb{E}\left[\mathbf{1}_{\{b>0\}} \left(1 + \sqrt{3} - \frac{(3+\sqrt{3})b}{2x+b}\right)\right] \\
&= (1+\sqrt{3})\mathbb{E}\left[\mathbf{1}_{\{b>0\}} \left(1 - \frac{b\sqrt{3}}{2} \cdot \frac{1}{x + b/2}\right)\right].
\end{aligned}$$

Note that for any given fixed $b > 0$, $x \sim \text{Binomial}(b, 0.5)$. Using Lemma C.1, we get $\mathbb{E}\Big[\frac{1}{x+b/2}\Big] \ge \Big(1 + \frac{1}{Cb}\Big) \cdot \frac{1}{b}$ for some universal constant C. Therefore,

$$\begin{aligned}
\mathbb{E}[\tilde{\boldsymbol{\beta}}]_1 &\le (1+\sqrt{3})\mathbb{E}\Big[\mathbf{1}_{\{b>0\}}\Big(1 - \frac{\sqrt{3}}{2}\Big(1 + \frac{1}{Cb}\Big)\Big)\Big] \\
&= (1+\sqrt{3})\Big[\Big(1 - \frac{\sqrt{3}}{2}\Big)\Pr(b>0) - \frac{\sqrt{3}}{2C}\mathbb{E}\Big[\frac{\mathbf{1}_{\{b>0\}}}{b}\Big]\Big] \\
&= (1+\sqrt{3})\Big[\Big(1 - \frac{\sqrt{3}}{2}\Big)\Pr(b>0) - \frac{\sqrt{3}}{2C}\sum_{t=1}^{m}\frac{1}{t}\Pr(b=t)\Big] \\
&\le (1+\sqrt{3}) \cdot \Big[\Big(1 - \frac{\sqrt{3}}{2}\Big)\Pr(b>0) - \frac{\sqrt{3}}{2C} \cdot \sum_{t=1}^{2m/p}\frac{1}{t}\Pr(b=t)\Big] \\
&\le (1+\sqrt{3}) \cdot \Big[\Big(1 - \frac{\sqrt{3}}{2}\Big)\Pr(b>0) - \frac{p\sqrt{3}}{4mC} \cdot \sum_{t=1}^{2m/p}\Pr(b=t)\Big] \\
&\le (1+\sqrt{3}) \cdot \Big[\Big(1 - \frac{\sqrt{3}}{2}\Big)\Pr(b>0) - \frac{p\sqrt{3}}{32mC}\Big],
\end{aligned}$$

where the last inequality holds because $\sum_{t=1}^{2m/p}\Pr(b=t) \ge \frac{1}{8}$, as $b \sim \text{Binomial}(m, \frac{1}{p})$. Therefore, we get,

$$\begin{aligned}
\mathbb{E}[\tilde{\boldsymbol{\beta}}]_1 &\le \Big(1 - \frac{\sqrt{3}}{2} + \sqrt{3} - \frac{3}{2}\Big)\Pr(b>0) - \frac{3+\sqrt{3}}{32C} \cdot \frac{p}{m} \\
&= [\boldsymbol{\beta}^*]_1 \Pr(b>0) - \frac{3+\sqrt{3}}{32C} \cdot \frac{p}{m}.
\end{aligned}$$

Similarly, we can show that,

$$\mathbb{E}[\tilde{\boldsymbol{\beta}}]_p \ge \Big([\boldsymbol{\beta}^*]_p \Pr(b>0) + \frac{3-\sqrt{3}}{32C} \cdot \frac{p}{m}\Big).$$

Let q denote $\Pr(b>0)$. Therefore, we get

$$\begin{aligned}
\mathbb{E}[\tilde{\boldsymbol{\beta}}]_i &= \Big(q[\boldsymbol{\beta}^*]_i - c_1\Big) && \text{for } i \le \frac{p}{2}, \\
&= \Big(q[\boldsymbol{\beta}^*]_i + c_2\Big) && \text{for } i > \frac{p}{2},
\end{aligned}$$

and some $c_1, c_2 = \Omega(\frac{p}{m})$. Let $\bar{\boldsymbol{\beta}} = \mathbb{E}[\tilde{\boldsymbol{\beta}}]$ and γ be any scalar factor. We have,

$$\|\mathbf{X}\gamma\bar{\boldsymbol{\beta}} - \mathbf{y}\|^2 = \|\mathbf{X}\boldsymbol{\beta}^* - \mathbf{y}\|^2 + \|\mathbf{X}(\gamma\bar{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)\|^2$$

$$= \|\mathbf{X}\boldsymbol{\beta}^* - \mathbf{y}\|^2 + \|\mathbf{X}_{[1:p,1:\frac{p}{2}]}(\gamma\bar{\boldsymbol{\beta}}_{[1:\frac{p}{2}]} - \boldsymbol{\beta}^*_{[1:\frac{p}{2}]})\|^2$$
$$+ \|\mathbf{X}_{[p+1:2p,\frac{p}{2}+1:p]}(\gamma\bar{\boldsymbol{\beta}}_{[\frac{p}{2}+1:p]} - \boldsymbol{\beta}^*_{[\frac{p}{2}+1:p]})\|^2, \tag{13}$$

where by $\mathbf{X}_{[p+1:2p,\frac{p}{2}+1:p]}$ we mean the sub matrix of $\mathbf{X}$ formed by $(p+1)^{th}$ to $(2p)^{th}$ rows and $(\frac{p}{2}+1)^{th}$ to p^{th} columns. The other subscripts can be understood similarly. First, consider the term $\|\mathbf{X}_{[1:p,1:\frac{p}{2}]}(\gamma\bar{\boldsymbol{\beta}}_{[1:\frac{p}{2}]} - \boldsymbol{\beta}^*_{[1:\frac{p}{2}]})\|^2$. We get,

$$\begin{aligned}\Big\|\mathbf{X}_{[1:p,1:\frac{p}{2}]}(\gamma\bar{\boldsymbol{\beta}}_{[1:\frac{p}{2}]} - \boldsymbol{\beta}^*_{[1:\frac{p}{2}]})\Big\|^2 &= \Big\|\mathbf{X}_{[1:p,1:\frac{p}{2}]}\big(\gamma\cdot(q\boldsymbol{\beta}^*_{[1:\frac{p}{2}]} - c_1\mathbf{1}) - \boldsymbol{\beta}^*_{[1:\frac{p}{2}]}\big)\Big\|^2 \\ &= \Big\|\mathbf{X}_{[1:p,1:\frac{p}{2}]}\big(\gamma\cdot c_1\mathbf{1} + (1-q\gamma)\cdot\boldsymbol{\beta}^*_{[1:\frac{p}{2}]}\big)\Big\|^2 \\ &= \Big(\gamma\cdot c_1 + (1-q\gamma)\cdot[\boldsymbol{\beta}^*]_1\Big)^2\cdot\|\mathbf{X}_{[1:p,1:\frac{p}{2}]}\mathbf{1}\|^2,\end{aligned}$$

where $\mathbf{1}$ denotes a vector of 1's. As $\|\mathbf{X}_{[1:p,1:\frac{p}{2}]}\mathbf{1}\|^2 = \Theta(p)$, and also recall $\|\mathbf{X}\boldsymbol{\beta}^* - \mathbf{y}\|^2 = \Theta(p)$, we get

$$\Big\|\mathbf{X}_{[1:p,1:\frac{p}{2}]}(\gamma\bar{\boldsymbol{\beta}}_{[1:\frac{p}{2}]} - \boldsymbol{\beta}^*_{[1:\frac{p}{2}]})\Big\|^2 = \Big(\gamma\cdot c_1 + (1-q\gamma)\cdot[\boldsymbol{\beta}^*]_1\Big)^2\cdot\Theta(\|\mathbf{X}\boldsymbol{\beta}^* - \mathbf{y}\|^2).$$

If $0 < q\gamma \le 1$, then we get

$$\Big(\gamma\cdot c_1 + (1-q\gamma)\cdot[\boldsymbol{\beta}^*]_1\Big)^2 \ge \max\Big\{\gamma^2\cdot\Omega\big(\frac{p^2}{m^2}\big), (1-q\gamma)^2\cdot[\boldsymbol{\beta}^*]_1^2\Big\}.$$

At least one of γ or $1-q\gamma$ is at least $\frac{1}{2}$. In either case we get,

$$\Big(\gamma\cdot c_1 + (1-q\gamma)\cdot[\boldsymbol{\beta}^*]_1\Big)^2 = \Omega\big(\frac{p^2}{m^2}\big).$$

In case $\gamma \le 0$ or $q\gamma > 1$, consider the last term in relation (13). We have,

$$\begin{aligned}\|\mathbf{X}_{[p+1:2p,\frac{p}{2}+1:p]}(\gamma\bar{\boldsymbol{\beta}}_{[\frac{p}{2}+1:p]} - \boldsymbol{\beta}^*_{[\frac{p}{2}+1:p]})\|^2 &= \|\mathbf{X}_{[p+1:2p,\frac{p}{2}+1:p]}(\gamma\cdot(q\boldsymbol{\beta}^* + c_2\mathbf{1})_{[\frac{p}{2}+1:p]} - \boldsymbol{\beta}^*_{[\frac{p}{2}+1:p]})\|^2 \\ &= \|\mathbf{X}_{[p+1:2p,\frac{p}{2}+1:p]}(\gamma\cdot c_2\mathbf{1} - (1-q\gamma)\cdot\boldsymbol{\beta}^*_{[\frac{p}{2}+1:p]})\|^2 \\ &= \Big(\gamma\cdot c_2 - (1-q\gamma)\cdot[\boldsymbol{\beta}^*]_p\Big)^2\cdot\Theta(\|\mathbf{X}\boldsymbol{\beta}^* - \mathbf{y}\|^2) \\ &= \Omega\big(\frac{p^2}{m^2}\big)\cdot\Theta(\|\mathbf{X}\boldsymbol{\beta}^* - \mathbf{y}\|^2).\end{aligned}$$

The last inequality holds as either $\gamma \le 0$ or $q\gamma > 1$. This completes the proof of Theorem 4.2.

D Proof of Theorem 4.3

Let $\mathbf{U} \in \mathbb{R}^{n\times p}$ have orthonormal columns spanning the column space of $\mathbf{X}$. Then, $\boldsymbol{\beta}_{\mathrm{OLS}} = \mathbf{U}^\top\mathbf{y}$ and $\tilde{\boldsymbol{\beta}} = (\mathbf{U}^\top\mathbf{S}^\top\mathbf{S}\mathbf{U})^\dagger\mathbf{U}^\top\mathbf{S}^\top\mathbf{S}\mathbf{y}$. Furthermore,

$$\begin{aligned}L(\tilde{\boldsymbol{\beta}}) - L(\boldsymbol{\beta}_{\mathrm{OLS}}) &= \big\|\mathbf{U}(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_{\mathrm{OLS}})\big\|^2 \\ &= \big\|\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_{\mathrm{OLS}}\big\|^2.\end{aligned}$$

Let ζ denote the event that

$$(1+\epsilon)^{-1}\cdot\mathbf{I} \preceq \mathbf{U}^\top\mathbf{S}^\top\mathbf{S}\mathbf{U} \preceq (1+\epsilon)\cdot\mathbf{I},$$

for $\epsilon = \sqrt{\frac{3p\theta_{\max}(\mathbf{X})\log(2p/\delta)}{m}}$. First, note that for $m > 12p\theta_{\max}\log(2p/\delta)$, $\Pr(\zeta) \ge 1-\delta$. This claim is essentially the subspace embedding guarantee provided by the approximate leverage sampling matrix $\mathbf{S}$ (Lemma B.3). In what follows, we will upper bound $\mathbb{E}_\zeta[L(\tilde{\boldsymbol{\beta}}) - L(\boldsymbol{\beta}_{\mathrm{OLS}})]$. We have,

$$\begin{aligned}\mathbb{E}_\zeta[L(\tilde{\boldsymbol{\beta}}) - L(\boldsymbol{\beta}_{\mathrm{OLS}})] &= \mathbb{E}_\zeta[\|\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_{\mathrm{OLS}}\|^2] \\ &= \mathbb{E}_\zeta\Big[\Big\|(\mathbf{U}^\top\mathbf{S}^\top\mathbf{S}\mathbf{U})^{-1}\big(\mathbf{U}^\top\mathbf{S}^\top\mathbf{S}\mathbf{y} - (\mathbf{U}^\top\mathbf{S}^\top\mathbf{S}\mathbf{U})\boldsymbol{\beta}_{\mathrm{OLS}}\big)\Big\|^2\Big]\end{aligned}$$

$$
\begin{aligned}
&= \mathbb{E}_\zeta\Big[\Big\|(\mathbf{U}^\top\mathbf{S}^\top\mathbf{S}\mathbf{U})^{-1}\mathbf{U}^\top\mathbf{S}^\top\mathbf{S}(\mathbf{y}-\mathbf{U}\boldsymbol{\beta}_{\mathrm{OLS}})\Big\|^2\Big]\\
&\le \mathbb{E}_\zeta\Big[\big\|(\mathbf{U}^\top\mathbf{S}^\top\mathbf{S}\mathbf{U})^{-1}\big\|^2\cdot\big\|\mathbf{U}^\top\mathbf{S}^\top\mathbf{S}\mathbf{r}\big\|^2\Big]\\
&\le (1+3\epsilon)^2\cdot\mathbb{E}_\zeta\Big[\|\mathbf{U}^\top\mathbf{S}^\top\mathbf{S}\mathbf{r}\|^2\Big],
\end{aligned}
$$

where $\mathbf{r}$ denotes $\mathbf{y}-\mathbf{U}\boldsymbol{\beta}_{\mathrm{OLS}}$. The remaining term to upper bound is $\mathbb{E}_\zeta\Big[\|\mathbf{U}^\top\mathbf{S}^\top\mathbf{S}\mathbf{r}\|^2\Big]$. It is straightforward to show that $\mathbb{E}\Big[\|\mathbf{U}^\top\mathbf{S}^\top\mathbf{S}\mathbf{r}\|^2] = \mathbf{r}^\top\operatorname{diag}\Big(\frac{\ell_i(\mathbf{X})}{m\pi_i}\Big)\mathbf{r}$. Note that,

$$
\mathbb{E}\Big[\|\mathbf{U}^\top\mathbf{S}^\top\mathbf{S}\mathbf{r}\|^2\Big] = \mathbb{E}_\zeta\Big[\|\mathbf{U}^\top\mathbf{S}^\top\mathbf{S}\mathbf{r}\|^2\Big]\cdot\Pr(\zeta) + \mathbb{E}_{\neg\zeta}\Big[\|\mathbf{U}^\top\mathbf{S}^\top\mathbf{S}\mathbf{r}\|^2\Big]\cdot\Pr(\neg\zeta).
$$

Therefore,

$$
\mathbb{E}_\zeta\Big[\|\mathbf{U}^\top\mathbf{S}^\top\mathbf{S}\mathbf{r}\|^2\Big] \le \frac{1}{\Pr(\zeta)}\cdot\mathbf{r}^\top\operatorname{diag}\Big(\frac{\ell_i(\mathbf{X})}{m\pi_i}\Big)\mathbf{r}.
$$

Combining everything we get,

$$
\begin{aligned}
\mathbb{E}_\zeta[L(\tilde{\boldsymbol{\beta}}) - L(\boldsymbol{\beta}_{\mathrm{OLS}})] &\le (1+3\epsilon)^2\cdot\frac{1}{\Pr(\zeta)}\cdot\mathbf{r}^\top\operatorname{diag}\Big\{\frac{\ell_i(\mathbf{X})}{m\pi_i}\Big\}_{i=1}^n\mathbf{r}\\
&\le (1+15\epsilon)\cdot\frac{1}{\Pr(\zeta)}\cdot\mathbf{r}^\top\operatorname{diag}\Big\{\frac{\ell_i(\mathbf{X})}{m\pi_i}\Big\}_{i=1}^n\mathbf{r}\\
&\le \Big(\Delta + 15\epsilon'\|\mathbf{r}\|^2\Big)\cdot\frac{1}{\Pr(\zeta)},
\end{aligned}
$$

where $\Delta = \mathbf{r}^\top\operatorname{diag}\Big\{\frac{\ell_i(\mathbf{X})}{m\pi_i}\Big\}_{i=1}^n\mathbf{r}$ and $\epsilon' = \sqrt{\frac{3p^3\theta_{\max}^3(\mathbf{X})\log(2p/\delta)}{m^3}}$. This proves Theorem 4.3.

E Proof of Theorem 4.4

In this section, we first provide RMT intuition for the bias of the debiased estimator in Appendix E.1. The corresponding intuition for the variance follows analogously and is omitted for brevity. We then present the detailed proof of Theorem 4.4 in Appendices E.2 and E.3. We next provide the proofs of auxiliary results related to Theorem 4.4 in Appendices E.4 and E.5, with additional lemmas from [45] deferred to Appendix E.6.

E.1 RMT intuition for Theorem 4.4

In this subsection, we provide a heuristic derivation of the bias of the debiased estimator in Theorem 4.4. We begin by recalling some notations from Theorem 4.4. Let $\mathbf{z}_s^\top = \mathbf{e}_{i_s}^\top/\sqrt{\pi_{i_s}}\mathbf{X} \in \mathbb{R}^p$ and $\mathbf{w}_s^\top = \mathbf{e}_{i_s}^\top/\sqrt{\pi_{i_s}}$ so that $\mathbb{E}[\mathbf{z}_s\mathbf{z}_s^\top] = \mathbf{X}^\top\mathbf{X}$ and $\mathbb{E}[\mathbf{z}_s\mathbf{w}_s^\top] = \mathbf{X}^\top$. Denote

$$
\mathbf{Q} = (\mathbf{X}^\top\check{\mathbf{S}}^\top\check{\mathbf{S}}\mathbf{X})^{-1} = \left(\sum_{s=1}^m\frac{1}{m}F_{i_si_s}\mathbf{z}_s\mathbf{z}_s^\top\right)^{-1},\quad \mathbf{Q}_{-s} = \left(\sum_{k\neq s}\frac{1}{m}F_{i_ki_k}\mathbf{z}_k\mathbf{z}_k^\top\right)^{-1},
$$

where $\mathbf{Q}_{-s}$ is independent of $\mathbf{z}_s$, and $F_{i_si_s}$ is a deterministic term whose explicit form will be specified later. Recalling $\operatorname{rank}(\mathbf{X}) = p$, we rewrite the estimator $(\check{\mathbf{S}}\mathbf{X})^\dagger\check{\mathbf{S}} = \mathbf{Q}\mathbf{X}^\top\check{\mathbf{S}}^\top\check{\mathbf{S}}$ and the pseudoinverse $\mathbf{X}^\dagger = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top$. Here, our objective is to choose the diagonal entries F_{ii} such that $\mathbb{E}[\mathbf{Q}\mathbf{X}^\top\check{\mathbf{S}}^\top\check{\mathbf{S}}] \simeq \mathbf{X}^\dagger$. Considering

$$
\sum_{s=1}^m\frac{1}{m}\mathbb{E}[F_{i_si_s}\mathbf{z}_s\mathbf{z}_s^\top] = \sum_{i=1}^n F_{ii}\mathbf{x}_i\mathbf{x}_i^\top,
$$

together with Sherman-Morrison formula in Lemma B.1, we first obtain

$$
\mathbb{E}[\mathbf{Q}\mathbf{X}^\top\check{\mathbf{S}}^\top\check{\mathbf{S}}] = \mathbb{E}\left[\frac{\mathbf{Q}_{-s}F_{i_si_s}\mathbf{z}_s\mathbf{w}_s^\top}{1+F_{i_si_s}\mathbf{z}_s^\top\mathbf{Q}_{-s}\mathbf{z}_s/m}\right] = \sum_{i=1}^n\mathbb{E}\left[\frac{\mathbf{Q}_{-s}F_{ii}\mathbf{X}^\top\mathbf{e}_i\mathbf{e}_i^\top}{1+F_{ii}\mathbf{e}_i^\top\mathbf{X}\mathbf{Q}_{-s}\mathbf{X}^\top\mathbf{e}_i/m\pi_i}\right]
$$

Let $\ell_i(\mathbf{X}) = \mathbf{e}_i^\top \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{e}_i$ be the ith leverage score of $\mathbf{X}$. Applying the rank-one perturbation formula for matrix inversion (see, e.g., [51, Lemma 2.6]), we obtain

$$\mathbb{E}[\mathbf{Q}\mathbf{X}^\top\check{\mathbf{S}}^\top\check{\mathbf{S}}] \simeq \sum_{i=1}^n \mathbb{E}\left[\frac{\mathbf{Q}_{-s}F_{ii}\mathbf{X}^\top\mathbf{e}_i\mathbf{e}_i^\top}{1+F_{ii}\mathbf{e}_i^\top\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{e}_i/m\pi_i}\right] \simeq \mathbb{E}[\mathbf{Q}]\mathbf{X}^\top\sum_{i=1}^n \frac{F_{ii}\mathbf{e}_i\mathbf{e}_i^\top}{1+F_{ii}\ell_i(\mathbf{X})/m\pi_i}.$$

We now choose

$$F_{ii} = \frac{m\pi_i}{m\pi_i - \ell_i(\mathbf{X})} \tag{14}$$

so that

$$\sum_{i=1}^n \frac{F_{ii}\mathbf{e}_i\mathbf{e}_i^\top}{1+F_{ii}\ell_i(\mathbf{X})/m\pi_i} = \mathbf{I}_n.$$

Combined with Proposition 3.2 in [45] and the choice of F_{ii} in (14), which together imply that $\mathbb{E}[\mathbf{Q}] \simeq (\mathbf{X}^\top\mathbf{X})^{-1}$, we get

$$\mathbb{E}[\mathbf{Q}\mathbf{X}^\top\check{\mathbf{S}}^\top\check{\mathbf{S}}] \simeq (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top.$$

This choice leads to the following debiased sampling matrix

$$\check{\mathbf{S}} = \operatorname{diag}\left\{\sqrt{m\pi_{i_s}/(m\pi_{i_s}-\ell_{i_s}(\mathbf{X}))}\right\}_{s=1}^m \cdot \mathbf{S}, \quad i_s \in \{1,\dots,n\}.$$

E.2 Detailed proof of the bias of debiased estimator in Theorem 4.4

For ease of the subsequent analysis, we first rewrite the bias of the debiased estimator:

$$\begin{aligned}
\textbf{Bias}(\check{\boldsymbol{\beta}}) &= \mathbb{E}_\zeta[\check{\boldsymbol{\beta}}]^\top\mathbf{X}^\top\mathbf{X}\mathbb{E}_\zeta[\check{\boldsymbol{\beta}}] - 2\mathbf{y}^\top\mathbf{X}\mathbb{E}_\zeta[\check{\boldsymbol{\beta}}] - \boldsymbol{\beta}_{OLS}^\top\mathbf{X}^\top\mathbf{X}\boldsymbol{\beta}_{OLS} + 2\mathbf{y}^\top\mathbf{X}\boldsymbol{\beta}_{OLS} \\
&= \mathbb{E}_\zeta[\check{\boldsymbol{\beta}}]^\top\mathbf{X}^\top\mathbf{X}\mathbb{E}_\zeta[\check{\boldsymbol{\beta}}] - 2\boldsymbol{\beta}_{OLS}^\top\mathbf{X}^\top\mathbf{X}\mathbb{E}_\zeta[\check{\boldsymbol{\beta}}] - \boldsymbol{\beta}_{OLS}^\top\mathbf{X}^\top\mathbf{X}\boldsymbol{\beta}_{OLS} + 2\boldsymbol{\beta}_{OLS}^\top\mathbf{X}^\top\mathbf{X}\boldsymbol{\beta}_{OLS} \\
&= \mathbb{E}_\zeta[\check{\boldsymbol{\beta}}]^\top\mathbf{X}^\top\mathbf{X}\mathbb{E}_\zeta[\check{\boldsymbol{\beta}}] - 2\boldsymbol{\beta}_{OLS}^\top\mathbf{X}^\top\mathbf{X}\mathbb{E}_\zeta[\check{\boldsymbol{\beta}}] + \boldsymbol{\beta}_{OLS}^\top\mathbf{X}^\top\mathbf{X}\boldsymbol{\beta}_{OLS} \\
&= (\mathbb{E}_\zeta[\check{\boldsymbol{\beta}}] - \boldsymbol{\beta}_{OLS})^\top\mathbf{X}^\top\mathbf{X}(\mathbb{E}_\zeta[\check{\boldsymbol{\beta}}] - \boldsymbol{\beta}_{OLS}) = (\mathbb{E}_\zeta[\check{\boldsymbol{\beta}}] - \boldsymbol{\beta}_{OLS})^\top\mathbf{H}(\mathbb{E}_\zeta[\check{\boldsymbol{\beta}}] - \boldsymbol{\beta}_{OLS}) \\
&= \left\|\mathbf{H}^{\frac{1}{2}}(\mathbb{E}_\zeta[\check{\boldsymbol{\beta}}] - \boldsymbol{\beta}_{OLS})\right\|^2,
\end{aligned} \tag{15}$$

with $\mathbf{H} = \mathbf{X}^\top\mathbf{X}$. This reformulation shows that the proof of the bias bound for the debiased estimator proceeds in two main steps:

1. construct a high probability event ζ, based on subspace-embedding-type results in Lemma B.3, under which the inverse $(\mathbf{X}^\top\check{\mathbf{S}}^\top\check{\mathbf{S}}\mathbf{X})^{-1}$ is well conditioned; and
2. conditioned on that event ζ, bound the quantity $\|\mathbf{H}^{1/2}(\mathbb{E}_\zeta[\check{\boldsymbol{\beta}}] - \boldsymbol{\beta}_{OLS})\|^2$ using a "leave-one-out" type analysis.

We begin by constructing a high-probability event ζ. Without loss of generality, assume that $t = m/4$ is an integer. We partition the index set $\{1,\dots,m\}$ into four blocks of size t, and define the events:

$$\zeta_j : \sum_{s=t(j-1)+1}^{tj} \frac{1}{t}\mathbf{z}_s\mathbf{z}_s^\top \succeq \frac{1}{2}\mathbf{X}^\top\mathbf{X}, \quad j=1,2,3,4, \quad \zeta = \bigcap_{j=1}^4 \zeta_j. \tag{16}$$

By $F_{i_s i_s} \geq 1$, each event ζ_j further implies

$$\sum_{s=t(j-1)+1}^{tj} \frac{1}{t}F_{i_s i_s}\mathbf{z}_s\mathbf{z}_s^\top \succeq \sum_{s=t(j-1)+1}^{tj} \frac{1}{t}\mathbf{z}_s\mathbf{z}_s^\top \succeq \frac{1}{2}\mathbf{X}^\top\mathbf{X}, \quad j=1,2,3,4.$$

Intuitively, the event ζ_j ensures that the (weighted) average of the rank-one matrices $\mathbf{z}_s\mathbf{z}_s^\top$ over the j-th block forms a sketch of size t that provides a "lower" $1/2$-spectral-approximation of $\mathbf{X}^\top\mathbf{X}$, in the sense of Definition 2.4.

Under random sampling in Definition 2.1, the events $\zeta_1, \zeta_2, \zeta_3, \zeta_4$ are mutually independent. Consequently, for any index $s \in \{1,\dots,m\}$, there exists a block index $j = j(s) \in \{1,2,3,4\}$ such that

1. the event ζ_j is independent of $\mathbf{z}_s$; and
2. conditioning on ζ_j, it follows that $\mathbf{Q} \preceq \mathbf{Q}_{-s} \preceq 8\mathbf{H}^{-1}$.

Denote $\check{\mathbf{z}}_s^\top = \mathbf{e}_{i_s}^\top/\sqrt{\pi_{i_s}}\mathbf{X_H} \in \mathbb{R}^p$ with $\mathbf{X_H} = \mathbf{X}\mathbf{H}^{-1/2}$, let

$$\check{\mathbf{Q}} = \mathbf{H}^{\frac{1}{2}}\mathbf{Q}\mathbf{H}^{\frac{1}{2}} = (\mathbf{X}_\mathbf{H}^\top\check{\mathbf{S}}^\top\check{\mathbf{S}}\mathbf{X}_\mathbf{H})^{-1} = \left(\sum_{s=1}^m \frac{1}{m}F_{i_si_s}\check{\mathbf{z}}_s\check{\mathbf{z}}_s^\top\right)^{-1},$$

and

$$\check{\mathbf{Q}}_{-s} = \mathbf{H}^{\frac{1}{2}}\mathbf{Q}_{-s}\mathbf{H}^{\frac{1}{2}} = \left(\sum_{k\neq s} \frac{1}{m}F_{i_k,i_k}\check{\mathbf{z}}_k\check{\mathbf{z}}_k^\top\right)^{-1}.$$

Letting $\mathbf{r} = \mathbf{y} - \mathbf{X}\boldsymbol{\beta}_{OLS}$ and $\check{a}_s = \mathbf{e}_{i_s}^\top/\sqrt{\pi_{i_s}}\mathbf{r}$, we now rewrite

$$\mathbb{E}_\zeta[\check{\boldsymbol{\beta}}] - \boldsymbol{\beta}_{OLS} = \mathbb{E}_\zeta[(\mathbf{X}^\top\check{\mathbf{S}}^\top\check{\mathbf{S}}\mathbf{X})^{-1}\mathbf{X}^\top\check{\mathbf{S}}^\top\check{\mathbf{S}}\mathbf{r}], \tag{17}$$

which gives

$$\begin{aligned}
\mathbf{H}^{\frac{1}{2}}(\mathbb{E}_\zeta[\check{\boldsymbol{\beta}}] - \boldsymbol{\beta}_{OLS}) &= \mathbf{H}^{\frac{1}{2}}\mathbb{E}_\zeta[(\mathbf{X}^\top\check{\mathbf{S}}^\top\check{\mathbf{S}}\mathbf{X})^{-1}\mathbf{X}^\top\check{\mathbf{S}}^\top\check{\mathbf{S}}\mathbf{r}] = \mathbb{E}_\zeta[(\mathbf{X}_\mathbf{H}^\top\check{\mathbf{S}}^\top\check{\mathbf{S}}\mathbf{X}_\mathbf{H})^{-1}\mathbf{X}_\mathbf{H}^\top\check{\mathbf{S}}^\top\check{\mathbf{S}}\mathbf{r}] \\
&= \mathbb{E}_\zeta[\check{\mathbf{Q}}\mathbf{X}_\mathbf{H}^\top\check{\mathbf{S}}^\top\check{\mathbf{S}}\mathbf{r}] = \frac{1}{m}\sum_{s=1}^m \mathbb{E}_\zeta[\check{\mathbf{Q}}F_{i_si_s}\check{\mathbf{z}}_s\check{a}_s] = \mathbb{E}_\zeta\left[\frac{\check{\mathbf{Q}}_{-s}F_{i_si_s}\check{\mathbf{z}}_s\check{a}_s}{1+\frac{1}{m}F_{i_si_s}\check{\mathbf{z}}_s^\top\check{\mathbf{Q}}_{-s}\check{\mathbf{z}}_s}\right] \\
&= \underbrace{\mathbb{E}_\zeta\left[\check{\mathbf{Q}}_{-s}\check{\mathbf{z}}_s\check{a}_s\left(\frac{F_{i_si_s}}{1+\frac{1}{m}F_{i_si_s}\check{\mathbf{z}}_s^\top\check{\mathbf{Q}}_{-s}\check{\mathbf{z}}_s} - 1\right)\right]}_{\mathbf{T}_1} + \underbrace{\mathbb{E}_\zeta[\check{\mathbf{Q}}_{-s}\check{\mathbf{z}}_s\check{a}_s]}_{\mathbf{T}_2}.
\end{aligned}$$

This decomposition yields

$$\|\mathbf{H}^{\frac{1}{2}}(\mathbb{E}_\zeta[\check{\boldsymbol{\beta}}] - \boldsymbol{\beta}_{OLS})\|^2 \leq \|\mathbf{T}_1\|^2 + \|\mathbf{T}_2\|^2 + 2\|\mathbf{T}_1\|\|\mathbf{T}_2\|. \tag{18}$$

Consequently, to establish the bias bound for the debiased estimator in Theorem 4.4, it suffices to derive upper bounds on the norms $\|\mathbf{T}_1\|$ and $\|\mathbf{T}_2\|$ in (18). We first consider to bound the term $\|\mathbf{T}_1\|$. Without loss of generality, assume that the events ζ_1, ζ_2, ζ_3 are independent of $\mathbf{z}_s$. Define $\zeta^{'} = \bigcap_{j=1}^3 \zeta_j$ and $\delta_4 = \Pr(\neg\zeta_4)$. We first consider the first term $\|\mathbf{T}_1\|$. Letting $\check{\gamma}_{i_s} = 1 + \frac{1}{m}F_{i_si_s}\check{\mathbf{z}}_s^\top\check{\mathbf{Q}}_{-s}\check{\mathbf{z}}_s$ and $\bar{\gamma}_i = 1 + \frac{1}{m\pi_i}F_{ii}\mathbf{x}_{\mathbf{H}_i}^\top\check{\mathbf{Q}}_{-s}\mathbf{x}_{\mathbf{H}_i}$ with $\mathbf{x}_{\mathbf{H}_i}^\top$ the ith row of $\mathbf{X}_\mathbf{H}$, along with the fact that the event $\zeta^{'}$ indicates $\check{\mathbf{Q}}_{-s} = \mathbf{H}^{1/2}\mathbf{Q}_{-s}\mathbf{H}^{1/2} \preceq 8\mathbf{I}_p$ and $\mathbf{X}_\mathbf{H}^\top\mathbf{X}_\mathbf{H} = \mathbf{I}_p$, we obtain

$$\begin{aligned}
\|\mathbf{T}_1\| &= \left\|\mathbb{E}_\zeta\left[\check{\mathbf{Q}}_{-s}\check{\mathbf{z}}_s\check{a}_s\left(\frac{F_{i_si_s}}{\check{\gamma}_{i_s}} - 1\right)\right]\right\| = \max_{\|\mathbf{u}\|=1}\mathbb{E}_\zeta\left[\mathbf{u}^\top\check{\mathbf{Q}}_{-s}\check{\mathbf{z}}_s\check{a}_s\left(\frac{F_{i_si_s}}{\check{\gamma}_{i_s}} - 1\right)\right] \\
&\overset{(a)}{\leq} \max_{\|\mathbf{u}\|=1}\sqrt{\mathbb{E}_\zeta\left[\mathbf{u}^\top\check{\mathbf{Q}}_{-s}\check{\mathbf{z}}_s\check{\mathbf{z}}_s^\top\check{\mathbf{Q}}_{-s}\mathbf{u}\left|\frac{F_{i_si_s}}{\check{\gamma}_{i_s}} - 1\right|\right]\mathbb{E}_\zeta\left[\check{a}_s^2\left|\frac{F_{i_si_s}}{\check{\gamma}_{i_s}} - 1\right|\right]} \\
&\overset{(b)}{\leq} \max_{\|\mathbf{u}\|=1} 2\sqrt{\mathbb{E}_{\zeta'}\left[\mathbf{u}^\top\check{\mathbf{Q}}_{-s}\check{\mathbf{z}}_s\check{\mathbf{z}}_s^\top\check{\mathbf{Q}}_{-s}\mathbf{u}\left|\frac{F_{i_si_s}}{\check{\gamma}_{i_s}} - 1\right|\right]\mathbb{E}_{\zeta'}\left[\check{a}_s^2\left|\frac{F_{i_si_s}}{\check{\gamma}_{i_s}} - 1\right|\right]} \\
&\leq \max_{\|\mathbf{u}\|=1} 2\|\mathbf{r}\|\sqrt{\mathbb{E}_{\zeta'}\left[\max_{1\leq i\leq n}\left|\frac{F_{ii}}{\bar{\gamma}_i} - 1\right|\mathbf{u}^\top\check{\mathbf{Q}}_{-s}\mathbf{X}_\mathbf{H}^\top\mathbf{X}_\mathbf{H}\check{\mathbf{Q}}_{-s}\mathbf{u}\right]\mathbb{E}_{\zeta'}\left[\max_{1\leq i\leq n}\left|\frac{F_{ii}}{\bar{\gamma}_i} - 1\right|\right]} \\
&\leq 16\|\mathbf{r}\|\mathbb{E}_{\zeta'}\left[\max_{1\leq i\leq n}\left|\frac{F_{ii}}{\bar{\gamma}_i} - 1\right|\right],
\end{aligned}$$

where in (a) we apply Cauchy-Schwarz inequality, in (b) we use Lemma E.2. Recalling the result (28) in Lemma E.1, we then get

$$\|\mathbf{T}_1\| = O\left(\frac{\log n}{\log\log n}\sqrt{\frac{\theta_{\max}^3(\mathbf{X})p^3}{m^3}}\right)\cdot\|\mathbf{r}\|.$$

We next explore the other term $\|\mathbf{T}_2\|$. Using the fact $\mathbf{X}_{\mathbf{H}}^\top \mathbf{r} = 0$, we get

$$\begin{aligned}
\|\mathbf{T}_2\| &= \|\mathbb{E}_\zeta[\check{\mathbf{Q}}_{-s}\check{\mathbf{z}}_s\check{a}_s]\| = \frac{1}{1-\delta_4}(\|\mathbb{E}_{\zeta'}[\check{\mathbf{Q}}_{-s}\check{\mathbf{z}}_s\check{a}_s] - \mathbb{E}_{\zeta'}[\check{\mathbf{Q}}_{-s}\check{\mathbf{z}}_s\check{a}_s \cdot \mathbf{1}_{\neg\zeta_4}]\|) \\
&= \frac{1}{1-\delta_4}(\|\mathbb{E}_{\zeta'}[\check{\mathbf{Q}}_{-s}\mathbf{X}_{\mathbf{H}}^\top \mathbf{r}] - \mathbb{E}_{\zeta'}[\check{\mathbf{Q}}_{-s}\check{\mathbf{z}}_s\check{a}_s \cdot \mathbf{1}_{\neg\zeta_4}]\|) = -\frac{1}{1-\delta_4}\|\mathbb{E}_{\zeta'}[\check{\mathbf{Q}}_{-s}\check{\mathbf{z}}_s\check{a}_s \cdot \mathbf{1}_{\neg\zeta_4}]\| \\
&\le 2\|\mathbb{E}_{\zeta'}[\check{\mathbf{Q}}_{-s}\check{\mathbf{z}}_s\check{a}_s \cdot \mathbf{1}_{\neg\zeta_4}]\| = \max_{\|\mathbf{u}\|=1} 2\mathbb{E}_{\zeta'}[\mathbf{u}^\top\check{\mathbf{Q}}_{-s}\check{\mathbf{z}}_s\check{a}_s \cdot \mathbf{1}_{\neg\zeta_4}] \\
&\le \max_{\|\mathbf{u}\|=1} 2\sqrt{\mathbb{E}_{\zeta'}[\mathbf{u}^\top\check{\mathbf{Q}}_{-s}\check{\mathbf{z}}_s\check{\mathbf{z}}_s^\top\check{\mathbf{Q}}_{-s}\mathbf{u} \cdot \mathbf{1}_{\neg\zeta_4}]\mathbb{E}_{\zeta'}[\check{a}_s^2]} \\
&\le \max_{\|\mathbf{u}\|=1} 2\|\mathbf{r}\|\sqrt{\mathbb{E}_{\zeta'}[\mathbf{u}^\top\check{\mathbf{Q}}_{-s}\check{\mathbf{z}}_s\check{\mathbf{z}}_s^\top\check{\mathbf{Q}}_{-s}\mathbf{u} \cdot \mathbf{1}_{\neg\zeta_4}]}.
\end{aligned}$$

Taking $\check{\mathbf{Q}}_{-s} = \mathbf{H}^{1/2}\mathbf{Q}_{-s}\mathbf{H}^{1/2} \preceq 8\mathbf{I}_p$ again, we derive

$$\mathbb{E}_{\zeta'}[\mathbf{u}^\top\check{\mathbf{Q}}_{-s}\check{\mathbf{z}}_s\check{\mathbf{z}}_s^\top\check{\mathbf{Q}}_{-s}\mathbf{u}] = \sum_{i=1}^n \mathbb{E}_{\zeta'}[\mathbf{u}^\top\check{\mathbf{Q}}_{-s}\mathbf{x}_{\mathbf{H}_i}\mathbf{x}_{\mathbf{H}_i}^\top\check{\mathbf{Q}}_{-s}\mathbf{u}] = \mathbb{E}_{\zeta'}[\mathbf{u}^\top\check{\mathbf{Q}}_{-s}\mathbf{X}_{\mathbf{H}}^\top\mathbf{X}_{\mathbf{H}}\check{\mathbf{Q}}_{-s}\mathbf{u}] \le 64,$$

and

$$\begin{aligned}
\mathrm{Var}_{\zeta'}[\mathbf{u}^\top\check{\mathbf{Q}}_{-s}\check{\mathbf{z}}_s\check{\mathbf{z}}_s^\top\check{\mathbf{Q}}_{-s}\mathbf{u}] &\le \mathbb{E}_{\zeta'}[(\mathbf{u}^\top\check{\mathbf{Q}}_{-s}\check{\mathbf{z}}_s\check{\mathbf{z}}_s^\top\check{\mathbf{Q}}_{-s}\mathbf{u})^2] = \sum_{i=1}^n \frac{1}{\pi_i}\mathbb{E}_{\zeta'}[(\mathbf{u}^\top\check{\mathbf{Q}}_{-s}\mathbf{x}_{\mathbf{H}_i}\mathbf{x}_{\mathbf{H}_i}^\top\check{\mathbf{Q}}_{-s}\mathbf{u})^2] \\
&\le 8^4 \max_{1\le i\le n}\frac{\ell_i(\mathbf{X})}{\pi_i} \le 8^4\theta_{\max}(\mathbf{X})p.
\end{aligned}$$

Applying the Chebyshev's inequality, we obtain, for $z > 64$,

$$\Pr(\mathbf{u}^\top\check{\mathbf{Q}}_{-s}\check{\mathbf{z}}_s\check{\mathbf{z}}_s^\top\check{\mathbf{Q}}_{-s}\mathbf{u} \ge z \mid \zeta^{'}) \le \frac{8^4\theta_{\max}(\mathbf{X})p}{z^2}.$$

This together with $\delta_4 \le n^{-(3+\alpha)}$ further leads to

$$\begin{aligned}
&\mathbb{E}_{\zeta'}[\mathbf{u}^\top\check{\mathbf{Q}}_{-s}\check{\mathbf{z}}_s\check{\mathbf{z}}_s^\top\check{\mathbf{Q}}_{-s}\mathbf{u} \cdot \mathbf{1}_{\neg\zeta_4}] = \int_0^\infty \Pr(\mathbf{u}^\top\check{\mathbf{Q}}_{-s}\check{\mathbf{z}}_s\check{\mathbf{z}}_s^\top\check{\mathbf{Q}}_{-s}\mathbf{u} \cdot \mathbf{1}_{\neg\zeta_4} \ge z \mid \zeta^{'})dz \\
&= \int_0^{2m^2} \Pr(\mathbf{u}^\top\check{\mathbf{Q}}_{-s}\check{\mathbf{z}}_s\check{\mathbf{z}}_s^\top\check{\mathbf{Q}}_{-s}\mathbf{u} \cdot \mathbf{1}_{\neg\zeta_4} \ge z \mid \zeta^{'})dz + \int_{2m^2}^\infty \Pr(\mathbf{u}^\top\check{\mathbf{Q}}_{-s}\check{\mathbf{z}}_s\check{\mathbf{z}}_s^\top\check{\mathbf{Q}}_{-s}\mathbf{u} \cdot \mathbf{1}_{\neg\zeta_4} \ge z \mid \zeta^{'})dz \\
&\le 2m^2\delta_4 + \int_{2m^2}^\infty \Pr(\mathbf{u}^\top\check{\mathbf{Q}}_{-s}\check{\mathbf{z}}_s\check{\mathbf{z}}_s^\top\check{\mathbf{Q}}_{-s}\mathbf{u} \ge z \mid \zeta^{'})dz \\
&\le \frac{2m^2}{n^{3+\alpha}} + \frac{8^4\theta_{\max}(\mathbf{X})p}{2m^2} = O\left(\frac{m^2}{n^{3+\alpha}} + \frac{\theta_{\max}(\mathbf{X})p}{m^2}\right).
\end{aligned}$$

We thus get

$$\|\mathbf{T}_2\| = O\left(\frac{m}{n^{(3+\alpha)/2}} + \frac{\sqrt{\theta_{\max}(\mathbf{X})p}}{m}\right)\cdot\|\mathbf{r}\|.$$

E.3 Detailed proof of the variance of debiased estimator in Theorem 4.4

A heuristic derivation of the variance of the debiased estimator proceeds analogously to that of the bias, and is omitted here for clarity.

Building on the proof strategy for the bias of the debiased estimator, the analysis of its variance proceeds in two main steps:

1. construct a high probability event ζ as in (16); and
2. conditioning on that event ζ, bound the quantity $\mathbf{Var}(\check{\boldsymbol{\beta}}) - \Delta(\mathbf{X})$ using a "leave-one/two-out" type analysis.

To complete the analysis of the variance of the debiased estimator, we begin by rewriting

$$\begin{aligned}
&\mathbf{Var}(\check{\boldsymbol{\beta}}) - \Delta(\mathbf{X}) = \mathbb{E}_\zeta[(\check{\boldsymbol{\beta}} - \boldsymbol{\beta}_{OLS})^\top \mathbf{X}^\top \mathbf{X}(\check{\boldsymbol{\beta}} - \boldsymbol{\beta}_{OLS})] - \Delta(\mathbf{X}) \\
&= \mathbb{E}_\zeta[\mathbf{r}^\top \check{\mathbf{S}}^\top \check{\mathbf{S}} \mathbf{X}_{\mathbf{H}} \check{\mathbf{Q}}^2 \mathbf{X}_{\mathbf{H}}^\top \check{\mathbf{S}}^\top \check{\mathbf{S}} \mathbf{r}] - \Delta(\mathbf{X}) = \frac{1}{m^2} \sum_{s=1}^m \sum_{k=1}^m \mathbb{E}_\zeta[\check{\mathbf{z}}_k^\top \check{\mathbf{Q}}^2 \check{\mathbf{z}}_s F_{i_s i_s} F_{i_k i_k} \check{a}_s \check{a}_k] - \Delta(\mathbf{X}) \\
&= \frac{1}{m^2} \sum_{s=k=1}^m \mathbb{E}_\zeta[\check{\mathbf{z}}_s^\top \check{\mathbf{Q}}^2 \check{\mathbf{z}}_s F_{i_s i_s}^2 \check{a}_s^2] + \frac{1}{m^2} \sum_{s \neq k}^m \mathbb{E}_\zeta[\check{\mathbf{z}}_k^\top \check{\mathbf{Q}}^2 \check{\mathbf{z}}_s F_{i_s i_s} F_{i_k i_k} \check{a}_s \check{a}_k] - \Delta(\mathbf{X}).
\end{aligned}$$

Recalling

$$\Delta(\mathbf{X}) = \mathbf{r}^\top \operatorname{diag}\left\{ \frac{\ell_i(\mathbf{X})}{m\pi_i} \right\}_{i=1}^n \mathbf{r},$$

we then consider

$$\begin{aligned}
&\frac{1}{m^2} \sum_{s=k=1}^m \mathbb{E}_\zeta[\check{\mathbf{z}}_s^\top \check{\mathbf{Q}}^2 \check{\mathbf{z}}_s F_{i_s i_s}^2 \check{a}_s^2] - \Delta(\mathbf{X}) = \sum_{s=k=1}^m \mathbb{E}_\zeta \left[\frac{\check{\mathbf{z}}_s^\top \check{\mathbf{Q}}_{-s}^2 \check{\mathbf{z}}_s F_{i_s i_s}^2 \check{a}_s^2}{m^2 \check{\gamma}_{i_s}^2} \right] - \Delta(\mathbf{X}) \\
&= \mathbb{E}_\zeta \left[\frac{\check{\mathbf{z}}_s^\top \check{\mathbf{Q}}_{-s}^2 \check{\mathbf{z}}_s F_{i_s i_s}^2 \check{a}_s^2}{m \check{\gamma}_{i_s}^2} \right] - \Delta(\mathbf{X}) \\
&= \underbrace{\frac{1}{m} \mathbb{E}_\zeta \left[\check{\mathbf{z}}_s^\top \check{\mathbf{Q}}_{-s}^2 \check{\mathbf{z}}_s \check{a}_s^2 \left(\frac{F_{i_s i_s}^2}{\check{\gamma}_{i_s}^2} - 1 \right) \right]}_{W_1} + \underbrace{\mathbb{E}_\zeta \left[\frac{1}{m} \check{\mathbf{z}}_s^\top \check{\mathbf{Q}}_{-s}^2 \check{\mathbf{z}}_s \check{a}_s^2 - \mathbf{r}^\top \operatorname{diag}\left\{ \frac{\mathbf{x}_{\mathbf{H}_i}^\top \check{\mathbf{Q}}_{-s}^2 \mathbf{x}_{\mathbf{H}_i}}{m\pi_i} \right\}_{i=1}^n \mathbf{r} \right]}_{W_2} \\
&+ \underbrace{\mathbf{r}^\top \operatorname{diag}\left\{ \frac{\mathbf{x}_{\mathbf{H}_i}^\top \mathbb{E}_\zeta[\check{\mathbf{Q}}_{-s}^2 - \check{\mathbf{Q}}^2] \mathbf{x}_{\mathbf{H}_i}}{m\pi_i} \right\}_{i=1}^n \mathbf{r}}_{W_3} + \underbrace{\mathbf{r}^\top \operatorname{diag}\left\{ \frac{\mathbf{x}_{\mathbf{H}_i}^\top \mathbb{E}_\zeta[\check{\mathbf{Q}}^2] \mathbf{x}_{\mathbf{H}_i} - \ell_i(\mathbf{X}) - \check{\Delta}_{\check{\mathbf{Q}}^2, i}}{m\pi_i} \right\}_{i=1}^n \mathbf{r}}_{W_4} \\
&+ \underbrace{\mathbf{r}^\top \operatorname{diag}\left\{ \frac{\check{\Delta}_{\check{\mathbf{Q}}^2, i}}{m\pi_i} \right\}_{i=1}^n \mathbf{r}}_{W_5},
\end{aligned}$$

where $\check{\gamma}_{i_s} = 1 + F_{i_s i_s} \check{\mathbf{z}}_s^\top \check{\mathbf{Q}}_{-s} \check{\mathbf{z}}_s / m$, $\check{\Delta}_{\check{\mathbf{Q}}^2, i} = \mathbf{x}_{\mathbf{H}_i}^\top \mathbf{X}_{\mathbf{H}}^\top \bar{\mathbf{F}} \mathbf{X}_{\mathbf{H}} \mathbf{x}_{\mathbf{H}_i}$, with diagonal matrix $\bar{\mathbf{F}} = \operatorname{diag}\{\bar{F}_{ii}\}_{i=1}^n$ and diagonal entries

$$\bar{F}_{ii} = \frac{\mathbf{x}_{\mathbf{H}_i}^\top \mathbb{E}_\zeta[\check{\mathbf{Q}}^2] \mathbf{x}_{\mathbf{H}_i}}{m\pi_i}.$$

In the following, we bound the term W_1. Considering $\mathbf{F}_{i_s i_s} < 2$ under the condition $m \geq 2\theta_{\max}(\mathbf{X})p$, together with Lemma E.2, we get

$$\begin{aligned}
W_1 &\leq \frac{1}{m} \mathbb{E}_\zeta \left[\check{\mathbf{z}}_s^\top \check{\mathbf{Q}}_{-s}^2 \check{\mathbf{z}}_s \check{a}_s^2 \left| \frac{F_{i_s i_s}^2}{\check{\gamma}_{i_s}^2} - 1 \right| \right] \leq \frac{2}{m} \mathbb{E}_{\zeta'} \left[\check{\mathbf{z}}_s^\top \check{\mathbf{Q}}_{-s}^2 \check{\mathbf{z}}_s \check{a}_s^2 \left| \frac{F_{i_s i_s}^2}{\check{\gamma}_{i_s}^2} - 1 \right| \right] \\
&\leq \frac{2}{m} \mathbb{E}_{\zeta'} \left[\check{\mathbf{z}}_s^\top \check{\mathbf{Q}}_{-s}^2 \check{\mathbf{z}}_s \check{a}_s^2 \frac{F_{i_s i_s}}{\check{\gamma}_{i_s}} \left| \frac{F_{i_s i_s}}{\check{\gamma}_{i_s}} - 1 \right| \right] + \frac{2}{m} \mathbb{E}_{\zeta'} \left[\check{\mathbf{z}}_s^\top \check{\mathbf{Q}}_{-s}^2 \check{\mathbf{z}}_s \check{a}_s^2 \left| \frac{F_{i_s i_s}}{\check{\gamma}_{i_s}} - 1 \right| \right] \\
&\leq \frac{4}{m} \mathbb{E}_{\zeta'} \left[\check{\mathbf{z}}_s^\top \check{\mathbf{Q}}_{-s}^2 \check{\mathbf{z}}_s \check{a}_s^2 \left| \frac{F_{i_s i_s}}{\check{\gamma}_{i_s}} - 1 \right| \right] + \frac{2}{m} \mathbb{E}_{\zeta'} \left[\check{\mathbf{z}}_s^\top \check{\mathbf{Q}}_{-s}^2 \check{\mathbf{z}}_s \check{a}_s^2 \left| \frac{F_{i_s i_s}}{\check{\gamma}_{i_s}} - 1 \right| \right] \\
&= \frac{6}{m} \mathbb{E}_{\zeta'} \left[\check{\mathbf{z}}_s^\top \check{\mathbf{Q}}_{-s}^2 \check{\mathbf{z}}_s \check{a}_s^2 \left| \frac{F_{i_s i_s}}{\check{\gamma}_{i_s}} - 1 \right| \right] = \sum_{i=1}^n \frac{6}{m\pi_i} \mathbb{E}_{\zeta'} \left[\mathbf{x}_{\mathbf{H}_i}^\top \check{\mathbf{Q}}_{-s}^2 \mathbf{x}_{\mathbf{H}_i} r_i^2 \left| \frac{F_{ii}}{\bar{\gamma}_i} - 1 \right| \right],
\end{aligned}$$

which combined with $\check{\mathbf{Q}}_{-s} \preceq 8\mathbf{I}_p$, $\mathbf{x}_{\mathbf{H}_i}^\top \check{\mathbf{Q}}_{-s} \mathbf{x}_{\mathbf{H}_i} / (m\pi_i) \leq 8\theta_{\max}(\mathbf{X})p/m$, and Lemma E.1 gives

$$W_1 \leq \sum_{i=1}^n \frac{384\ell_i(\mathbf{X})}{m\pi_i} \mathbb{E}_{\zeta'} \left[r_i^2 \left| \frac{F_{ii}}{\bar{\gamma}_i} - 1 \right| \right] \leq \frac{384\theta_{\max}(\mathbf{X})p}{m} \mathbb{E}_{\zeta'} \left[\max_{1 \leq i \leq n} \left| \frac{F_{ii}}{\bar{\gamma}_i} - 1 \right| \sum_{i=1}^n r_i^2 \right]$$

$$\leq \frac{384\theta_{\max}(\mathbf{X})p\|\mathbf{r}\|^2}{m}\mathbb{E}_{\zeta'}\left[\max_{1\leq i\leq n}\left|\frac{F_{ii}}{\bar{\gamma}_i}-1\right|\right] = O\left(\frac{\log n}{\log\log n}\sqrt{\frac{\theta^3_{\max}(\mathbf{X})p^3}{m^3}}\right)\cdot\|\mathbf{r}\|^2.$$

Next, we explore the term W_2,

$$\begin{aligned}
W_2 &= \frac{1}{1-\delta_4}\mathbb{E}_{\zeta'}\left[\frac{1}{m}\check{\mathbf{z}}_s^\top\check{\mathbf{Q}}^2_{-s}\check{\mathbf{z}}_s\check{a}_s^2 - \mathbf{r}^\top\operatorname{diag}\left\{\frac{\mathbf{x}_{\mathbf{H}_i}^\top\check{\mathbf{Q}}^2_{-s}\mathbf{x}_{\mathbf{H}_i}}{m\pi_i}\right\}_{i=1}^n\mathbf{r}\right]\\
&\quad - \frac{1}{1-\delta_4}\mathbb{E}_{\zeta'}\left[\left(\frac{1}{m}\check{\mathbf{z}}_s^\top\check{\mathbf{Q}}^2_{-s}\check{\mathbf{z}}_s\check{a}_s^2 - \mathbf{r}^\top\operatorname{diag}\left\{\frac{\mathbf{x}_{\mathbf{H}_i}^\top\check{\mathbf{Q}}^2_{-s}\mathbf{x}_{\mathbf{H}_i}}{m\pi_i}\right\}_{i=1}^n\mathbf{r}\right)\cdot\mathbf{1}_{\neg\zeta_4}\right]\\
&\leq 2\mathbb{E}_{\zeta'}\left[\frac{1}{m}\check{\mathbf{z}}_s^\top\check{\mathbf{Q}}^2_{-s}\check{\mathbf{z}}_s\check{a}_s^2\cdot\mathbf{1}_{\neg\zeta_4}\right] + 2\mathbb{E}_{\zeta'}\left[\mathbf{r}^\top\operatorname{diag}\left\{\frac{\mathbf{x}_{\mathbf{H}_i}^\top\check{\mathbf{Q}}^2_{-s}\mathbf{x}_{\mathbf{H}_i}}{m\pi_i}\right\}_{i=1}^n\mathbf{r}\cdot\mathbf{1}_{\neg\zeta_4}\right]\\
&\leq 2\mathbb{E}_{\zeta'}\left[\frac{1}{m}\check{\mathbf{z}}_s^\top\check{\mathbf{Q}}^2_{-s}\check{\mathbf{z}}_s\check{a}_s^2\cdot\mathbf{1}_{\neg\zeta_4}\right] + \frac{128\theta_{\max}(\mathbf{X})p\|\mathbf{r}\|^2\delta_4}{m}\\
&= 2\mathbb{E}_{\zeta'}\left[\frac{1}{m}\check{\mathbf{z}}_s^\top\check{\mathbf{Q}}^2_{-s}\check{\mathbf{z}}_s\check{a}_s^2\cdot\mathbf{1}_{\neg\zeta_4}\right] + O\left(\frac{\theta_{\max}(\mathbf{X})p}{m^2}\right)\|\mathbf{r}\|^2.
\end{aligned}$$

Using $\check{\mathbf{Q}}_{-s}\preceq 8\mathbf{I}_p$, we obtain

$$\mathbb{E}_{\zeta'}\left[\frac{1}{m}\check{\mathbf{z}}_s^\top\check{\mathbf{Q}}^2_{-s}\check{\mathbf{z}}_s\check{a}_s^2\right] = \sum_{i=1}^n\frac{1}{m\pi_i}\mathbb{E}_{\zeta'}\left[\mathbf{x}_{\mathbf{H}_i}^\top\check{\mathbf{Q}}^2_{-s}\mathbf{x}_{\mathbf{H}_i}r_i^2\right] \leq \frac{64\theta_{\max}(\mathbf{X})p}{m}\|\mathbf{r}\|^2,$$

and

$$\begin{aligned}
\operatorname{Var}_{\zeta'}\left[\frac{1}{m}\check{\mathbf{z}}_s^\top\check{\mathbf{Q}}^2_{-s}\check{\mathbf{z}}_s\check{a}_s^2\right] &\leq \frac{1}{m^2}\mathbb{E}_{\zeta'}\left[(\check{\mathbf{z}}_s^\top\check{\mathbf{Q}}^2_{-s}\check{\mathbf{z}}_s\check{a}_s^2)^2\right] \leq \sum_{i=1}^n\frac{1}{m^2\pi_i^3}\mathbb{E}_{\zeta'}\left[(\mathbf{x}_{\mathbf{H}_i}^\top\check{\mathbf{Q}}^2_{-s}\mathbf{x}_{\mathbf{H}_i}r_i^2)^2\right]\\
&\leq \frac{8^4\theta^3_{\max}(\mathbf{X})p^3}{m^2}\sum_{i=1}^n\frac{r_i^4}{\ell_i} \leq \frac{8^4\theta^3_{\max}(\mathbf{X})p^3}{m^2\ell_{\min}(\mathbf{X})}\|\mathbf{r}\|^4,
\end{aligned}$$

where we denote $\ell_{\min}(\mathbf{X}) = \min_{1\leq i\leq n,\ell_i(\mathbf{X})\neq 0}\ell_i(\mathbf{X})$. Following the Chebyshev's inequality results in, for $z > 128\theta_{\max}(\mathbf{X})p\|\mathbf{r}\|^2/m$, we get

$$\Pr\left(\frac{1}{m}\check{\mathbf{z}}_s^\top\check{\mathbf{Q}}^2_{-s}\check{\mathbf{z}}_s\check{a}_s^2 \geq z \mid \zeta'\right) \leq \frac{8^4\theta^3_{\max}(\mathbf{X})p^3}{m^2\ell_{\min}(\mathbf{X})z^2}\|\mathbf{r}\|^4.$$

Considering

$$\ell_{\min}(\mathbf{X}) = \min_{i,\|\mathbf{x}_i\|>0}\mathbf{x}_i^\top(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{x}_i \geq \min_{i,\|\mathbf{x}_i\|>0}\frac{\|\mathbf{x}_i\|^2}{\|\mathbf{X}\|^2} \geq \min_{i,\|\mathbf{x}_i\|>0}\frac{\|\mathbf{x}_i\|^2}{\|\mathbf{X}\|_F^2} \geq \frac{1}{n^\alpha},$$

for some constant $\alpha > 0$, together with $\delta_4 \leq n^{-2-\alpha}$, we further derive

$$\begin{aligned}
\mathbb{E}_{\zeta'}\left[\frac{1}{m}\check{\mathbf{z}}_s^\top\check{\mathbf{Q}}^2_{-s}\check{\mathbf{z}}_s\check{a}_s^2\cdot\mathbf{1}_{\neg\zeta_4}\right] &= \int_0^\infty\Pr\left(\frac{1}{m}\check{\mathbf{z}}_s^\top\check{\mathbf{Q}}^2_{-s}\check{\mathbf{z}}_s\check{a}_s^2\cdot\mathbf{1}_{\neg\zeta_4}\geq z\mid\zeta'\right)dz\\
&= \int_0^{2m\|\mathbf{r}\|^2/\ell_{\min}(\mathbf{X})}\Pr\left(\frac{1}{m}\check{\mathbf{z}}_s^\top\check{\mathbf{Q}}^2_{-s}\check{\mathbf{z}}_s\check{a}_s^2\cdot\mathbf{1}_{\neg\zeta_4}\geq z\mid\zeta'\right)dz\\
&\quad + \int_{2m\|\mathbf{r}\|^2/\ell_{\min}(\mathbf{X})}^\infty\Pr\left(\frac{1}{m}\check{\mathbf{z}}_s^\top\check{\mathbf{Q}}^2_{-s}\check{\mathbf{z}}_s\check{a}_s^2\cdot\mathbf{1}_{\neg\zeta_4}\geq z\mid\zeta'\right)dz\\
&\leq \frac{2m\|\mathbf{r}\|^2\delta_4}{\ell_{\min}(\mathbf{X})} + \int_{2m\|\mathbf{r}\|^2/\ell_{\min}(\mathbf{X})}^\infty\Pr\left(\frac{1}{m}\check{\mathbf{z}}_s^\top\check{\mathbf{Q}}^2_{-s}\check{\mathbf{z}}_s\check{a}_s^2\geq z\mid\zeta'\right)dz\\
&\leq 2mn^\alpha\delta_4\|\mathbf{r}\|^2 + \frac{8^4\theta^3_{\max}(\mathbf{X})p^3\|\mathbf{r}\|^2}{2m^3} = O\left(\frac{1}{m}+\frac{\theta^3_{\max}(\mathbf{X})p^3}{m^3}\right)\cdot\|\mathbf{r}\|^2.
\end{aligned}$$

This further yields

$$W_2 = O\left(\frac{1}{m} + \frac{\theta^3_{\max}(\mathbf{X})p^3}{m^3}\right) \cdot \|\mathbf{r}\|^2.$$

Recalling the result (27) in Lemma E.1, it follows that

$$W_3 \leq \frac{\theta_{\max}(\mathbf{X})p}{m}\|\mathbf{r}\|^2\|\mathbb{E}_\zeta[\check{\mathbf{Q}}^2_{-s} - \check{\mathbf{Q}}^2]\| = O\left(\frac{\theta_{\max}(\mathbf{X})p}{m^2}\right) \cdot \|\mathbf{r}\|^2.$$

Then, applying the result (26) in Lemma E.1 leads to

$$\begin{aligned} W_4 &\leq \max_{1\leq i\leq n} \frac{\|\mathbf{r}\|^2\ell_i(\mathbf{X})}{m\pi_i} \left\|\mathbb{E}_\zeta[\check{\mathbf{Q}}^2] - \mathbf{I}_p - \mathbf{X}_\mathbf{H}^\top \bar{\mathbf{F}}\mathbf{X}_\mathbf{H}\right\| \\ &= O\left(\frac{\log n}{\text{loglog}\, n}\sqrt{\frac{\theta^3_{\max}(\mathbf{X})p^3}{m^3}} \cdot \frac{\theta_{\max}(\mathbf{X})p}{m}\right) \cdot \|\mathbf{r}\|^2 \\ &= O\left(\frac{\log n}{\text{loglog}\, n}\sqrt{\frac{\theta^5_{\max}(\mathbf{X})p^5}{m^5}}\right) \cdot \|\mathbf{r}\|^2. \end{aligned}$$

Then, we turn to bound the term W_5. Taking $\check{\mathbf{Q}} \preceq 8\mathbf{I}_p$ and

$$\|\mathbf{X}_\mathbf{H}^\top \bar{\mathbf{F}}\mathbf{X}_\mathbf{H}\| \leq \max_{1\leq i\leq n} \bar{F}_{ii} \leq \frac{64\theta_{\max}(\mathbf{X})p}{m},$$

it follows that

$$W_5 \leq \max_{1\leq i\leq n} \frac{\ell_i(\mathbf{X})}{m\pi_i}\|\mathbf{r}\|^2 \left\|\mathbf{X}_\mathbf{H}^\top \bar{\mathbf{F}}\mathbf{X}_\mathbf{H}\right\| = O\left(\frac{\theta^2_{\max}(\mathbf{X})p^2}{m^2}\right) \cdot \|\mathbf{r}\|^2.$$

Now, it remains to bound the cross term $m^{-2}\sum_{s\neq k}^m \mathbb{E}_\zeta[\check{\mathbf{z}}_k^\top \check{\mathbf{Q}}^2 \check{\mathbf{z}}_s F_{i_s i_s} F_{i_k i_k} \check{a}_s \check{a}_k]$. Without loss of generality, assume that the events ζ_1, ζ_2 are independent of both $\mathbf{z}_s$ and $\mathbf{z}_k$. Define $\zeta'' = \zeta_1 \bigcap \zeta_2$, $\zeta''' = \zeta_3 \bigcap \zeta_4$ and $\delta_{\zeta'''} = \Pr(\neg\zeta''')$. Let $\check{\gamma}_{i_{s,k}} = 1 + \frac{1}{m}F_{i_k i_k}\check{\mathbf{z}}_k^\top \check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_k$. We first rewrite

$$\begin{aligned} &\frac{1}{m^2}\sum_{s\neq k}^m \mathbb{E}_\zeta[\check{\mathbf{z}}_k^\top \check{\mathbf{Q}}^2 \check{\mathbf{z}}_s F_{i_s i_s} F_{i_k i_k} \check{a}_s \check{a}_k] = \sum_{s\neq k}^m \mathbb{E}_\zeta\left[\frac{\check{\mathbf{z}}_k^\top \check{\mathbf{Q}}^2_{-s}\check{\mathbf{z}}_s F_{i_s i_s} F_{i_k i_k}\check{a}_s\check{a}_k}{m^2\check{\gamma}_{i_s}}\right] \\ &- \sum_{s\neq k}^m \mathbb{E}_\zeta\left[\frac{\check{\mathbf{z}}_k^\top \check{\mathbf{Q}}_{-s}\check{\mathbf{z}}_s\check{\mathbf{z}}_s^\top \check{\mathbf{Q}}^2_{-s}\check{\mathbf{z}}_s F^2_{i_s i_s} F_{i_k i_k}\check{a}_s\check{a}_k}{m^3\check{\gamma}^2_{i_s}}\right] = \sum_{s\neq k}^m \mathbb{E}_\zeta\left[\frac{\check{\mathbf{z}}_k^\top \check{\mathbf{Q}}^2_{-sk}\check{\mathbf{z}}_s F_{i_s i_s} F_{i_k i_k}\check{a}_s\check{a}_k}{m^2\check{\gamma}_{i_s}\check{\gamma}_{i_{s,k}}}\right] \\ &- \sum_{s\neq k}^m \mathbb{E}_\zeta\left[\frac{\check{\mathbf{z}}_k^\top \check{\mathbf{Q}}^2_{-sk}\check{\mathbf{z}}_k\check{\mathbf{z}}_k^\top \check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_s F_{i_s i_s} F^2_{i_k i_k}\check{a}_s\check{a}_k}{m^3\check{\gamma}_{i_s}\check{\gamma}^2_{i_{s,k}}}\right] - \sum_{s\neq k}^m \mathbb{E}_\zeta\left[\frac{\check{\mathbf{z}}_k^\top \check{\mathbf{Q}}_{-s}\check{\mathbf{z}}_s\check{\mathbf{z}}_s^\top \check{\mathbf{Q}}^2_{-s}\check{\mathbf{z}}_s F^2_{i_s i_s} F_{i_k i_k}\check{a}_s\check{a}_k}{m^3\check{\gamma}^2_{i_s}}\right] \\ &= \frac{m(m-1)}{m^2}\mathbb{E}_\zeta\left[\frac{\check{\mathbf{z}}_k^\top \check{\mathbf{Q}}^2_{-sk}\check{\mathbf{z}}_s F_{i_s i_s} F_{i_k i_k}\check{a}_s\check{a}_k}{\check{\gamma}_{i_s}\check{\gamma}_{i_{s,k}}}\right] \\ &- \frac{m(m-1)}{m^2}\mathbb{E}_\zeta\left[\frac{\check{\mathbf{z}}_k^\top \check{\mathbf{Q}}^2_{-sk}\check{\mathbf{z}}_k\check{\mathbf{z}}_k^\top \check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_s F_{i_s i_s} F^2_{i_k i_k}\check{a}_s\check{a}_k}{m\check{\gamma}_{i_s}\check{\gamma}^2_{i_{s,k}}}\right] - \sum_{s\neq k}^m \mathbb{E}_\zeta\left[\frac{\check{\mathbf{z}}_k^\top \check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_s\check{\mathbf{z}}_s^\top \check{\mathbf{Q}}^2_{-s}\check{\mathbf{z}}_s F^2_{i_s i_s} F_{i_k i_k}\check{a}_s\check{a}_k}{m^3\check{\gamma}^2_{i_s}\check{\gamma}_{i_{s,k}}}\right] \\ &= \frac{m(m-1)}{m^2(1-\delta_{\zeta'''})}\left(\mathbb{E}_{\zeta''}\left[\frac{\check{\mathbf{z}}_k^\top \check{\mathbf{Q}}^2_{-sk}\check{\mathbf{z}}_s F_{i_s i_s} F_{i_k i_k}\check{a}_s\check{a}_k}{\check{\gamma}_{i_s}\check{\gamma}_{i_{s,k}}}\right] - \mathbb{E}_{\zeta''}\left[\frac{\check{\mathbf{z}}_k^\top \check{\mathbf{Q}}^2_{-sk}\check{\mathbf{z}}_s F_{i_s i_s} F_{i_k i_k}\check{a}_s\check{a}_k}{\check{\gamma}_{i_s}\check{\gamma}_{i_{s,k}}} \cdot \mathbf{1}_{\neg\zeta'''}\right]\right) \\ &- \frac{m(m-1)}{m^2(1-\delta_{\zeta'''})}\left(\mathbb{E}_{\zeta''}\left[\frac{\check{\mathbf{z}}_k^\top \check{\mathbf{Q}}^2_{-sk}\check{\mathbf{z}}_k\check{\mathbf{z}}_k^\top \check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_s F_{i_s i_s} F^2_{i_k i_k}\check{a}_s\check{a}_k}{m\check{\gamma}_{i_s}\check{\gamma}^2_{i_{s,k}}}\right]\right. \\ &\left.-\mathbb{E}_{\zeta''}\left[\frac{\check{\mathbf{z}}_k^\top \check{\mathbf{Q}}^2_{-sk}\check{\mathbf{z}}_k\check{\mathbf{z}}_k^\top \check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_s F_{i_s i_s} F^2_{i_k i_k}\check{a}_s\check{a}_k}{m\check{\gamma}_{i_s}\check{\gamma}^2_{i_{s,k}}} \cdot \mathbf{1}_{\neg\zeta'''}\right]\right) \end{aligned}$$

$$
\begin{aligned}
&-\sum_{s\neq k}^{m}\mathbb{E}_{\zeta}\left[\frac{\check{\mathbf{z}}_k^\top\check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_s\check{\mathbf{z}}_s^\top\check{\mathbf{Q}}_{-s}^2\check{\mathbf{z}}_s F_{i_si_s}^2F_{i_ki_k}\check{a}_s\check{a}_k}{m^3\check{\gamma}_{i_s}^2\check{\gamma}_{i_{s,k}}}\right]=\underbrace{\frac{m(m-1)}{m^2(1-\delta_{\zeta'''})}\mathbb{E}_{\zeta''}\left[\frac{\check{\mathbf{z}}_k^\top\check{\mathbf{Q}}_{-sk}^2\check{\mathbf{z}}_s F_{i_si_s}F_{i_ki_k}\check{a}_s\check{a}_k}{\check{\gamma}_{i_s}\check{\gamma}_{i_{s,k}}}\right]}_{W_6}\\
&\underbrace{-\frac{m(m-1)}{m^2(1-\delta_{\zeta'''})}\mathbb{E}_{\zeta''}\left[\frac{\check{\mathbf{z}}_k^\top\check{\mathbf{Q}}_{-sk}^2\check{\mathbf{z}}_s F_{i_si_s}F_{i_ki_k}\check{a}_s\check{a}_k}{\check{\gamma}_{i_s}\check{\gamma}_{i_{s,k}}}\cdot\mathbf{1}_{\neg\zeta'''}\right]}_{W_7}\\
&\underbrace{-\frac{m(m-1)}{m^2(1-\delta_{\zeta'''})}\mathbb{E}_{\zeta''}\left[\frac{\check{\mathbf{z}}_k^\top\check{\mathbf{Q}}_{-sk}^2\check{\mathbf{z}}_k\check{\mathbf{z}}_k^\top\check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_s F_{i_si_s}F_{i_ki_k}^2\check{a}_s\check{a}_k}{m\check{\gamma}_{i_s}\check{\gamma}_{i_{s,k}}^2}\right]}_{W_8}\\
&\underbrace{+\frac{m(m-1)}{m^2(1-\delta_{\zeta'''})}\mathbb{E}_{\zeta''}\left[\frac{\check{\mathbf{z}}_k^\top\check{\mathbf{Q}}_{-sk}^2\check{\mathbf{z}}_k\check{\mathbf{z}}_k^\top\check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_s F_{i_si_s}F_{i_ki_k}^2\check{a}_s\check{a}_k}{m\check{\gamma}_{i_s}\check{\gamma}_{i_{s,k}}^2}\cdot\mathbf{1}_{\neg\zeta'''}\right]}_{W_9}\\
&\underbrace{-\sum_{s\neq k}^{m}\frac{1}{1-\delta_{\zeta'''}}\mathbb{E}_{\zeta''}\left[\frac{\check{\mathbf{z}}_k^\top\check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_s\check{\mathbf{z}}_s^\top\check{\mathbf{Q}}_{-s}^2\check{\mathbf{z}}_s F_{i_si_s}^2F_{i_ki_k}\check{a}_s\check{a}_k}{m^3\check{\gamma}_{i_s}^2\check{\gamma}_{i_{s,k}}}\right]}_{W_{10}}\\
&\underbrace{+\sum_{s\neq k}^{m}\frac{1}{1-\delta_{\zeta'''}}\mathbb{E}_{\zeta''}\left[\frac{\check{\mathbf{z}}_k^\top\check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_s\check{\mathbf{z}}_s^\top\check{\mathbf{Q}}_{-s}^2\check{\mathbf{z}}_s F_{i_si_s}^2F_{i_ki_k}\check{a}_s\check{a}_k}{m^3\check{\gamma}_{i_s}^2\check{\gamma}_{i_{s,k}}}\cdot\mathbf{1}_{\neg\zeta'''}\right]}_{W_{11}}.
\end{aligned}
$$

Next, we first bound the term W_6. Define $\bar{\gamma}_{i_{s,k,j}}=1+\frac{1}{m\pi_j}F_{jj}\mathbf{x}_{\mathbf{H}_j}^\top\check{\mathbf{Q}}_{-sk}\mathbf{x}_{\mathbf{H}_j}$ and $\check{\gamma}_{i_{k,s}}=1+\frac{1}{m}F_{i_si_s}\check{\mathbf{z}}_s^\top\check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_s$. We rewrite

$$
\begin{aligned}
W_6=&\underbrace{\frac{m(m-1)}{m^2(1-\delta_{\zeta'''})}\mathbb{E}_{\zeta''}\left[\frac{\check{\mathbf{z}}_k^\top\check{\mathbf{Q}}_{-sk}^2\check{\mathbf{z}}_s F_{i_si_s}F_{i_ki_k}\check{a}_s\check{a}_k}{\check{\gamma}_{i_{k,s}}\check{\gamma}_{i_{s,k}}}\right]}_{W_{61}}\\
&+\underbrace{\frac{m(m-1)}{m^2(1-\delta_{\zeta'''})}\mathbb{E}_{\zeta''}\left[\frac{\check{\mathbf{z}}_k^\top\check{\mathbf{Q}}_{-sk}^2\check{\mathbf{z}}_s F_{i_si_s}F_{i_ki_k}\check{a}_s\check{a}_k}{\check{\gamma}_{i_{s,k}}}\left(\frac{1}{\check{\gamma}_{i_s}}-\frac{1}{\check{\gamma}_{i_{k,s}}}\right)\right]}_{W_{62}}.
\end{aligned}
$$

For the term W_{61}, we get

$$
\begin{aligned}
W_{61}=&\underbrace{\frac{m(m-1)}{m^2(1-\delta_{\zeta'''})}\mathbb{E}_{\zeta''}\left[\frac{\check{\mathbf{z}}_k^\top\check{\mathbf{Q}}_{-sk}^2\check{\mathbf{z}}_s F_{i_ki_k}\check{a}_s\check{a}_k}{\check{\gamma}_{i_{s,k}}}\left(\frac{F_{i_si_s}}{\check{\gamma}_{i_{k,s}}}-1\right)\right]}_{W_{611}}\\
&+\underbrace{\frac{m(m-1)}{m^2(1-\delta_{\zeta'''})}\mathbb{E}_{\zeta''}\left[\check{\mathbf{z}}_k^\top\check{\mathbf{Q}}_{-sk}^2\check{\mathbf{z}}_s\check{a}_s\check{a}_k\left(\frac{F_{i_ki_k}}{\check{\gamma}_{i_{s,k}}}-1\right)\right]}_{W_{612}}\\
&+\underbrace{\frac{m(m-1)}{m^2(1-\delta_{\zeta'''})}\mathbb{E}_{\zeta''}\left[\check{\mathbf{z}}_k^\top\check{\mathbf{Q}}_{-sk}^2\check{\mathbf{z}}_s\check{a}_s\check{a}_k\right]}_{W_{613}}.
\end{aligned}
$$

Denote $\bar{\gamma}_{i_{k,s,i}}=1+\frac{1}{m\pi_i}F_{ii}\mathbf{x}_{\mathbf{H}_i}^\top\check{\mathbf{Q}}_{-sk}\mathbf{x}_{\mathbf{H}_i}$. Together with $\check{\mathbf{Q}}_{-sk}\preceq 8\mathbf{I}_p$ and $F_{jj}<2$, we then obtain

$$
W_{611}=\sum_{j=1}^{n}\sum_{i=1}^{n}\frac{m(m-1)}{m^2(1-\delta_{\zeta'''})}\mathbb{E}_{\zeta''}\left[\frac{\mathbf{x}_{\mathbf{H}_j}^\top\check{\mathbf{Q}}_{-sk}^2\mathbf{x}_{\mathbf{H}_i}F_{jj}r_ir_j}{\bar{\gamma}_{i_{s,k,j}}}\left(\frac{F_{ii}}{\bar{\gamma}_{i_{k,s,i}}}-1\right)\right]
$$

$$= \frac{m(m-1)}{m^2(1-\delta_{\zeta'''})}\mathbb{E}_{\zeta''}\left[\mathbf{r}^\top \operatorname{diag}\left\{\frac{F_{jj}}{\bar{\gamma}_{i_{s,k,j}}}\right\}_{j=1}^n \mathbf{X}_{\mathbf{H}}\check{\mathbf{Q}}_{-sk}^2\mathbf{X}_{\mathbf{H}}^\top \operatorname{diag}\left\{\left(\frac{F_{ii}}{\bar{\gamma}_{i_{k,s,i}}}-1\right)\right\}_{i=1}^n \mathbf{r}\right]$$
$$\leq 256\mathbb{E}_{\zeta''}\left[\max_{1\leq i\leq n}\left|\frac{F_{ii}}{\bar{\gamma}_{i_{k,s,i}}}-1\right|\right]\|\mathbf{r}\|^2 = O\left(\frac{\log n}{\operatorname{loglog} n}\sqrt{\frac{\theta_{\max}^3(\mathbf{X})p^3}{m^3}}\right)\|\mathbf{r}\|^2.$$

where the last line follows from (29) in Lemma E.1. Using $\mathbf{X}_{\mathbf{H}}^\top\mathbf{r}=0$, we next consider the term W_{612}:

$$W_{612} = \sum_{j=1}^n\sum_{i=1}^n \frac{m(m-1)}{m^2(1-\delta_{\zeta'''})}\mathbb{E}_{\zeta''}\left[\mathbf{x}_{\mathbf{H}_j}^\top\check{\mathbf{Q}}_{-sk}^2\mathbf{x}_{\mathbf{H}_i}r_ir_j\left(\frac{F_{jj}}{\bar{\gamma}_{i_{s,k,j}}}-1\right)\right]$$
$$= \frac{m(m-1)}{m^2(1-\delta_{\zeta'''})}\mathbb{E}_{\zeta''}\left[\mathbf{r}^\top \operatorname{diag}\left\{\left(\frac{F_{jj}}{\bar{\gamma}_{i_{s,k,j}}}-1\right)\right\}_{j=1}^n \mathbf{X}_{\mathbf{H}}\check{\mathbf{Q}}_{-sk}^2\mathbf{X}_{\mathbf{H}}^\top\mathbf{r}\right] = 0.$$

Recalling $\mathbf{X}_{\mathbf{H}}^\top\mathbf{r}=0$ again, we now consider W_{613}:

$$W_{613} = \frac{m(m-1)}{m^2(1-\delta_{\zeta'''})}\sum_{j=1}^n\sum_{i=1}^n\mathbb{E}_{\zeta''}\left[\mathbf{x}_{\mathbf{H}_j}^\top\check{\mathbf{Q}}_{-sk}^2\mathbf{x}_{\mathbf{H}_i}r_ir_j\right] = \frac{m(m-1)}{m^2(1-\delta_{\zeta'''})}\mathbb{E}_{\zeta''}\left[\mathbf{r}^\top\mathbf{X}_{\mathbf{H}}\check{\mathbf{Q}}_{-sk}^2\mathbf{X}_{\mathbf{H}}^\top\mathbf{r}\right] = 0.$$

Now, we turn to explore the term W_{62}. Noting $1<F_{ii}<2$, $\check{\mathbf{Q}}_{-sk}\preceq 8\mathbf{I}_p$, and

$$\check{\gamma}_{i_s} = 1+\frac{F_{i_si_s}\check{\mathbf{z}}_s^\top\check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_s}{m} - \frac{F_{i_si_s}F_{i_ki_k}\check{\mathbf{z}}_s^\top\check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_k\check{\mathbf{z}}_k^\top\check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_s}{m^2\check{\gamma}_{i_{s,k}}} \geq 1/2,$$

under $m\geq 256\theta_{\max}p$, we get

$$W_{62} = \frac{m(m-1)}{m^2(1-\delta_{\zeta'''})}\mathbb{E}_{\zeta''}\left[\frac{\check{\mathbf{z}}_k^\top\check{\mathbf{Q}}_{-sk}^2\check{\mathbf{z}}_s\check{\mathbf{z}}_s^\top\check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_k\check{\mathbf{z}}_k^\top\check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_sF_{i_si_s}^2F_{i_ki_k}^2\check{a}_s\check{a}_k}{m^2\check{\gamma}_{i_{s,k}}^2\check{\gamma}_{i_{k,s}}\check{\gamma}_{i_s}}\right]$$
$$= \frac{m(m-1)}{m^2(1-\delta_{\zeta'''})}\sum_{i=1}^n\mathbb{E}_{\zeta''}\left[\frac{\check{\mathbf{z}}_k^\top\check{\mathbf{Q}}_{-sk}^2\mathbf{x}_{\mathbf{H}_i}\mathbf{x}_{\mathbf{H}_i}^\top\check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_k\check{\mathbf{z}}_k^\top\check{\mathbf{Q}}_{-sk}\mathbf{x}_{\mathbf{H}_i}F_{ii}^2F_{i_ki_k}^2r_i\check{a}_k}{m^2\bar{\gamma}_{i_{k,s,i}}\check{\gamma}_{i_{s,k}}^2\bar{\gamma}_i\pi_i}\right]$$
$$\overset{(a)}{\leq}\sum_{i=1}^n\mathbb{E}_{\zeta''}\left[\frac{\mathbf{x}_{\mathbf{H}_i}^\top\check{\mathbf{Q}}_{-sk}^2\check{\mathbf{z}}_k\check{\mathbf{z}}_k^\top\check{\mathbf{Q}}_{-sk}^2\mathbf{x}_{\mathbf{H}_i}(\mathbf{x}_{\mathbf{H}_i}^\top\check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_k\check{\mathbf{z}}_k^\top\check{\mathbf{Q}}_{-sk}\mathbf{x}_{\mathbf{H}_i})^{1/2}F_{ii}^2F_{i_ki_k}^2r_i^2}{m^2\bar{\gamma}_{i_{k,s,i}}\check{\gamma}_{i_{s,k}}^2\bar{\gamma}_i\pi_i^{3/2}}\right]$$
$$+\sum_{i=1}^n\mathbb{E}_{\zeta''}\left[\frac{F_{ii}^2F_{i_ki_k}^2\check{a}_k^2(\mathbf{x}_{\mathbf{H}_i}^\top\check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_k\check{\mathbf{z}}_k^\top\check{\mathbf{Q}}_{-sk}\mathbf{x}_{\mathbf{H}_i})^{3/2}}{m^2\bar{\gamma}_{i_{k,s,i}}\check{\gamma}_{i_{s,k}}^2\bar{\gamma}_i\pi_i^{1/2}}\right]$$
$$\leq\frac{128\theta_{\max}(\mathbf{X})p}{m}\sum_{i=1}^n\mathbb{E}_{\zeta''}\left[\frac{\mathbf{x}_{\mathbf{H}_i}^\top\check{\mathbf{Q}}_{-sk}^2\check{\mathbf{z}}_k\check{\mathbf{z}}_k^\top\check{\mathbf{Q}}_{-sk}^2\mathbf{x}_{\mathbf{H}_i}r_i^2}{m\pi_i}\right]$$
$$+\frac{128\theta_{\max}(\mathbf{X})p}{m}\sum_{i=1}^n\mathbb{E}_{\zeta''}\left[\frac{\check{a}_k^2\check{\mathbf{z}}_k^\top\check{\mathbf{Q}}_{-sk}\mathbf{x}_{\mathbf{H}_i}\mathbf{x}_{\mathbf{H}_i}^\top\check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_k}{m}\right]$$
$$\leq\frac{128\theta_{\max}(\mathbf{X})p}{m}\sum_{i=1}^n\sum_{j=1}^n\mathbb{E}_{\zeta''}\left[\frac{\mathbf{x}_{\mathbf{H}_i}^\top\check{\mathbf{Q}}_{-sk}^2\mathbf{x}_{\mathbf{H}_j}\mathbf{x}_{\mathbf{H}_j}^\top\check{\mathbf{Q}}_{-sk}^2\mathbf{x}_{\mathbf{H}_i}r_i^2}{m\pi_i}\right]$$
$$+\frac{128\theta_{\max}(\mathbf{X})p}{m}\mathbb{E}_{\zeta''}\left[\frac{\check{a}_k^2\check{\mathbf{z}}_k^\top\check{\mathbf{Q}}_{-sk}\mathbf{X}_{\mathbf{H}}^\top\mathbf{X}_{\mathbf{H}}\check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_k}{m}\right]$$
$$\leq\frac{128\theta_{\max}(\mathbf{X})p}{m}\sum_{i=1}^n\mathbb{E}_{\zeta''}\left[\frac{\mathbf{x}_{\mathbf{H}_i}^\top\check{\mathbf{Q}}_{-sk}^2\mathbf{X}_{\mathbf{H}}^\top\mathbf{X}_{\mathbf{H}}\check{\mathbf{Q}}_{-sk}^2\mathbf{x}_{\mathbf{H}_i}r_i^2}{m\pi_i}\right]+\frac{8192\theta_{\max}^2(\mathbf{X})p^2}{m^2}\mathbb{E}_{\zeta''}\left[\check{a}_k^2\right]$$
$$\leq\frac{2\times 8^6\theta_{\max}^2(\mathbf{X})p^2}{m^2}\sum_{i=1}^n r_i^2+\frac{8192\theta_{\max}^2(\mathbf{X})p^2}{m^2}\sum_{j=1}^n r_j^2 = O\left(\frac{\theta_{\max}^2(\mathbf{X})p^2}{m^2}\right)\cdot\|\mathbf{r}\|^2. \tag{19}$$

where in (a) we use the inequality $2xy \leq x^2 + y^2$. Thus, we conclude

$$W_6 = O\left(\frac{\log n}{\log\log n}\sqrt{\frac{\theta^3_{\max}(\mathbf{X})p^3}{m^3}}\right)\|\mathbf{r}\|^2.$$

Next, we proceed to analyze the term W_7. Applying

$$\begin{aligned}
&\mathbb{E}_{\zeta''}\left[\left|\frac{\check{\mathbf{z}}_k^\top \check{\mathbf{Q}}^2_{-sk}\check{\mathbf{z}}_s F_{i_s i_s}F_{i_k i_k}\check{a}_s\check{a}_k}{\check{\gamma}_{i_s}\check{\gamma}_{i_{s,k}}}\right|\right] \leq 4\mathbb{E}_{\zeta''}\left[|\check{\mathbf{z}}_k^\top \check{\mathbf{Q}}^2_{-sk}\check{\mathbf{z}}_s\check{a}_s\check{a}_k|\right] \\
&\leq 4\sqrt{\mathbb{E}_{\zeta''}\left[\check{\mathbf{z}}_k^\top \check{\mathbf{Q}}^2_{-sk}\check{\mathbf{z}}_s\check{\mathbf{z}}_s^\top \check{\mathbf{Q}}^2_{-sk}\check{\mathbf{z}}_k\right]\mathbb{E}_{\zeta''}\left[\check{a}_s^2\check{a}_k^2\right]} = 4\|\mathbf{r}\|^2\sqrt{\mathbb{E}_{\zeta''}\left[\check{\mathbf{z}}_k^\top \check{\mathbf{Q}}^2_{-sk}\mathbf{X}_{\mathbf{H}}^\top\mathbf{X}_{\mathbf{H}}\check{\mathbf{Q}}^2_{-sk}\check{\mathbf{z}}_k\right]} \\
&\leq 4\|\mathbf{r}\|^2\sqrt{\sum_{j=1}^n \mathbb{E}_{\zeta''}\left[\mathbf{x}_{\mathbf{H}_j}^\top \check{\mathbf{Q}}^2_{-sk}\mathbf{X}_{\mathbf{H}}^\top\mathbf{X}_{\mathbf{H}}\check{\mathbf{Q}}^2_{-sk}\mathbf{x}_{\mathbf{H}_j}\right]} \leq 256p^{1/2}\|\mathbf{r}\|^2, \qquad (20)
\end{aligned}$$

and

$$\begin{aligned}
&\mathrm{Var}_{\zeta''}\left[\frac{\check{\mathbf{z}}_k^\top \check{\mathbf{Q}}^2_{-sk}\check{\mathbf{z}}_s F_{i_s i_s}F_{i_k i_k}\check{a}_s\check{a}_k}{\check{\gamma}_{i_s}\check{\gamma}_{i_{s,k}}}\right] \leq \sum_{j=1}^n\sum_{i=1}^n \mathbb{E}_{\zeta''}\left[\frac{(\mathbf{x}_{\mathbf{H}_j}^\top \check{\mathbf{Q}}^2_{-sk}\mathbf{x}_{\mathbf{H}_i})^2 F_{ii}^2F_{jj}^2r_i^2r_j^2}{\pi_j\pi_i\bar{\gamma}_i^2\bar{\gamma}^2_{i_{s,k},j}}\right] \\
&\leq \sum_{j=1}^n\sum_{i=1}^n \frac{16}{\pi_j\pi_i}\mathbb{E}_{\zeta''}\left[(\mathbf{x}_{\mathbf{H}_j}^\top \check{\mathbf{Q}}^2_{-sk}\mathbf{x}_{\mathbf{H}_i})^2 r_i^2r_j^2\right] \leq 2\times 8^5\theta^2_{\max}(\mathbf{X})p^2\|\mathbf{r}\|^4, \qquad (21)
\end{aligned}$$

together with the Chebyshev's inequality, it follows that for $z > 512p^{1/2}\|\mathbf{r}\|^2$,

$$\Pr\left(\left|\frac{\check{\mathbf{z}}_k^\top \check{\mathbf{Q}}^2_{-sk}\check{\mathbf{z}}_s F_{i_s i_s}F_{i_k i_k}\check{a}_s\check{a}_k}{\check{\gamma}_{i_s}\check{\gamma}_{i_{s,k}}}\right| \geq z \mid \zeta''\right) \leq \frac{2\times 8^5\theta^2_{\max}(\mathbf{X})p^2\|\mathbf{r}\|^4}{z^2}.$$

With $\delta_{\zeta''} \leq m^{-3}$, we then deduce

$$\begin{aligned}
&\mathbb{E}_{\zeta''}\left[\frac{\check{\mathbf{z}}_k^\top \check{\mathbf{Q}}^2_{-sk}\check{\mathbf{z}}_s F_{i_s i_s}F_{i_k i_k}\check{a}_s\check{a}_k}{\check{\gamma}_{i_s}\check{\gamma}_{i_{s,k}}}\cdot\mathbf{1}_{\neg\zeta'''}\right] \leq \mathbb{E}_{\zeta''}\left[\left|\frac{\check{\mathbf{z}}_k^\top \check{\mathbf{Q}}^2_{-sk}\check{\mathbf{z}}_s F_{i_s i_s}F_{i_k i_k}\check{a}_s\check{a}_k}{\check{\gamma}_{i_s}\check{\gamma}_{i_{s,k}}}\right|\cdot\mathbf{1}_{\neg\zeta'''}\right] \\
&= \int_0^\infty \Pr\left(\left|\frac{\check{\mathbf{z}}_k^\top \check{\mathbf{Q}}^2_{-sk}\check{\mathbf{z}}_s F_{i_s i_s}F_{i_k i_k}\check{a}_s\check{a}_k}{\check{\gamma}_{i_s}\check{\gamma}_{i_{s,k}}}\right|\cdot\mathbf{1}_{\neg\zeta'''} \geq z \mid \zeta''\right)dz \\
&= \int_0^{2m^2\|\mathbf{r}\|^2} \Pr\left(\left|\frac{\check{\mathbf{z}}_k^\top \check{\mathbf{Q}}^2_{-sk}\check{\mathbf{z}}_s F_{i_s i_s}F_{i_k i_k}\check{a}_s\check{a}_k}{\check{\gamma}_{i_s}\check{\gamma}_{i_{s,k}}}\right|\cdot\mathbf{1}_{\neg\zeta'''} \geq z \mid \zeta''\right)dz \\
&+ \int_{2m^2\|\mathbf{r}\|^2}^\infty \Pr\left(\left|\frac{\check{\mathbf{z}}_k^\top \check{\mathbf{Q}}^2_{-sk}\check{\mathbf{z}}_s F_{i_s i_s}F_{i_k i_k}\check{a}_s\check{a}_k}{\check{\gamma}_{i_s}\check{\gamma}_{i_{s,k}}}\right|\cdot\mathbf{1}_{\neg\zeta'''} \geq z \mid \zeta''\right)dz \\
&\leq \frac{2\|\mathbf{r}\|^2}{m} + \int_{2m^2\|\mathbf{r}\|^2}^\infty \Pr\left(\left|\frac{\check{\mathbf{z}}_k^\top \check{\mathbf{Q}}^2_{-sk}\check{\mathbf{z}}_s F_{i_s i_s}F_{i_k i_k}\check{a}_s\check{a}_k}{\check{\gamma}_{i_s}\check{\gamma}_{i_{s,k}}}\right|\cdot\mathbf{1}_{\neg\zeta'''} \geq z \mid \zeta''\right)dz \\
&\leq \frac{2\|\mathbf{r}\|^2}{m} + \frac{2\times 8^5\theta^2_{\max}(\mathbf{X})p^2\|\mathbf{r}\|^2}{2m^2} = O\left(\frac{1}{m} + \frac{\theta^2_{\max}(\mathbf{X})p^2}{m^2}\right)\cdot\|\mathbf{r}\|^2.
\end{aligned}$$

This thus gives

$$W_7 = O\left(\frac{1}{m} + \frac{\theta^2_{\max}(\mathbf{X})p^2}{m^2}\right)\cdot\|\mathbf{r}\|^2.$$

Subsequently, we continue to bound the term W_8. We get

$$W_8 = \underbrace{\frac{m(m-1)}{m^2(1-\delta_{\zeta'''})}\mathbb{E}_{\zeta''}\left[\frac{\check{\mathbf{z}}_k^\top \check{\mathbf{Q}}^2_{-sk}\check{\mathbf{z}}_k\check{\mathbf{z}}_k^\top \check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_s F_{i_s i_s}F^2_{i_k i_k}\check{a}_s\check{a}_k}{m\check{\gamma}_{i_{k,s}}\check{\gamma}^2_{i_{s,k}}}\right]}_{W_{81}}$$

$$+\underbrace{\frac{m(m-1)}{m^2(1-\delta_{\zeta'''})}\mathbb{E}_{\zeta''}\left[\frac{\check{\mathbf{z}}_k^\top\check{\mathbf{Q}}_{-sk}^2\check{\mathbf{z}}_k\check{\mathbf{z}}_k^\top\check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_s F_{i_s i_s}F_{i_k i_k}^2\check{a}_s\check{a}_k}{m\check{\gamma}_{i_{s,k}}^2}\left(\frac{1}{\check{\gamma}_{i_s}}-\frac{1}{\check{\gamma}_{i_{k,s}}}\right)\right]}_{W_{82}}.$$

In the following, we consider the first term W_{81}:

$$\begin{aligned}W_{81}=&\underbrace{\frac{m(m-1)}{m^2(1-\delta_{\zeta'''})}\mathbb{E}_{\zeta''}\left[\frac{\check{\mathbf{z}}_k^\top\check{\mathbf{Q}}_{-sk}^2\check{\mathbf{z}}_k\check{\mathbf{z}}_k^\top\check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_s F_{i_k i_k}^2\check{a}_s\check{a}_k}{m\check{\gamma}_{i_{s,k}}^2}\left(\frac{F_{i_s i_s}}{\check{\gamma}_{i_{k,s}}}-1\right)\right]}_{W_{811}}\\&+\underbrace{\frac{m(m-1)}{m^3(1-\delta_{\zeta'''})}\mathbb{E}_{\zeta''}\left[\check{\mathbf{z}}_k^\top\check{\mathbf{Q}}_{-sk}^2\check{\mathbf{z}}_k\check{\mathbf{z}}_k^\top\check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_s\check{a}_s\check{a}_k\left(\frac{F_{i_k i_k}^2}{\check{\gamma}_{i_{s,k}}^2}-1\right)\right]}_{W_{812}}\\&+\underbrace{\frac{m(m-1)}{m^3(1-\delta_{\zeta'''})}\mathbb{E}_{\zeta''}\left[\check{\mathbf{z}}_k^\top\check{\mathbf{Q}}_{-sk}^2\check{\mathbf{z}}_k\check{\mathbf{z}}_k^\top\check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_s\check{a}_s\check{a}_k\right]}_{W_{813}}.\end{aligned}$$

Recalling $F_{ii}\le 2$ and $\check{\mathbf{Q}}_{-sk}\preceq 8\mathbf{I}_p$ again, together with the result (29) in Lemma E.1, we get

$$\begin{aligned}W_{811}&=\sum_{j=1}^n\sum_{i=1}^n\frac{m(m-1)}{m^2(1-\delta_{\zeta'''})}\mathbb{E}_{\zeta''}\left[\frac{\mathbf{x}_{\mathbf{H}_j}^\top\check{\mathbf{Q}}_{-sk}^2\mathbf{x}_{\mathbf{H}_j}\mathbf{x}_{\mathbf{H}_j}^\top\check{\mathbf{Q}}_{-sk}\mathbf{x}_{\mathbf{H}_i}F_{jj}^2 r_i r_j}{m\pi_j\bar{\gamma}_{i_{s,k,j}}^2}\left(\frac{F_{ii}}{\bar{\gamma}_{i_{k,s,i}}}-1\right)\right]\\&=\sum_{j=1}^n\sum_{i=1}^n\frac{m(m-1)}{m^2(1-\delta_{\zeta'''})}\mathbb{E}_{\zeta''}\left[\frac{\mathbf{x}_{\mathbf{H}_j}^\top\check{\mathbf{Q}}_{-sk}^2\mathbf{x}_{\mathbf{H}_j}\mathbf{x}_{\mathbf{H}_j}^\top\check{\mathbf{Q}}_{-sk}\mathbf{x}_{\mathbf{H}_i}F_{jj}^2 r_i r_j}{m\pi_j\bar{\gamma}_{i_{s,k,j}}^2}\left(\frac{F_{ii}}{\bar{\gamma}_{i_{k,s,i}}}-1\right)\right]\\&=\frac{m(m-1)}{m^2(1-\delta_{\zeta'''})}\mathbb{E}_{\zeta''}\left[\mathbf{r}^\top\operatorname{diag}\left\{\frac{F_{jj}^2\mathbf{x}_{\mathbf{H}_j}^\top\check{\mathbf{Q}}_{-sk}^2\mathbf{x}_{\mathbf{H}_j}}{m\pi_j\bar{\gamma}_{i_{s,k,j}}^2}\right\}_{j=1}^n\mathbf{X}_{\mathbf{H}}\check{\mathbf{Q}}_{-sk}\mathbf{X}_{\mathbf{H}}^\top\operatorname{diag}\left\{\frac{F_{ii}}{\bar{\gamma}_{i_{k,s,i}}}-1\right\}_{i=1}^n\mathbf{r}\right]\\&\le\frac{4096\theta_{\max}(\mathbf{X})p}{m}\mathbb{E}_{\zeta''}\left[\max_{1\le i\le n}\left|\frac{F_{ii}}{\bar{\gamma}_{i_{k,s,i}}}-1\right|\right]\cdot\|\mathbf{r}\|^2=O\left(\frac{\log n}{\operatorname{loglog} n}\sqrt{\frac{\theta_{\max}^5(\mathbf{X})p^5}{m^5}}\right)\cdot\|\mathbf{r}\|^2.\end{aligned}$$

Now, we turn to analyze the term W_{812}. Applying $\mathbf{X}_{\mathbf{H}}^\top\mathbf{r}=0$ again leads to

$$W_{812}=\sum_{j=1}^n\sum_{i=1}^n\frac{m(m-1)}{m^2(1-\delta_{\zeta'''})}\mathbb{E}_{\zeta''}\left[\frac{\mathbf{x}_{\mathbf{H}_j}^\top\check{\mathbf{Q}}_{-sk}^2\mathbf{x}_{\mathbf{H}_j}\mathbf{x}_{\mathbf{H}_j}^\top\check{\mathbf{Q}}_{-sk}\mathbf{x}_{\mathbf{H}_i}r_i r_j}{m\pi_j}\left(\frac{F_{jj}^2}{\bar{\gamma}_{i_{s,k,j}}^2}-1\right)\right]=0.$$

Similarly, we get

$$W_{813}=\frac{m(m-1)}{m^2(1-\delta_{\zeta'''})}\sum_{j=1}^n\sum_{i=1}^n\frac{1}{m\pi_j}\mathbb{E}_{\zeta''}\left[\mathbf{x}_{\mathbf{H}_j}^\top\check{\mathbf{Q}}_{-sk}^2\mathbf{x}_{\mathbf{H}_j}\mathbf{x}_{\mathbf{H}_j}^\top\check{\mathbf{Q}}_{-sk}\mathbf{x}_{\mathbf{H}_i}r_i r_j\right]=0.$$

We then consider the term W_{82}, with $F_{ii}<2$ and $\check{\mathbf{Q}}_{-sk}\preceq 8\mathbf{I}_p$,

$$\begin{aligned}W_{82}&\le\frac{m(m-1)}{m^2(1-\delta_{\zeta'''})}\mathbb{E}_{\zeta''}\left[\frac{\check{\mathbf{z}}_k^\top\check{\mathbf{Q}}_{-sk}^2\check{\mathbf{z}}_k\check{\mathbf{z}}_k^\top\check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_s\check{\mathbf{z}}_s^\top\check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_k\check{\mathbf{z}}_k^\top\check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_s F_{i_s i_s}^2F_{i_k i_k}^3\check{a}_s\check{a}_k}{m^3\check{\gamma}_{i_s}\check{\gamma}_{i_{k,s}}\check{\gamma}_{i_{s,k}}^3}\right]\\&=\frac{m(m-1)}{m^2(1-\delta_{\zeta'''})}\sum_{i=1}^n\mathbb{E}_{\zeta''}\left[\frac{\check{\mathbf{z}}_k^\top\check{\mathbf{Q}}_{-sk}^2\check{\mathbf{z}}_k\check{\mathbf{z}}_k^\top\check{\mathbf{Q}}_{-sk}\mathbf{x}_{\mathbf{H}_i}\mathbf{x}_{\mathbf{H}_i}^\top\check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_k\check{\mathbf{z}}_k^\top\check{\mathbf{Q}}_{-sk}\mathbf{x}_{\mathbf{H}_i}F_{ii}^2F_{i_k i_k}^3r_i\check{a}_k}{m^3\bar{\gamma}_{i_{k,s,i}}\check{\gamma}_{i_{s,k}}^3\bar{\gamma}_i\pi_i}\right]\\&\le\sum_{i=1}^n\mathbb{E}_{\zeta''}\left[\frac{\check{\mathbf{z}}_k^\top\check{\mathbf{Q}}_{-sk}^2\check{\mathbf{z}}_k\mathbf{x}_{\mathbf{H}_i}^\top\check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_k\check{\mathbf{z}}_k^\top\check{\mathbf{Q}}_{-sk}\mathbf{x}_{\mathbf{H}_i}(\mathbf{x}_{\mathbf{H}_i}^\top\check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_k\check{\mathbf{z}}_k^\top\check{\mathbf{Q}}_{-sk}\mathbf{x}_{\mathbf{H}_i})^{1/2}F_{ii}^2F_{i_k i_k}^3r_i^2}{m^3\bar{\gamma}_{i_{k,s,i}}\check{\gamma}_{i_{s,k}}^3\bar{\gamma}_i\pi_i^{3/2}}\right]\\&+\sum_{i=1}^n\mathbb{E}_{\zeta''}\left[\frac{F_{ii}^2F_{i_k i_k}^3\check{a}_k^2\check{\mathbf{z}}_k^\top\check{\mathbf{Q}}_{-sk}^2\check{\mathbf{z}}_k(\mathbf{x}_{\mathbf{H}_i}^\top\check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_k\check{\mathbf{z}}_k^\top\check{\mathbf{Q}}_{-sk}\mathbf{x}_{\mathbf{H}_i})^{3/2}}{m^3\bar{\gamma}_{i_{k,s,i}}\check{\gamma}_{i_{s,k}}^3\bar{\gamma}_i\pi_i^{1/2}}\right]\end{aligned}$$

$$
\begin{aligned}
&\leq \frac{16384\theta_{\max}^2(\mathbf{X})p^2}{m^2}\sum_{i=1}^{n}\mathbb{E}_{\zeta''}\left[\frac{\mathbf{x}_{\mathbf{H}_i}^\top\check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_k\check{\mathbf{z}}_k^\top\check{\mathbf{Q}}_{-sk}\mathbf{x}_{\mathbf{H}_i}r_i^2}{m\pi_i}\right]\\
&+\frac{16384\theta_{\max}^2(\mathbf{X})p^2}{m^2}\sum_{i=1}^{n}\mathbb{E}_{\zeta''}\left[\frac{\check{a}_k^2\check{\mathbf{z}}_k^\top\check{\mathbf{Q}}_{-sk}\mathbf{x}_{\mathbf{H}_i}\mathbf{x}_{\mathbf{H}_i}^\top\check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_k}{m}\right]\\
&=O\left(\frac{\theta_{\max}^3(\mathbf{X})p^3}{m^3}\right)\cdot\|\mathbf{r}\|^2. \qquad (22)
\end{aligned}
$$

We now proceed to bound the term W_9. Similar to (20) and (21), we have

$$
\begin{aligned}
&\mathbb{E}_{\zeta''}\left[\left|\frac{\check{\mathbf{z}}_k^\top\check{\mathbf{Q}}_{-sk}^2\check{\mathbf{z}}_k\check{\mathbf{z}}_k^\top\check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_sF_{i_si_s}F_{i_ki_k}^2\check{a}_s\check{a}_k}{m\check{\gamma}_{i_s}\check{\gamma}_{i_{s,k}}^2}\right|\right]\leq\frac{512\theta_{\max}(\mathbf{X})p}{m}\mathbb{E}_{\zeta''}\left[|\check{\mathbf{z}}_k^\top\check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_s\check{a}_s\check{a}_k|\right]\\
&\leq\frac{512\theta_{\max}(\mathbf{X})p}{m}\sqrt{\mathbb{E}_{\zeta''}\left[\check{\mathbf{z}}_k^\top\check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_s\check{\mathbf{z}}_s^\top\check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_k\right]\mathbb{E}_{\zeta''}\left[\check{a}_s^2\check{a}_k^2\right]}\\
&=\frac{512\theta_{\max}(\mathbf{X})p\|\mathbf{r}\|^2}{m}\sqrt{\mathbb{E}_{\zeta''}\left[\check{\mathbf{z}}_k^\top\check{\mathbf{Q}}_{-sk}\mathbf{X}_{\mathbf{H}}^\top\mathbf{X}_{\mathbf{H}}\check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_k\right]}\\
&\leq\frac{512\theta_{\max}(\mathbf{X})p\|\mathbf{r}\|^2}{m}\sqrt{\sum_{j=1}^{n}\mathbb{E}_{\zeta''}\left[\mathbf{x}_{\mathbf{H}_j}^\top\check{\mathbf{Q}}_{-sk}\mathbf{X}_{\mathbf{H}}^\top\mathbf{X}_{\mathbf{H}}\check{\mathbf{Q}}_{-sk}\mathbf{x}_{\mathbf{H}_j}\right]}\leq\frac{4096\theta_{\max}(\mathbf{X})p^{3/2}\|\mathbf{r}\|^2}{m},
\end{aligned}
$$

and

$$
\begin{aligned}
&\mathrm{Var}_{\zeta''}\left[\frac{\check{\mathbf{z}}_k^\top\check{\mathbf{Q}}_{-sk}^2\check{\mathbf{z}}_k\check{\mathbf{z}}_k^\top\check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_sF_{i_si_s}F_{i_ki_k}^2\check{a}_s\check{a}_k}{m\check{\gamma}_{i_s}\check{\gamma}_{i_{s,k}}^2}\right]\\
&\leq\sum_{j=1}^{n}\sum_{i=1}^{n}\mathbb{E}_{\zeta''}\left[\frac{(\mathbf{x}_{\mathbf{H}_j}^\top\check{\mathbf{Q}}_{-sk}^2\mathbf{x}_{\mathbf{H}_j}\mathbf{x}_{\mathbf{H}_j}^\top\check{\mathbf{Q}}_{-sk}\mathbf{x}_{\mathbf{H}_i})^2F_{ii}^2F_{jj}^4r_i^2r_j^2}{m^2\pi_j^3\pi_i\bar{\gamma}_i^2\bar{\gamma}_{i_{s,k,j}}^4}\right]\\
&\leq 64\sum_{j=1}^{n}\sum_{i=1}^{n}\frac{1}{m^2\pi_j^3\pi_i}\mathbb{E}_{\zeta''}\left[(\mathbf{x}_{\mathbf{H}_j}^\top\check{\mathbf{Q}}_{-sk}^2\mathbf{x}_{\mathbf{H}_j}\mathbf{x}_{\mathbf{H}_j}^\top\check{\mathbf{Q}}_{-sk}\mathbf{x}_{\mathbf{H}_i})^2r_i^2r_j^2\right]\leq\frac{8^8\theta_{\max}^4(\mathbf{X})p^4}{m^2}\|\mathbf{r}\|^4,
\end{aligned}
$$

Similarly to the bound of W_7, we obtain

$$
W_9=O\left(\frac{1}{m}+\frac{\theta_{\max}^4(\mathbf{X})p^4}{m^4}\right)\cdot\|\mathbf{r}\|^2.
$$

In the following, we turn to explore the term W_{10}. We first rewrite

$$
\begin{aligned}
W_{10}&=\sum_{s\neq k}^{m}\frac{1}{1-\delta_{\zeta'''}}\mathbb{E}_{\zeta''}\left[\frac{\check{\mathbf{z}}_k^\top\check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_s\check{\mathbf{z}}_s^\top\check{\mathbf{Q}}_{-s}^2\check{\mathbf{z}}_sF_{i_si_s}^2F_{i_ki_k}\check{a}_s\check{a}_k}{m^3\check{\gamma}_{i_s}^2\check{\gamma}_{i_{s,k}}}\right]\\
&=\frac{m(m-1)}{m^2(1-\delta_{\zeta'''})}\mathbb{E}_{\zeta''}\left[\frac{\check{\mathbf{z}}_k^\top\check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_s\check{\mathbf{z}}_s^\top\check{\mathbf{Q}}_{-s}^2\check{\mathbf{z}}_sF_{i_si_s}^2F_{i_ki_k}\check{a}_s\check{a}_k}{m\check{\gamma}_{i_s}^2\check{\gamma}_{i_{s,k}}}\right]\\
&=\underbrace{\frac{m(m-1)}{m^2(1-\delta_{\zeta'''})}\mathbb{E}_{\zeta''}\left[\frac{\check{\mathbf{z}}_k^\top\check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_s\check{\mathbf{z}}_s^\top\check{\mathbf{Q}}_{-s}^2\check{\mathbf{z}}_sF_{i_si_s}^2F_{i_ki_k}\check{a}_s\check{a}_k}{m\check{\gamma}_{i_{k,s}}^2\check{\gamma}_{i_{s,k}}}\right]}_{W_{101}}\\
&+\underbrace{\frac{m(m-1)}{m^2(1-\delta_{\zeta'''})}\mathbb{E}_{\zeta''}\left[\frac{\check{\mathbf{z}}_k^\top\check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_s\check{\mathbf{z}}_s^\top\check{\mathbf{Q}}_{-s}^2\check{\mathbf{z}}_sF_{i_si_s}^2F_{i_ki_k}\check{a}_s\check{a}_k}{m\check{\gamma}_{i_{s,k}}}\left(\frac{1}{\check{\gamma}_{i_s}^2}-\frac{1}{\check{\gamma}_{i_{k,s}}^2}\right)\right]}_{W_{102}}.
\end{aligned}
$$

Then, we write the term W_{101}:

$$
W_{101}=\underbrace{\frac{m(m-1)}{m^2(1-\delta_{\zeta'''})}\mathbb{E}_{\zeta''}\left[\frac{\check{\mathbf{z}}_k^\top\check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_s\check{\mathbf{z}}_s^\top\check{\mathbf{Q}}_{-s}^2\check{\mathbf{z}}_sF_{i_ki_k}\check{a}_s\check{a}_k}{m\check{\gamma}_{i_{s,k}}}\left(\frac{F_{i_si_s}^2}{\check{\gamma}_{i_{k,s}}^2}-1\right)\right]}_{W_{1011}}
$$

$$+\underbrace{\frac{m(m-1)}{m^2(1-\delta_{\zeta'''})}\mathbb{E}_{\zeta''}\left[\frac{\check{\mathbf{z}}_k^\top\check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_s\check{\mathbf{z}}_s^\top\check{\mathbf{Q}}_{-s}^2\check{\mathbf{z}}_s\check{a}_s\check{a}_k}{m}\left(\frac{F_{i_ki_k}}{\check{\gamma}_{i_{s,k}}}-1\right)\right]}_{W_{1012}}$$
$$+\underbrace{\frac{m(m-1)}{m^2(1-\delta_{\zeta'''})}\mathbb{E}_{\zeta''}\left[\frac{\check{\mathbf{z}}_k^\top\check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_s\check{\mathbf{z}}_s^\top\check{\mathbf{Q}}_{-s}^2\check{\mathbf{z}}_s\check{a}_s\check{a}_k}{m}\right]}_{W_{1013}}.$$

For the term W_{1013}, we have

$$\begin{aligned}
W_{1013} &= \frac{m(m-1)}{m^2(1-\delta_{\zeta'''})}\mathbb{E}_{\zeta''}\left[\frac{\check{\mathbf{z}}_k^\top\check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_s\check{\mathbf{z}}_s^\top\check{\mathbf{Q}}_{-sk}\check{\mathbf{Q}}_{-s}\check{\mathbf{z}}_s\check{a}_s\check{a}_k}{m}\right]\\
&\quad-\frac{m(m-1)}{m^2(1-\delta_{\zeta'''})}\mathbb{E}_{\zeta''}\left[\frac{\check{\mathbf{z}}_k^\top\check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_s\check{\mathbf{z}}_s^\top\check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_k\check{\mathbf{z}}_k^\top\check{\mathbf{Q}}_{-sk}\check{\mathbf{Q}}_{-s}\check{\mathbf{z}}_sF_{i_ki_k}\check{a}_s\check{a}_k}{m^2\check{\gamma}_{i_{s,k}}}\right]\\
&=\underbrace{-\frac{m(m-1)}{m^2(1-\delta_{\zeta'''})}\mathbb{E}_{\zeta''}\left[\frac{\check{\mathbf{z}}_k^\top\check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_s\check{\mathbf{z}}_s^\top\check{\mathbf{Q}}_{-sk}^2\check{\mathbf{z}}_k\check{\mathbf{z}}_k^\top\check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_sF_{i_ki_k}\check{a}_s\check{a}_k}{m^2\check{\gamma}_{i_{s,k}}}\right]}_{W_{10131}}\\
&\quad+\underbrace{\frac{m(m-1)}{m^2(1-\delta_{\zeta'''})}\mathbb{E}_{\zeta''}\left[\frac{\check{\mathbf{z}}_k^\top\check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_s\check{\mathbf{z}}_s^\top\check{\mathbf{Q}}_{-sk}^2\check{\mathbf{z}}_s\check{a}_s\check{a}_k}{m}\right]}_{W_{10132}}\\
&\quad+\underbrace{\frac{m(m-1)}{m^2(1-\delta_{\zeta'''})}\mathbb{E}_{\zeta''}\left[\frac{\check{\mathbf{z}}_k^\top\check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_s\check{\mathbf{z}}_s^\top\check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_k\check{\mathbf{z}}_k^\top\check{\mathbf{Q}}_{-sk}^2\check{\mathbf{z}}_k\check{\mathbf{z}}_k^\top\check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_sF_{i_ki_k}^2\check{a}_s\check{a}_k}{m^3\check{\gamma}_{i_{s,k}}^2}\right]}_{W_{10133}}\\
&\quad\underbrace{-\frac{m(m-1)}{m^2(1-\delta_{\zeta'''})}\mathbb{E}_{\zeta''}\left[\frac{\check{\mathbf{z}}_k^\top\check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_s\check{\mathbf{z}}_s^\top\check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_k\check{\mathbf{z}}_k^\top\check{\mathbf{Q}}_{-sk}^2\check{\mathbf{z}}_s\check{\mathbf{z}}_sF_{i_ki_k}\check{a}_s\check{a}_k}{m^2\check{\gamma}_{i_{s,k}}}\right]}_{W_{10134}}. \qquad (23)
\end{aligned}$$

Applying (19) yields

$$\begin{aligned}
W_{10131} = W_{10134} &= -\frac{m(m-1)}{m^2(1-\delta_{\zeta'''})}\mathbb{E}_{\zeta''}\left[\frac{\check{\mathbf{z}}_k^\top\check{\mathbf{Q}}_{-sk}^2\check{\mathbf{z}}_s\check{\mathbf{z}}_s^\top\check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_k\check{\mathbf{z}}_k^\top\check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_sF_{i_ki_k}\check{a}_s\check{a}_k}{m^2\check{\gamma}_{i_{s,k}}}\right]\\
&= O\left(\frac{\theta_{\max}^2(\mathbf{X})p^2}{m^2}\right)\cdot\|\mathbf{r}\|^2.
\end{aligned}$$

Taking the fact $\mathbf{X}_{\mathbf{H}}^\top\mathbf{r}=0$, we further obtain

$$W_{10132} = \frac{m(m-1)}{m^2(1-\delta_{\zeta'''})}\mathbb{E}_{\zeta''}\left[\mathbf{r}^\top\mathbf{X}_{\mathbf{H}}\check{\mathbf{Q}}_{-sk}\mathbf{X}_{\mathbf{H}}^\top\operatorname{diag}\left\{\frac{\mathbf{x}_{\mathbf{H}_i}^\top\check{\mathbf{Q}}_{-sk}^2\mathbf{x}_{\mathbf{H}_i}}{m\pi_i}\right\}_{i=1}^n\mathbf{r}\right]=0.$$

Recalling (22), we bound the term W_{10133}:

$$\begin{aligned}
W_{10133} &= \frac{m(m-1)}{m^2(1-\delta_{\zeta'''})}\mathbb{E}_{\zeta''}\left[\frac{\check{\mathbf{z}}_k^\top\check{\mathbf{Q}}_{-sk}^2\check{\mathbf{z}}_k\check{\mathbf{z}}_k^\top\check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_s\check{\mathbf{z}}_s^\top\check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_k\check{\mathbf{z}}_k^\top\check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_sF_{i_ki_k}^2\check{a}_s\check{a}_k}{m^3\check{\gamma}_{i_{s,k}}^2}\right]\\
&= O\left(\frac{\theta_{\max}^3(\mathbf{X})p^3}{m^3}\right)\cdot\|\mathbf{r}\|^2.
\end{aligned}$$

We thus conclude

$$W_{1013} = O\left(\frac{\theta_{\max}^2(\mathbf{X})p^2}{m^2}\right)\cdot\|\mathbf{r}\|^2.$$

Subsequently, we continue to bound the term W_{1011}. We rewrite

$$W_{1011} = \underbrace{-\frac{m(m-1)}{m^2(1-\delta_{\zeta'''})}\mathbb{E}_{\zeta''}\left[\frac{\check{\mathbf{z}}_k^\top \check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_s\check{\mathbf{z}}_s^\top \check{\mathbf{Q}}_{-sk}^2\check{\mathbf{z}}_k\check{\mathbf{z}}_k^\top \check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_s F_{i_k i_k}^2 \check{a}_s\check{a}_k}{m^2\check{\gamma}_{i_{s,k}}^2}\left(\frac{F_{i_s i_s}^2}{\check{\gamma}_{i_{k,s}}^2}-1\right)\right]}_{W_{10111}}$$
$$+\underbrace{\frac{m(m-1)}{m^2(1-\delta_{\zeta'''})}\mathbb{E}_{\zeta''}\left[\frac{\check{\mathbf{z}}_k^\top \check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_s\check{\mathbf{z}}_s^\top \check{\mathbf{Q}}_{-sk}^2\check{\mathbf{z}}_s F_{i_k i_k}\check{a}_s\check{a}_k}{m\check{\gamma}_{i_{s,k}}}\left(\frac{F_{i_s i_s}^2}{\check{\gamma}_{i_{k,s}}^2}-1\right)\right]}_{W_{10112}}$$
$$+\underbrace{\frac{m(m-1)}{m^2(1-\delta_{\zeta'''})}\mathbb{E}_{\zeta''}\left[\frac{\check{\mathbf{z}}_k^\top \check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_s\check{\mathbf{z}}_s^\top \check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_k\check{\mathbf{z}}_k^\top \check{\mathbf{Q}}_{-sk}^2\check{\mathbf{z}}_k\check{\mathbf{z}}_k^\top \check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_s F_{i_k i_k}^3 \check{a}_s\check{a}_k}{m^3\check{\gamma}_{i_{s,k}}^3}\left(\frac{F_{i_s i_s}^2}{\check{\gamma}_{i_{k,s}}^2}-1\right)\right]}_{W_{10113}}$$
$$\underbrace{-\frac{m(m-1)}{m^2(1-\delta_{\zeta'''})}\mathbb{E}_{\zeta''}\left[\frac{\check{\mathbf{z}}_k^\top \check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_s\check{\mathbf{z}}_s^\top \check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_k\check{\mathbf{z}}_k^\top \check{\mathbf{Q}}_{-sk}^2\check{\mathbf{z}}_s F_{i_k i_k}^2 \check{a}_s\check{a}_k}{m^2\check{\gamma}_{i_{s,k}}^2}\left(\frac{F_{i_s i_s}^2}{\check{\gamma}_{i_{k,s}}^2}-1\right)\right]}_{W_{10114}}.$$

Following (29) in Lemma E.1 and the bound on (22), we explore the terms W_{10111} and W_{10114}:

$$W_{10111} = W_{10114} = -\frac{m(m-1)}{m^2(1-\delta_{\zeta'''})}\mathbb{E}_{\zeta''}\left[\frac{\check{\mathbf{z}}_k^\top \check{\mathbf{Q}}_{-sk}^2\check{\mathbf{z}}_s\check{\mathbf{z}}_s^\top \check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_k\check{\mathbf{z}}_k^\top \check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_s F_{i_k i_k}^2 \check{a}_s\check{a}_k}{m^2\check{\gamma}_{i_{s,k}}^2}\left(\frac{F_{i_s i_s}^2}{\check{\gamma}_{i_{k,s}}^2}-1\right)\right]$$
$$= -\frac{m(m-1)}{m^2(1-\delta_{\zeta'''})}\sum_{i=1}^n \mathbb{E}_{\zeta''}\left[\frac{\check{\mathbf{z}}_k^\top \check{\mathbf{Q}}_{-sk}^2\mathbf{x}_{\mathbf{H}_i}\mathbf{x}_{\mathbf{H}_i}^\top \check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_k\check{\mathbf{z}}_k^\top \check{\mathbf{Q}}_{-sk}\mathbf{x}_{\mathbf{H}_i} F_{i_k i_k}^2 r_i\check{a}_k}{m^2\check{\gamma}_{i_{s,k}}^2\pi_i}\left(\frac{F_{ii}^2}{\bar{\gamma}_{i_{k,s,i}}^2}-1\right)\right]$$
$$\le \sum_{i=1}^n \mathbb{E}_{\zeta''}\left[\frac{\mathbf{x}_{\mathbf{H}_i}\check{\mathbf{Q}}_{-sk}^2\check{\mathbf{z}}_k\check{\mathbf{z}}_k^\top \check{\mathbf{Q}}_{-sk}^2\mathbf{x}_{\mathbf{H}_i}(\mathbf{x}_{\mathbf{H}_i}^\top \check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_k\check{\mathbf{z}}_k^\top \check{\mathbf{Q}}_{-sk}\mathbf{x}_{\mathbf{H}_i})^{1/2}F_{i_k i_k}^2 r_i^2}{m^2\check{\gamma}_{i_{s,k}}^2\pi_i^{3/2}}\left|\frac{F_{ii}^2}{\bar{\gamma}_{i_{k,s,i}}^2}-1\right|\right]$$
$$+\sum_{i=1}^n \mathbb{E}_{\zeta''}\left[\frac{F_{i_k i_k}^2\check{a}_k^2(\mathbf{x}_{\mathbf{H}_i}^\top \check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_k\check{\mathbf{z}}_k^\top \check{\mathbf{Q}}_{-sk}\mathbf{x}_{\mathbf{H}_i})^{3/2}}{m^2\check{\gamma}_{i_{s,k}}^2\pi_i^{1/2}}\left|\frac{F_{ii}^2}{\bar{\gamma}_{i_{k,s,i}}^2}-1\right|\right]$$
$$\le \sum_{i=1}^n \mathbb{E}_{\zeta''}\left[\frac{\mathbf{x}_{\mathbf{H}_i}\check{\mathbf{Q}}_{-sk}^2\check{\mathbf{z}}_k\check{\mathbf{z}}_k^\top \check{\mathbf{Q}}_{-sk}^2\mathbf{x}_{\mathbf{H}_i}(\mathbf{x}_{\mathbf{H}_i}^\top \check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_k\check{\mathbf{z}}_k^\top \check{\mathbf{Q}}_{-sk}\mathbf{x}_{\mathbf{H}_i})^{1/2}F_{i_k i_k}^2 r_i^2}{m^2\check{\gamma}_{i_{s,k}}^2\pi_i^{3/2}}\left(\frac{F_{ii}}{\bar{\gamma}_{i_{k,s,i}}}\left|\frac{F_{ii}}{\bar{\gamma}_{i_{k,s,i}}}-1\right|+\left|\frac{F_{ii}}{\bar{\gamma}_{i_{k,s,i}}}-1\right|\right)\right]$$
$$+\sum_{i=1}^n \mathbb{E}_{\zeta''}\left[\frac{F_{i_k i_k}^2\check{a}_k^2(\mathbf{x}_{\mathbf{H}_i}^\top \check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_k\check{\mathbf{z}}_k^\top \check{\mathbf{Q}}_{-sk}\mathbf{x}_{\mathbf{H}_i})^{3/2}}{m^2\check{\gamma}_{i_{s,k}}^2\pi_i^{1/2}}\left(\frac{F_{ii}}{\bar{\gamma}_{i_{k,s,i}}}\left|\frac{F_{ii}}{\bar{\gamma}_{i_{k,s,i}}}-1\right|+\left|\frac{F_{ii}}{\bar{\gamma}_{i_{k,s,i}}}-1\right|\right)\right]$$
$$\le 3\sum_{i=1}^n \mathbb{E}_{\zeta''}\left[\frac{\mathbf{x}_{\mathbf{H}_i}\check{\mathbf{Q}}_{-sk}^2\check{\mathbf{z}}_k\check{\mathbf{z}}_k^\top \check{\mathbf{Q}}_{-sk}^2\mathbf{x}_{\mathbf{H}_i}(\mathbf{x}_{\mathbf{H}_i}^\top \check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_k\check{\mathbf{z}}_k^\top \check{\mathbf{Q}}_{-sk}\mathbf{x}_{\mathbf{H}_i})^{1/2}F_{i_k i_k}^2 r_i^2}{m^2\check{\gamma}_{i_{s,k}}^2\pi_i^{3/2}}\left|\frac{F_{ii}}{\bar{\gamma}_{i_{k,s,i}}}-1\right|\right]$$
$$+3\sum_{i=1}^n \mathbb{E}_{\zeta''}\left[\frac{F_{i_k i_k}^2\check{a}_k^2(\mathbf{x}_{\mathbf{H}_i}^\top \check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_k\check{\mathbf{z}}_k^\top \check{\mathbf{Q}}_{-sk}\mathbf{x}_{\mathbf{H}_i})^{3/2}}{m^2\check{\gamma}_{i_{s,k}}^2\pi_i^{1/2}}\left|\frac{F_{ii}}{\bar{\gamma}_{i_{k,s,i}}}-1\right|\right]$$
$$\le \frac{24\theta_{\max}(\mathbf{X})p}{m}\sum_{i=1}^n \mathbb{E}_{\zeta''}\left[\frac{\mathbf{x}_{\mathbf{H}_i}\check{\mathbf{Q}}_{-sk}^2\check{\mathbf{z}}_k\check{\mathbf{z}}_k^\top \check{\mathbf{Q}}_{-sk}^2\mathbf{x}_{\mathbf{H}_i}F_{i_k i_k}^2 r_i^2}{m\pi_i}\left|\frac{F_{ii}}{\bar{\gamma}_{i_{k,s,i}}}-1\right|\right]$$
$$+\frac{24\theta_{\max}(\mathbf{X})p}{m}\sum_{i=1}^n \mathbb{E}_{\zeta''}\left[\frac{F_{i_k i_k}^2\check{a}_k^2\check{\mathbf{z}}_k^\top \check{\mathbf{Q}}_{-sk}\mathbf{x}_{\mathbf{H}_i}\mathbf{x}_{\mathbf{H}_i}^\top \check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_k}{m}\left|\frac{F_{ii}}{\bar{\gamma}_{i_{k,s,i}}}-1\right|\right]$$
$$\le \frac{96\theta_{\max}(\mathbf{X})p}{m}\sum_{i=1}^n\sum_{j=1}^n \mathbb{E}_{\zeta''}\left[\frac{\mathbf{x}_{\mathbf{H}_i}\check{\mathbf{Q}}_{-sk}^2\mathbf{x}_{\mathbf{H}_j}\mathbf{x}_{\mathbf{H}_j}^\top \check{\mathbf{Q}}_{-sk}^2\mathbf{x}_{\mathbf{H}_i}r_i^2}{m\pi_i}\left|\frac{F_{ii}}{\bar{\gamma}_{i_{k,s,i}}}-1\right|\right]$$

$$
\begin{aligned}
&+\frac{24\theta_{\max}(\mathbf{X})p}{m}\mathbb{E}_{\zeta''}\left[\frac{F_{i_ki_k}^2\check{a}_k^2\check{\mathbf{z}}_k^\top\check{\mathbf{Q}}_{-sk}\mathbf{X}_{\mathbf{H}}^\top\mathbf{X}_{\mathbf{H}}\check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_k}{m}\max_{1\le i\le n}\left|\frac{F_{ii}}{\bar{\gamma}_{i_{k,s,i}}}-1\right|\right]\\
&\le\frac{96\theta_{\max}(\mathbf{X})p}{m}\sum_{i=1}^n\mathbb{E}_{\zeta''}\left[\frac{\mathbf{x}_{\mathbf{H}_i}\check{\mathbf{Q}}_{-sk}^2\mathbf{X}_{\mathbf{H}}^\top\mathbf{X}_{\mathbf{H}}\check{\mathbf{Q}}_{-sk}^2\mathbf{x}_{\mathbf{H}_i}r_i^2}{m\pi_i}\left|\frac{F_{ii}}{\bar{\gamma}_{i_{k,s,i}}}-1\right|\right]\\
&+\frac{1536\theta_{\max}^2(\mathbf{X})p^2}{m^2}\mathbb{E}_{\zeta''}\left[F_{i_ki_k}^2\check{a}_k^2\max_{1\le i\le n}\left|\frac{F_{ii}}{\bar{\gamma}_{i_{k,s,i}}}-1\right|\right]\\
&\le\frac{12\times8^5\theta_{\max}^2(\mathbf{X})p^2\|\mathbf{r}\|^2}{m^2}\mathbb{E}_{\zeta''}\left[\max_{1\le i\le n}\left|\frac{F_{ii}}{\bar{\gamma}_{i_{k,s,i}}}-1\right|\right]+\frac{6144\theta_{\max}^2(\mathbf{X})p^2}{m^2}\mathbb{E}_{\zeta''}\left[\check{a}_k^2\max_{1\le i\le n}\left|\frac{F_{ii}}{\bar{\gamma}_{i_{k,s,i}}}-1\right|\right]\\
&\le\frac{12\times8^5\theta_{\max}^2(\mathbf{X})p^2\|\mathbf{r}\|^2}{m^2}\mathbb{E}_{\zeta''}\left[\max_{1\le i\le n}\left|\frac{F_{ii}}{\bar{\gamma}_{i_{k,s,i}}}-1\right|\right]+\frac{6144\theta_{\max}^2(\mathbf{X})p^2\|\mathbf{r}\|^2}{m^2}\mathbb{E}_{\zeta''}\left[\max_{1\le i\le n}\left|\frac{F_{ii}}{\bar{\gamma}_{i_{k,s,i}}}-1\right|\right]\\
&=O\left(\frac{\log n}{\log\log n}\sqrt{\frac{\theta_{\max}^7(\mathbf{X})p^7}{m^7}}\right)\cdot\|\mathbf{r}\|^2. \qquad (24)
\end{aligned}
$$

Similarly, we get

$$
\begin{aligned}
W_{10113}&=\frac{m(m-1)}{m^2(1-\delta_{\zeta'''})}\mathbb{E}_{\zeta''}\left[\frac{\check{\mathbf{z}}_k^\top\check{\mathbf{Q}}_{-sk}^2\check{\mathbf{z}}_k\check{\mathbf{z}}_k^\top\check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_s\check{\mathbf{z}}_s^\top\check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_k\check{\mathbf{z}}_k^\top\check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_sF_{i_ki_k}^3\check{a}_s\check{a}_k}{m^3\check{\gamma}_{i_{s,k}}^3}\left(\frac{F_{i_si_s}^2}{\check{\gamma}_{i_{k,s}}^2}-1\right)\right]\\
&=O\left(\frac{\log n}{\log\log n}\sqrt{\frac{\theta_{\max}^9(\mathbf{X})p^9}{m^9}}\right)\cdot\|\mathbf{r}\|^2.
\end{aligned}
$$

We then derive

$$
\begin{aligned}
W_{10112}&=\frac{m(m-1)}{m^2(1-\delta_{\zeta'''})}\mathbb{E}_{\zeta''}\left[\mathbf{r}^\top\operatorname{diag}\left\{\frac{F_{jj}}{\bar{\gamma}_{i_{s,k,j}}}\right\}_{j=1}^n\mathbf{X}_{\mathbf{H}}\check{\mathbf{Q}}_{-sk}\mathbf{X}_{\mathbf{H}}^\top\operatorname{diag}\left\{\frac{\mathbf{x}_{\mathbf{H}_i}^\top\check{\mathbf{Q}}_{-sk}\mathbf{x}_{\mathbf{H}_i}}{m\pi_i}\left(\frac{F_{ii}^2}{\bar{\gamma}_{i_{k,s,i}}^2}-1\right)\right\}_{i=1}^n\mathbf{r}\right]\\
&\le\frac{256\theta_{\max}(\mathbf{X})p\|\mathbf{r}\|^2}{m}\mathbb{E}_{\zeta''}\left[\max_{1\le i\le n}\left|\frac{F_{ii}^2}{\bar{\gamma}_{i_{k,s,i}}^2}-1\right|\right]\\
&\le\frac{256\theta_{\max}(\mathbf{X})p\|\mathbf{r}\|^2}{m}\left(\mathbb{E}_{\zeta''}\left[\max_{1\le i\le n}\frac{F_{ii}}{\bar{\gamma}_{i_{k,s,i}}}\left|\frac{F_{ii}}{\bar{\gamma}_{i_{k,s,i}}}-1\right|\right]+\mathbb{E}_{\zeta''}\left[\max_{1\le i\le n}\left|\frac{F_{ii}}{\bar{\gamma}_{i_{k,s,i}}}-1\right|\right]\right)\\
&\le\frac{768\theta_{\max}(\mathbf{X})p\|\mathbf{r}\|^2}{m}\mathbb{E}_{\zeta''}\left[\max_{1\le i\le n}\left|\frac{F_{ii}}{\bar{\gamma}_{i_{k,s,i}}}-1\right|\right]=O\left(\frac{\log n}{\log\log n}\sqrt{\frac{\theta_{\max}^5(\mathbf{X})p^5}{m^5}}\right)\cdot\|\mathbf{r}\|^2. \qquad (25)
\end{aligned}
$$

Thus, we get

$$
W_{1011}=O\left(\frac{\log n}{\log\log n}\sqrt{\frac{\theta_{\max}^5(\mathbf{X})p^5}{m^5}}\right)\cdot\|\mathbf{r}\|^2.
$$

Now, we turn to bound the term W_{1012}:

$$
\begin{aligned}
W_{1012}=&\underbrace{-\frac{m(m-1)}{m^2(1-\delta_{\zeta'''})}\mathbb{E}_{\zeta''}\left[\frac{\check{\mathbf{z}}_k^\top\check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_s\check{\mathbf{z}}_s^\top\check{\mathbf{Q}}_{-sk}^2\check{\mathbf{z}}_k\check{\mathbf{z}}_k^\top\check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_sF_{i_ki_k}\check{a}_s\check{a}_k}{m^2\check{\gamma}_{i_{s,k}}}\left(\frac{F_{i_ki_k}}{\check{\gamma}_{i_{s,k}}}-1\right)\right]}_{W_{10121}}\\
&+\underbrace{\frac{m(m-1)}{m^2(1-\delta_{\zeta'''})}\mathbb{E}_{\zeta''}\left[\frac{\check{\mathbf{z}}_k^\top\check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_s\check{\mathbf{z}}_s^\top\check{\mathbf{Q}}_{-sk}^2\check{\mathbf{z}}_s\check{a}_s\check{a}_k}{m}\left(\frac{F_{i_ki_k}}{\check{\gamma}_{i_{s,k}}}-1\right)\right]}_{W_{10122}}\\
&+\underbrace{\frac{m(m-1)}{m^2(1-\delta_{\zeta'''})}\mathbb{E}_{\zeta''}\left[\frac{\check{\mathbf{z}}_k^\top\check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_s\check{\mathbf{z}}_s^\top\check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_k\check{\mathbf{z}}_k^\top\check{\mathbf{Q}}_{-sk}^2\check{\mathbf{z}}_k\check{\mathbf{z}}_k^\top\check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_sF_{i_ki_k}^2\check{a}_s\check{a}_k}{m^3\check{\gamma}_{i_{s,k}}^2}\left(\frac{F_{i_ki_k}}{\check{\gamma}_{i_{s,k}}}-1\right)\right]}_{W_{10123}}
\end{aligned}
$$

$$\underbrace{-\frac{m(m-1)}{m^2(1-\delta_{\zeta'''})}\mathbb{E}_{\zeta''}\left[\frac{\check{\mathbf{z}}_k^\top\check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_s\check{\mathbf{z}}_s^\top\check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_k\check{\mathbf{z}}_k^\top\check{\mathbf{Q}}_{-sk}^2\check{\mathbf{z}}_s F_{i_ki_k}\check{a}_s\check{a}_k}{m^2\check{\gamma}_{i_{s,k}}}\left(\frac{F_{i_ki_k}}{\check{\gamma}_{i_{s,k}}}-1\right)\right]}_{W_{10124}}.$$

We begin by bounding the term W_{10121}. Using (29) in Lemma E.1 and (24), we get

$$\begin{aligned}
W_{10121} &= W_{10124} = \frac{m(m-1)}{m^2(1-\delta_{\zeta'''})}\mathbb{E}_{\zeta''}\left[\frac{\check{\mathbf{z}}_k^\top\check{\mathbf{Q}}_{-sk}^2\check{\mathbf{z}}_s\check{\mathbf{z}}_s^\top\check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_k\check{\mathbf{z}}_k^\top\check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_s F_{i_ki_k}\check{a}_s\check{a}_k}{m^2\check{\gamma}_{i_{s,k}}}\left(\frac{F_{i_ki_k}}{\check{\gamma}_{i_{s,k}}}-1\right)\right] \\
&= \frac{m(m-1)}{m^2(1-\delta_{\zeta'''})}\sum_{i=1}^n\mathbb{E}_{\zeta''}\left[\frac{\check{\mathbf{z}}_k^\top\check{\mathbf{Q}}_{-sk}^2\mathbf{x}_{\mathbf{H}_i}\mathbf{x}_{\mathbf{H}_i}^\top\check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_k\check{\mathbf{z}}_k^\top\check{\mathbf{Q}}_{-sk}\mathbf{x}_{\mathbf{H}_i}F_{i_ki_k}r_i\check{a}_k}{m^2\check{\gamma}_{i_{s,k}}\pi_i}\left(\frac{F_{i_ki_k}}{\check{\gamma}_{i_{s,k}}}-1\right)\right] \\
&\le \sum_{i=1}^n\mathbb{E}_{\zeta''}\left[\frac{\mathbf{x}_{\mathbf{H}_i}\check{\mathbf{Q}}_{-sk}^2\check{\mathbf{z}}_k\check{\mathbf{z}}_k^\top\check{\mathbf{Q}}_{-sk}^2\mathbf{x}_{\mathbf{H}_i}(\mathbf{x}_{\mathbf{H}_i}^\top\check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_k\check{\mathbf{z}}_k^\top\check{\mathbf{Q}}_{-sk}\mathbf{x}_{\mathbf{H}_i})^{1/2}F_{i_ki_k}r_i^2}{m^2\check{\gamma}_{i_{s,k}}\pi_i^{3/2}}\left|\frac{F_{i_ki_k}}{\check{\gamma}_{i_{s,k}}}-1\right|\right] \\
&+ \sum_{i=1}^n\mathbb{E}_{\zeta''}\left[\frac{F_{i_ki_k}\check{a}_k^2(\mathbf{x}_{\mathbf{H}_i}^\top\check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_k\check{\mathbf{z}}_k^\top\check{\mathbf{Q}}_{-sk}\mathbf{x}_{\mathbf{H}_i})^{3/2}}{m^2\check{\gamma}_{i_{s,k}}\pi_i^{1/2}}\left|\frac{F_{i_ki_k}}{\check{\gamma}_{i_{s,k}}}-1\right|\right] \\
&\le \frac{8\theta_{\max}(\mathbf{X})p}{m}\sum_{i=1}^n\mathbb{E}_{\zeta''}\left[\frac{\mathbf{x}_{\mathbf{H}_i}\check{\mathbf{Q}}_{-sk}^2\check{\mathbf{z}}_k\check{\mathbf{z}}_k^\top\check{\mathbf{Q}}_{-sk}^2\mathbf{x}_{\mathbf{H}_i}F_{i_ki_k}r_i^2}{m\pi_i}\left|\frac{F_{i_ki_k}}{\check{\gamma}_{i_{s,k}}}-1\right|\right] \\
&+ \frac{8\theta_{\max}(\mathbf{X})p}{m}\sum_{i=1}^n\mathbb{E}_{\zeta''}\left[\frac{F_{i_ki_k}\check{a}_k^2\check{\mathbf{z}}_k^\top\check{\mathbf{Q}}_{-sk}\mathbf{x}_{\mathbf{H}_i}\mathbf{x}_{\mathbf{H}_i}^\top\check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_k}{m}\left|\frac{F_{i_ki_k}}{\check{\gamma}_{i_{s,k}}}-1\right|\right] \\
&\le \frac{8\theta_{\max}(\mathbf{X})p}{m}\sum_{i=1}^n\sum_{j=1}^n\mathbb{E}_{\zeta''}\left[\frac{\mathbf{x}_{\mathbf{H}_i}\check{\mathbf{Q}}_{-sk}^2\mathbf{x}_{\mathbf{H}_j}\mathbf{x}_{\mathbf{H}_j}^\top\check{\mathbf{Q}}_{-sk}^2\mathbf{x}_{\mathbf{H}_i}F_{jj}r_i^2}{m\pi_i}\left|\frac{F_{jj}}{\bar{\gamma}_{i_{s,k,j}}}-1\right|\right] \\
&+ \frac{8\theta_{\max}(\mathbf{X})p}{m}\mathbb{E}_{\zeta''}\left[\frac{F_{i_ki_k}\check{a}_k^2\check{\mathbf{z}}_k^\top\check{\mathbf{Q}}_{-sk}\mathbf{X}_{\mathbf{H}}^\top\mathbf{X}_{\mathbf{H}}\check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_k}{m}\left|\frac{F_{i_ki_k}}{\check{\gamma}_{i_{s,k}}}-1\right|\right] \\
&\le \frac{16\theta_{\max}(\mathbf{X})p}{m}\sum_{i=1}^n\mathbb{E}_{\zeta''}\left[\frac{\mathbf{x}_{\mathbf{H}_i}\check{\mathbf{Q}}_{-sk}^2\mathbf{X}_{\mathbf{H}}^\top\mathbf{X}_{\mathbf{H}}\check{\mathbf{Q}}_{-sk}^2\mathbf{x}_{\mathbf{H}_i}r_i^2}{m\pi_i}\max_{1\le j\le n}\left|\frac{F_{jj}}{\bar{\gamma}_{i_{s,k,j}}}-1\right|\right] \\
&+ \frac{512\theta_{\max}^2(\mathbf{X})p^2}{m^2}\mathbb{E}_{\zeta''}\left[F_{i_ki_k}\check{a}_k^2\left|\frac{F_{i_ki_k}}{\check{\gamma}_{i_{s,k}}}-1\right|\right] \\
&\le \frac{2\times 8^5\theta_{\max}^2(\mathbf{X})p^2\|\mathbf{r}\|^2}{m^2}\mathbb{E}_{\zeta''}\left[\max_{1\le j\le n}\left|\frac{F_{jj}}{\bar{\gamma}_{i_{s,k,j}}}-1\right|\right]+\frac{512\theta_{\max}^2(\mathbf{X})p^2}{m^2}\mathbb{E}_{\zeta''}\left[F_{i_ki_k}\check{a}_k^2\left|\frac{F_{i_ki_k}}{\check{\gamma}_{i_{s,k}}}-1\right|\right] \\
&= O\left(\frac{\log n}{\operatorname{loglog} n}\sqrt{\frac{\theta_{\max}^7(\mathbf{X})p^7}{m^7}}\right)\|\mathbf{r}\|^2.
\end{aligned}$$

Analogously, we derive

$$\begin{aligned}
W_{10123} &= \frac{m(m-1)}{m^2(1-\delta_{\zeta'''})}\mathbb{E}_{\zeta''}\left[\frac{\check{\mathbf{z}}_k^\top\check{\mathbf{Q}}_{-sk}^2\check{\mathbf{z}}_k\check{\mathbf{z}}_k^\top\check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_s\check{\mathbf{z}}_s^\top\check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_k\check{\mathbf{z}}_k^\top\check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_s F_{i_ki_k}^2\check{a}_s\check{a}_k}{m^3\check{\gamma}_{i_{s,k}}^2}\left(\frac{F_{i_ki_k}}{\check{\gamma}_{i_{s,k}}}-1\right)\right] \\
&= O\left(\frac{\log n}{\operatorname{loglog} n}\sqrt{\frac{\theta_{\max}^9(\mathbf{X})p^9}{m^9}}\right)\cdot\|\mathbf{r}\|^2.
\end{aligned}$$

Recalling $\mathbf{X}_{\mathbf{H}}^\top\mathbf{r}=\mathbf{0}$, together with the bound on W_{10112} in (25), we attain

$$\begin{aligned}
W_{10122} &= \frac{m(m-1)}{m^2(1-\delta_{\zeta'''})}\mathbb{E}_{\zeta''}\left[\mathbf{r}^\top\operatorname{diag}\left\{\frac{F_{jj}}{\bar{\gamma}_{i_{s,k,j}}}-1\right\}_{j=1}^n\mathbf{X}_{\mathbf{H}}\check{\mathbf{Q}}_{-sk}\mathbf{X}_{\mathbf{H}}^\top\operatorname{diag}\left\{\frac{\mathbf{x}_{\mathbf{H}_i}^\top\check{\mathbf{Q}}_{-sk}\mathbf{x}_{\mathbf{H}_i}}{m\pi_i}\right\}_{i=1}^n\mathbf{r}\right] \\
&\le \frac{128\theta_{\max}(\mathbf{X})p\|\mathbf{r}\|^2}{m}\mathbb{E}_{\zeta''}\left[\max_{1\le j\le n}\left|\frac{F_{jj}}{\bar{\gamma}_{i_{s,k,j}}}-1\right|\right]=O\left(\frac{\log n}{\operatorname{loglog} n}\sqrt{\frac{\theta_{\max}^5(\mathbf{X})p^5}{m^5}}\right)\cdot\|\mathbf{r}\|^2.
\end{aligned}$$

Then, we get

$$W_{1012} = O\left(\frac{\log n}{\log\log n}\sqrt{\frac{\theta^5_{\max}(\mathbf{X})p^5}{m^5}}\right)\cdot\|\mathbf{r}\|^2.$$

Consequently, it follows that

$$W_{101} = O\left(\frac{\log n}{\log\log n}\cdot\frac{\theta^2_{\max}(\mathbf{X})p^2}{m^2}\right)\cdot\|\mathbf{r}\|^2.$$

Now, we proceed to explore the term W_{102}:

$$\begin{aligned}
W_{102} &= \frac{m(m-1)}{m^2(1-\delta_{\zeta'''})}\mathbb{E}_{\zeta''}\left[\frac{\check{\mathbf{z}}_k^\top\check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_s\check{\mathbf{z}}_s^\top\check{\mathbf{Q}}^2_{-s}\check{\mathbf{z}}_sF^2_{i_si_s}F_{i_ki_k}\check{a}_s\check{a}_k}{m\check{\gamma}_{i_{s,k}}}\left(\frac{1}{\check{\gamma}^2_{i_s}}-\frac{1}{\check{\gamma}^2_{i_{k,s}}}\right)\right]\\
&= \underbrace{\frac{m(m-1)}{m^2(1-\delta_{\zeta'''})}\mathbb{E}_{\zeta''}\left[\frac{2\check{\mathbf{z}}_k^\top\check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_s\check{\mathbf{z}}_s^\top\check{\mathbf{Q}}^2_{-s}\check{\mathbf{z}}_sF^3_{i_si_s}F^2_{i_ki_k}\check{a}_s\check{a}_k\check{\mathbf{z}}_s^\top\check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_k\check{\mathbf{z}}_k^\top\check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_s}{m^3\check{\gamma}^2_{i_{s,k}}\check{\gamma}^2_{i_s}\check{\gamma}_{i_{k,s}}}\right]}_{W_{1021}}\\
&\quad\underbrace{-\frac{m(m-1)}{m^2(1-\delta_{\zeta'''})}\mathbb{E}_{\zeta''}\left[\frac{\check{\mathbf{z}}_k^\top\check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_s\check{\mathbf{z}}_s^\top\check{\mathbf{Q}}^2_{-s}\check{\mathbf{z}}_sF^4_{i_si_s}F^3_{i_ki_k}\check{a}_s\check{a}_k(\check{\mathbf{z}}_s^\top\check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_k\check{\mathbf{z}}_k^\top\check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_s)^2}{m^5\check{\gamma}^3_{i_{s,k}}\check{\gamma}^2_{i_s}\check{\gamma}^2_{i_{k,s}}}\right]}_{W_{1022}}.
\end{aligned}$$

Considering the bound on W_{82} in (22), we have

$$\begin{aligned}
W_{1021} &= \frac{2m(m-1)}{m^2(1-\delta_{\zeta'''})}\sum_{i=1}^n\mathbb{E}_{\zeta''}\left[\frac{\check{\mathbf{z}}_k^\top\check{\mathbf{Q}}_{-sk}\mathbf{x}_{\mathbf{H}_i}\mathbf{x}_{\mathbf{H}_i}^\top\check{\mathbf{Q}}^2_{-s}\mathbf{x}_{\mathbf{H}_i}\mathbf{x}_{\mathbf{H}_i}^\top\check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_k\check{\mathbf{z}}_k^\top\check{\mathbf{Q}}_{-sk}\mathbf{x}_{\mathbf{H}_i}F^3_{ii}F^2_{i_ki_k}r_i\check{a}_k}{m^3\bar{\gamma}_{i_{k,s,i}}\check{\gamma}^2_{i_{s,k}}\bar{\gamma}_i^2\pi_i^2}\right]\\
&\le 2\sum_{i=1}^n\mathbb{E}_{\zeta''}\left[\frac{\mathbf{x}_{\mathbf{H}_i}\check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_k\check{\mathbf{z}}_k^\top\check{\mathbf{Q}}_{-sk}\mathbf{x}_{\mathbf{H}_i}\mathbf{x}_{\mathbf{H}_i}^\top\check{\mathbf{Q}}^2_{-s}\mathbf{x}_{\mathbf{H}_i}(\mathbf{x}_{\mathbf{H}_i}^\top\check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_k\check{\mathbf{z}}_k^\top\check{\mathbf{Q}}_{-sk}\mathbf{x}_{\mathbf{H}_i})^{1/2}F^3_{ii}F^2_{i_ki_k}r_i^2}{m^3\bar{\gamma}_{i_{k,s,i}}\check{\gamma}^2_{i_{s,k}}\bar{\gamma}_i^2\pi_i^{5/2}}\right]\\
&\quad+2\sum_{i=1}^n\mathbb{E}_{\zeta''}\left[\frac{F^3_{ii}F^2_{i_ki_k}\check{a}_k^2\mathbf{x}_{\mathbf{H}_i}^\top\check{\mathbf{Q}}^2_{-s}\mathbf{x}_{\mathbf{H}_i}(\mathbf{x}_{\mathbf{H}_i}^\top\check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_k\check{\mathbf{z}}_k^\top\check{\mathbf{Q}}_{-sk}\mathbf{x}_{\mathbf{H}_i})^{3/2}}{m^3\bar{\gamma}_{i_{k,s,i}}\check{\gamma}^2_{i_{s,k}}\bar{\gamma}_i^2\pi_i^{3/2}}\right]\\
&\le \frac{8^5\theta^2_{\max}(\mathbf{X})p^2}{m^2}\sum_{i=1}^n\mathbb{E}_{\zeta''}\left[\frac{\mathbf{x}_{\mathbf{H}_i}^\top\check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_k\check{\mathbf{z}}_k^\top\check{\mathbf{Q}}_{-sk}\mathbf{x}_{\mathbf{H}_i}r_i^2}{m\pi_i}\right]\\
&\quad+\frac{8^5\theta^2_{\max}(\mathbf{X})p^2}{m^2}\sum_{i=1}^n\mathbb{E}_{\zeta''}\left[\frac{\check{a}_k^2\check{\mathbf{z}}_k^\top\check{\mathbf{Q}}_{-sk}\mathbf{x}_{\mathbf{H}_i}\mathbf{x}_{\mathbf{H}_i}^\top\check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_k}{m}\right] = O\left(\frac{\theta^3_{\max}(\mathbf{X})p^3}{m^3}\right)\cdot\|\mathbf{r}\|^2.
\end{aligned}$$

Similarly, we bound the term W_{1022}:

$$\begin{aligned}
W_{1022} &= \frac{m(m-1)}{m^2(1-\delta_{\zeta'''})}\sum_{i=1}^n\mathbb{E}_{\zeta''}\left[\frac{\check{\mathbf{z}}_k^\top\check{\mathbf{Q}}_{-sk}\mathbf{x}_{\mathbf{H}_i}\mathbf{x}_{\mathbf{H}_i}^\top\check{\mathbf{Q}}^2_{-s}\mathbf{x}_{\mathbf{H}_i}(\mathbf{x}_{\mathbf{H}_i}^\top\check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_k\check{\mathbf{z}}_k^\top\check{\mathbf{Q}}_{-sk}\mathbf{x}_{\mathbf{H}_i})^2F^4_{ii}F^3_{i_ki_k}r_i\check{a}_k}{m^5\bar{\gamma}^2_{i_{k,s,i}}\check{\gamma}^3_{i_{s,k}}\bar{\gamma}_i^2\pi_i^3}\right]\\
&\le \sum_{i=1}^n\mathbb{E}_{\zeta''}\left[\frac{\mathbf{x}_{\mathbf{H}_i}^\top\check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_k\check{\mathbf{z}}_k^\top\check{\mathbf{Q}}_{-sk}\mathbf{x}_{\mathbf{H}_i}\mathbf{x}_{\mathbf{H}_i}^\top\check{\mathbf{Q}}^2_{-s}\mathbf{x}_{\mathbf{H}_i}(\mathbf{x}_{\mathbf{H}_i}^\top\check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_k\check{\mathbf{z}}_k^\top\check{\mathbf{Q}}_{-sk}\mathbf{x}_{\mathbf{H}_i})^{3/2}F^4_{ii}F^3_{i_ki_k}r_i^2}{m^5\bar{\gamma}^2_{i_{k,s,i}}\check{\gamma}^3_{i_{s,k}}\bar{\gamma}_i^2\pi_i^{7/2}}\right]\\
&\quad+\sum_{i=1}^n\mathbb{E}_{\zeta''}\left[\frac{F^4_{ii}F^3_{i_ki_k}\check{a}_k^2\mathbf{x}_{\mathbf{H}_i}^\top\check{\mathbf{Q}}^2_{-s}\mathbf{x}_{\mathbf{H}_i}(\mathbf{x}_{\mathbf{H}_i}^\top\check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_k\check{\mathbf{z}}_k^\top\check{\mathbf{Q}}_{-sk}\mathbf{x}_{\mathbf{H}_i})^{5/2}}{m^5\bar{\gamma}^2_{i_{k,s,i}}\check{\gamma}^3_{i_{s,k}}\bar{\gamma}_i^2\pi_i^{5/2}}\right]\\
&\le \frac{2\times8^7\theta^4_{\max}(\mathbf{X})p^4}{m^4}\sum_{i=1}^n\mathbb{E}_{\zeta''}\left[\frac{\mathbf{x}_{\mathbf{H}_i}^\top\check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_k\check{\mathbf{z}}_k^\top\check{\mathbf{Q}}_{-sk}\mathbf{x}_{\mathbf{H}_i}r_i^2}{m\pi_i}\right]\\
&\quad+\frac{2\times8^7\theta^4_{\max}(\mathbf{X})p^4}{m^4}\sum_{i=1}^n\mathbb{E}_{\zeta''}\left[\frac{\check{a}_k^2\check{\mathbf{z}}_k^\top\check{\mathbf{Q}}_{-sk}\mathbf{x}_{\mathbf{H}_i}\mathbf{x}_{\mathbf{H}_i}^\top\check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_k}{m}\right]
\end{aligned}$$

$$= O\left(\frac{\theta_{\max}^5(\mathbf{X})p^5}{m^5}\right)\cdot\|\mathbf{r}\|^2.$$

This yields

$$W_{102} = O\left(\frac{\theta_{\max}^3(\mathbf{X})p^3}{m^3}\right)\cdot\|\mathbf{r}\|^2.$$

We further conclude

$$W_{10} = O\left(\frac{\log n}{\log\log n}\cdot\frac{\theta_{\max}^2(\mathbf{X})p^2}{m^2}\right)\cdot\|\mathbf{r}\|^2.$$

Now, it remains to derive the bound of the term W_{11}. Recalling (20) and (21) again, we get

$$\begin{aligned}
&\mathbb{E}_{\zeta''}\left[\left|\frac{\check{\mathbf{z}}_k^\top\check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_s\check{\mathbf{z}}_s^\top\check{\mathbf{Q}}_{-s}^2\check{\mathbf{z}}_s F_{i_s i_s}^2 F_{i_k i_k}\check{a}_s\check{a}_k}{m\check{\gamma}_{i_s}^2\check{\gamma}_{i_{s,k}}}\right|\right] \le \frac{8^3\theta_{\max}(\mathbf{X})p}{m}\mathbb{E}_{\zeta''}\left[|\check{\mathbf{z}}_k^\top\check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_s\check{a}_s\check{a}_k|\right]\\
&\le \frac{8^3\theta_{\max}(\mathbf{X})p}{m}\sqrt{\mathbb{E}_{\zeta''}\left[\check{\mathbf{z}}_k^\top\check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_s\check{\mathbf{z}}_s^\top\check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_k\right]\mathbb{E}_{\zeta''}\left[\check{a}_s^2\check{a}_k^2\right]}\\
&= \frac{8^3\theta_{\max}(\mathbf{X})p\|\mathbf{r}\|^2}{m}\sqrt{\mathbb{E}_{\zeta''}\left[\check{\mathbf{z}}_k^\top\check{\mathbf{Q}}_{-sk}\mathbf{X}_{\mathbf{H}}^\top\mathbf{X}_{\mathbf{H}}\check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_k\right]}\\
&= \frac{8^3\theta_{\max}(\mathbf{X})p\|\mathbf{r}\|^2}{m}\sqrt{\sum_{j=1}^n\mathbb{E}_{\zeta''}\left[\mathbf{x}_{\mathbf{H}_j}^\top\check{\mathbf{Q}}_{-sk}\mathbf{X}_{\mathbf{H}}^\top\mathbf{X}_{\mathbf{H}}\check{\mathbf{Q}}_{-sk}\mathbf{x}_{\mathbf{H}_j}\right]} \le \frac{8^4\theta_{\max}(\mathbf{X})p^{3/2}\|\mathbf{r}\|^2}{m},
\end{aligned}$$

and

$$\begin{aligned}
&\mathrm{Var}_{\zeta''}\left[\frac{\check{\mathbf{z}}_k^\top\check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_s\check{\mathbf{z}}_s^\top\check{\mathbf{Q}}_{-s}^2\check{\mathbf{z}}_s F_{i_s i_s}^2 F_{i_k i_k}\check{a}_s\check{a}_k}{m\check{\gamma}_{i_s}^2\check{\gamma}_{i_{s,k}}}\right] \le \mathbb{E}_{\zeta''}\left[\frac{(\check{\mathbf{z}}_k^\top\check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_s\check{\mathbf{z}}_s^\top\check{\mathbf{Q}}_{-s}^2\check{\mathbf{z}}_s F_{i_s i_s}^2 F_{i_k i_k}\check{a}_s\check{a}_k)^2}{m^2\check{\gamma}_{i_s}^4\check{\gamma}_{i_{s,k}}^2}\right]\\
&\le \frac{8^6\theta_{\max}^2(\mathbf{X})p^2}{m^2}\mathbb{E}_{\zeta''}\left[\check{\mathbf{z}}_k^\top\check{\mathbf{Q}}_{-sk}\check{\mathbf{z}}_s\check{\mathbf{z}}_s^\top\check{\mathbf{Q}}_{-s}\check{\mathbf{z}}_k\check{a}_s^2\check{a}_k^2\right] \le \frac{8^8\theta_{\max}^4(\mathbf{X})p^4}{m^2}\|\mathbf{r}\|^4.
\end{aligned}$$

Following the bound on the term W_7, we similarly obtain, for $\delta_{\zeta'''}\le m^{-3}$,

$$W_{11} = O\left(\frac{1}{m}+\frac{\theta_{\max}^4(\mathbf{X})p^4}{m^4}\right)\cdot\|\mathbf{r}\|^2.$$

We thus complete the proof.

E.4 Proof of Corollary 4.5

The proof of Corollary 4.5 is based on the same proof strategy developed for Theorem 4.4. For completeness, we briefly outline the necessary modifications. The only substantive departure from the proof of Theorem 4.4 lies in the auxiliary estimates used to control several key terms. In the leverage score sampling setting, we rely on the refined bounds $\left\|\mathbb{E}_\zeta[\check{\mathbf{Q}}^2]-\mathbf{I}_p-\mathbf{X}_{\mathbf{H}}^\top\bar{\mathbf{F}}\mathbf{X}_{\mathbf{H}}\right\|$, $\|\mathbb{E}_\zeta[\check{\mathbf{Q}}^2]-\mathbb{E}_\zeta[\check{\mathbf{Q}}_{-s}^2]\|$, $\mathbb{E}_{\zeta'}\left[\max_{1\le i\le n}\left|\frac{F_{ii}}{\bar{\gamma}_i}-1\right|\right]$ and $\mathbb{E}_{\zeta''}\left[\max_{1\le i\le n}\left|\frac{F_{ii}}{\bar{\gamma}_{i_{k,s,i}}}-1\right|\right]$, which are established in Lemma E.3. Once these bounds are invoked, the remainder of the argument proceeds identically to that of Theorem 4.4.

E.5 Proof of Corollary 4.7

The proof of Corollary 4.7 follows the same overall strategy as that of Theorem 4.4. While the general structure of the argument remains unchanged, the technical distinctions arise from the auxiliary bounds employed. In the SRHT setting, the required estimates for $\left\|\mathbb{E}_\zeta[\check{\mathbf{Q}}^2]-\mathbf{I}_p-\mathbf{X}_{\mathbf{H}}^\top\bar{\mathbf{F}}\mathbf{X}_{\mathbf{H}}\right\|$, $\|\mathbb{E}_\zeta[\check{\mathbf{Q}}^2]-\mathbb{E}_\zeta[\check{\mathbf{Q}}_{-s}^2]\|$, $\mathbb{E}_{\zeta'}\left[\max_{1\le i\le n}\left|\frac{F_{ii}}{\bar{\gamma}_i}-1\right|\right]$ and $\mathbb{E}_{\zeta''}\left[\max_{1\le i\le n}\left|\frac{F_{ii}}{\bar{\gamma}_{i_{k,s,i}}}-1\right|\right]$, are taken from Lemma E.4. Aside from substituting these bounds, the remainder of the proof follows the same sequence of arguments as in Theorem 4.4.

E.6 Auxiliary results from [45]

Here, we recall several auxiliary lemmas from [45], which serve as key technical ingredients in the proof of Theorem 4.4.

We first present the following two lemmas, which are derived from Propositions 3.2 and F.1, along with their corresponding proofs given in Sections E.2 and F.1 of [45].

Lemma E.1 ([45]). *Under the settings and notations of Theorem 4.4, let $\mathbf{X_H} = \mathbf{X}\mathbf{H}^{-1/2}$. Then, there exists $C > 0$ independent of n, p so that for $m \geq C\theta_{\max}(\mathbf{X})p\log(p/\delta)$, $\delta \leq m^{-3}$, $\theta_{\max}(\mathbf{X})$ in Definition 2.3, when conditioned on an event ζ that holds with probability at least $1-\delta$, we have*

$$\left\|\mathbb{E}_\zeta[\check{\mathbf{Q}}^2] - \mathbf{I}_p - \mathbf{X}_\mathbf{H}^\top \bar{\mathbf{F}} \mathbf{X}_\mathbf{H}\right\| = O\left(\frac{\log n}{\operatorname{loglog} n}\sqrt{\frac{\theta_{\max}^3(\mathbf{X})p^3}{m^3}}\right), \tag{26}$$

$$\|\mathbb{E}_\zeta[\check{\mathbf{Q}}^2] - \mathbb{E}_\zeta[\check{\mathbf{Q}}_{-s}^2]\| = O\left(\frac{1}{m}\right), \tag{27}$$

$$\mathbb{E}_{\zeta'}\left[\max_{1\leq i\leq n}\left|\frac{F_{ii}}{\bar{\gamma}_i} - 1\right|\right] = O\left(\frac{\log n}{\operatorname{loglog} n}\sqrt{\frac{\theta_{\max}^3(\mathbf{X})p^3}{m^3}}\right), \tag{28}$$

$$\mathbb{E}_{\zeta''}\left[\max_{1\leq i\leq n}\left|\frac{F_{ii}}{\bar{\gamma}_{i_{k,s,i}}} - 1\right|\right] = O\left(\frac{\log n}{\operatorname{loglog} n}\sqrt{\frac{\theta_{\max}^3(\mathbf{X})p^3}{m^3}}\right). \tag{29}$$

Lemma E.2 ([45]). *Under the settings and notations of Theorem 4.4, suppose that the events ζ_1, ζ_2, ζ_3 are independent of $\mathbf{z}_s$. Define $\zeta^{'} = \bigcap_{j=1}^3 \zeta_j$ and $\delta_4 = \Pr(\neg\zeta_4)$. For a p.s.d. random matrix $\mathbf{K}$ (or any non-negative random variable) depending only on random sampling $\mathbf{S}$, we obtain,*

$$\mathbb{E}_\zeta[\mathbf{K}] = \frac{\mathbb{E}[(\prod_{j=1}^4 \mathbf{1}_{\zeta_j})\mathbf{K}]}{\Pr(\zeta)} \preceq \frac{1}{1-\delta}\mathbb{E}[\mathbf{1}_{\zeta'}\mathbf{K}] \preceq 2\mathbb{E}_{\zeta'}[\mathbf{K}], \tag{30}$$

where $\mathbf{1}_{\zeta_j}$ is the indicator of the event ζ_j.

We next state two lemmas specializing Lemma E.1 to the settings of leverage score sampling and the subsampled randomized Hadamard transform (SRHT). These results follow from Corollaries 3.4, 3.7, F.2, and F.3, and their proofs appear in Sections E.3.2 and E.3.4 of [45].

Lemma E.3 ([45]). *Under the settings and notations of Corollary 4.5, then, there exists $C > 0$ independent of n, p so that for $m \geq C\theta_{\max}(\mathbf{X})p\log(p/\delta)$, $\delta \leq m^{-3}$, $\theta_{\max}(\mathbf{X})$ in Definition 2.3, when conditioned on an event ζ that holds with probability at least $1-\delta$, we have*

$$\left\|\mathbb{E}_\zeta[\check{\mathbf{Q}}^2] - \mathbf{I}_p - \mathbf{X}_\mathbf{H}^\top \bar{\mathbf{F}} \mathbf{X}_\mathbf{H}\right\| = O\left(\frac{\log n}{\operatorname{loglog} n}\sqrt{\frac{\theta_{\max}^3(\mathbf{X})p^3}{m^3}} + \frac{\epsilon_\theta\theta_{\max}(\mathbf{X})p}{m}\right),$$

$$\|\mathbb{E}_\zeta[\check{\mathbf{Q}}^2] - \mathbb{E}_\zeta[\check{\mathbf{Q}}_{-s}^2]\| = O\left(\frac{1}{m}\right),$$

$$\mathbb{E}_{\zeta'}\left[\max_{1\leq i\leq n}\left|\frac{F_{ii}}{\bar{\gamma}_i} - 1\right|\right] = O\left(\frac{\log n}{\operatorname{loglog} n}\sqrt{\frac{\theta_{\max}^3(\mathbf{X})p^3}{m^3}} + \frac{\epsilon_\theta\theta_{\max}(\mathbf{X})p}{m}\right),$$

$$\mathbb{E}_{\zeta''}\left[\max_{1\leq i\leq n}\left|\frac{F_{ii}}{\bar{\gamma}_{i_{k,s,i}}} - 1\right|\right] = O\left(\frac{\log n}{\operatorname{loglog} n}\sqrt{\frac{\theta_{\max}^3(\mathbf{X})p^3}{m^3}} + \frac{\epsilon_\theta\theta_{\max}(\mathbf{X})p}{m}\right),$$

where $F_{ii} = m/(m-p)$.

Lemma E.4 ([45]). *Under the settings and notations of Corollary 4.7, there exists $C > 0$, $n\exp(-p) < \delta \leq m^{-3}$ such that for $m \geq Cp\log(p/\delta)$, when conditioned on the event ζ that holds with the probability at least $1-\delta$, we have*

$$\left\|\mathbb{E}_\zeta[\check{\mathbf{Q}}^2] - \mathbf{I}_p - \mathbf{X}_\mathbf{H}^\top \bar{\mathbf{F}} \mathbf{X}_\mathbf{H}\right\| = O\left(\frac{\log n}{\operatorname{loglog} n}\sqrt{\frac{p^3}{m^3}} + \frac{\sqrt{p\log(n/\delta)}}{m}\right),$$

$$\|\mathbb{E}_\zeta[\check{\mathbf{Q}}^2] - \mathbb{E}_\zeta[\check{\mathbf{Q}}_{-s}^2]\| = O\left(\frac{1}{m}\right),$$

$$\mathbb{E}_{\zeta'}\left[\max_{1\le i\le n}\left|\frac{F_{ii}}{\bar{\gamma}_i}-1\right|\right]=O\left(\frac{\log n}{\operatorname{loglog} n}\sqrt{\frac{p^3}{m^3}}+\frac{\sqrt{p\log(n/\delta)}}{m}\right),$$
$$\mathbb{E}_{\zeta''}\left[\max_{1\le i\le n}\left|\frac{F_{ii}}{\bar{\gamma}_{i_{k,s,i}}}-1\right|\right]=O\left(\frac{\log n}{\operatorname{loglog} n}\sqrt{\frac{p^3}{m^3}}+\frac{\sqrt{p\log(n/\delta)}}{m}\right),$$

where $F_{ii}=m/(m-p)$.

F Proof and discussion of Theorem 3.1

F.1 Detailed proof of Theorem 3.1

We begin by recalling some notations used in Appendix E. Let $\check{\mathbf{z}}_s^\top=\mathbf{e}_{i_s}^\top/\sqrt{\pi_{i_s}}\mathbf{X}_\mathbf{H}\in\mathbb{R}^p$ with $\mathbf{X}_\mathbf{H}=\mathbf{X}\mathbf{H}^{-1/2}$ and $\mathbf{H}=\mathbf{X}^\top\mathbf{X}$, and let

$$\check{\mathbf{Q}}=(\mathbf{X}_\mathbf{H}^\top\check{\mathbf{S}}^\top\check{\mathbf{S}}\mathbf{X}_\mathbf{H})^{-1}=\left(\sum_{s=1}^m\frac{1}{m}F_{i_si_s}\check{\mathbf{z}}_s\check{\mathbf{z}}_s^\top\right)^{-1},\ \ \check{\mathbf{Q}}_{-s}=\left(\sum_{k\neq s}\frac{1}{m}F_{i_k,i_k}\check{\mathbf{z}}_k\check{\mathbf{z}}_k^\top\right)^{-1}.$$

Note that for $\operatorname{rank}(\mathbf{X})=p$,

$$\mathbf{X}(\check{\mathbf{S}}\mathbf{X})^\dagger\check{\mathbf{S}}=\mathbf{X}(\mathbf{X}^\top\check{\mathbf{S}}^\top\check{\mathbf{S}}\mathbf{X})^{-1}\mathbf{X}^\top\check{\mathbf{S}}^\top\check{\mathbf{S}},\ \ \mathbf{X}\mathbf{X}^\dagger=\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top,$$

which implies

$$\begin{aligned}&\|\mathbb{E}_\zeta[\mathbf{X}(\check{\mathbf{S}}\mathbf{X})^\dagger\check{\mathbf{S}}]-\mathbf{X}\mathbf{X}^\dagger\|_F^2=\|\mathbf{H}^{\frac{1}{2}}\mathbb{E}_\zeta[(\mathbf{X}^\top\check{\mathbf{S}}^\top\check{\mathbf{S}}\mathbf{X})^{-1}\mathbf{X}^\top\check{\mathbf{S}}^\top\check{\mathbf{S}}\mathbf{P}_\perp]\|_F^2\\&=\|\mathbb{E}_\zeta[\check{\mathbf{Q}}\mathbf{X}_\mathbf{H}^\top\check{\mathbf{S}}^\top\check{\mathbf{S}}\mathbf{P}_\perp]\|_F^2=\sum_{l=1}^n\|\mathbb{E}_\zeta[\check{\mathbf{Q}}\mathbf{X}_\mathbf{H}^\top\check{\mathbf{S}}^\top\check{\mathbf{S}}\mathbf{r}_l]\|^2\end{aligned}\tag{31}$$

and

$$\begin{aligned}&\mathbf{E}_\zeta[\|\mathbf{X}(\check{\mathbf{S}}\mathbf{X})^\dagger\check{\mathbf{S}}-\mathbf{X}\mathbf{X}^\dagger\|_F^2]=\mathbf{E}_\zeta[\|\mathbf{H}^{\frac{1}{2}}(\mathbf{X}^\top\check{\mathbf{S}}^\top\check{\mathbf{S}}\mathbf{X})^{-1}\mathbf{X}^\top\check{\mathbf{S}}^\top\check{\mathbf{S}}\mathbf{P}_\perp\|_F^2]\\&=\mathbf{E}_\zeta[\|\check{\mathbf{Q}}\mathbf{X}_\mathbf{H}^\top\check{\mathbf{S}}^\top\check{\mathbf{S}}\mathbf{P}_\perp\|_F^2]=\sum_{l=1}^n\mathbb{E}_\zeta[\|\check{\mathbf{Q}}\mathbf{X}_\mathbf{H}^\top\check{\mathbf{S}}^\top\check{\mathbf{S}}\mathbf{r}_l\|^2]\end{aligned}$$

with $\mathbf{r}_l$ the lth column of $\mathbf{P}_\perp$. Following the same line of arguments as in the proof of Theorem 4.4, we obtain that, for each column $\mathbf{r}_l$,

$$\|\mathbb{E}_\zeta[\check{\mathbf{Q}}\mathbf{X}_\mathbf{H}^\top\check{\mathbf{S}}^\top\check{\mathbf{S}}\mathbf{r}_l]\|^2=\epsilon^2\|\mathbf{r}_l\|^2,$$

and

$$\mathbb{E}_\zeta[\|\check{\mathbf{Q}}\mathbf{X}_\mathbf{H}^\top\check{\mathbf{S}}^\top\check{\mathbf{S}}\mathbf{r}_l\|^2]=\mathbf{r}_l^\top\operatorname{diag}\left\{\frac{\ell_i(\mathbf{X})}{m\pi_i}\right\}_{i=1}^n\mathbf{r}_l+\epsilon\|\mathbf{r}_l\|_2^2.$$

This implies

$$\|\mathbb{E}_\zeta[\mathbf{X}(\check{\mathbf{S}}\mathbf{X})^\dagger\check{\mathbf{S}}]-\mathbf{X}\mathbf{X}^\dagger\|_F^2=\epsilon^2\sum_{l=1}^n\|\mathbf{r}_l\|^2=\epsilon^2\|\mathbf{P}_\perp\|_F^2,$$

and

$$\begin{aligned}\mathbb{E}_\zeta[\|\mathbf{X}(\check{\mathbf{S}}\mathbf{X})^\dagger\check{\mathbf{S}}-\mathbf{X}\mathbf{X}^\dagger\|_F^2]&=\sum_{l=1}^n\mathbf{r}_l^\top\operatorname{diag}\left\{\frac{\ell_i(\mathbf{X})}{m\pi_i}\right\}_{i=1}^n\mathbf{r}_l+\epsilon\sum_{l=1}^n\|\mathbf{r}_l\|_2^2\\&=\operatorname{tr}\left(\mathbf{P}_\perp\operatorname{diag}\left\{\frac{\ell_i(\mathbf{X})}{m\pi_i}\right\}_{i=1}^n\right)+\epsilon\|\mathbf{P}_\perp\|_F^2.\end{aligned}$$

This completes the proof.

F.2 Discussion on debiasing $\check{\mathbf{S}}$ with approximate leverage scores

Computing the exact leverage scores in (1) can be costly. The following corollary shows that replacing them with efficiently computable approximations incurs only a marginal increase in bias.

Corollary F.1. *Under the settings and notations of Theorem 3.1, if one uses approximate leverage scores $\check{\ell}_{i_s}(\mathbf{X})$ in the debiasing matrix, instead of the exact leverage scores $\ell_{i_s}(\mathbf{X})$ in* (1)*, that is*

$$\check{\mathbf{S}} = \operatorname{diag}\left\{1/\sqrt{1-\check{\ell}_{i_s}(\mathbf{X})/(m\pi_{i_s})}\right\}_{s=1}^{m}\cdot\mathbf{S},\ (1-\omega)\ell_{i_s}(\mathbf{X}) \le \check{\ell}_{i_s}(\mathbf{X}) \le (1+\omega)\ell_{i_s}(\mathbf{X}),\ \omega\in[0,1]. \tag{32}$$

Then, there exists $C>0$ independent of n,p so that for $m \ge C\theta_{\max}(\mathbf{X})p\log(p/\delta)$, $\delta \le n^{-(3+\alpha)}$, $\theta_{\max}(\mathbf{X})$ in Definition 2.3, when conditioned on an event ζ that holds with probability at least $1-\delta$, the debiased oblique projection $\check{\mathbf{P}} \equiv \mathbf{X}(\check{\mathbf{S}}\mathbf{X})^{\dagger}\check{\mathbf{S}}$ satisfies

$$\begin{aligned}
\left\|\mathbb{E}_{\zeta}[\check{\mathbf{P}}] - \mathbf{P}\right\|_F^2 &= \epsilon^2\cdot\|\mathbf{P}_{\perp}\|_F^2,\\
\mathbb{E}_{\zeta}\left[\|\check{\mathbf{P}}-\mathbf{P}\|_F^2\right] &= \operatorname{tr}\left(\mathbf{P}_{\perp}\operatorname{diag}\left\{\ell_i(\mathbf{X})/(m\pi_i)\right\}_{i=1}^{n}\right) + \epsilon\cdot\|\mathbf{P}_{\perp}\|_F^2,
\end{aligned}$$

with $\epsilon = O(\sqrt{(\log n/\operatorname{loglog} n)^2\cdot\theta_{\max}^3(\mathbf{X})p^3/m^3} + \omega\theta_{\max}(\mathbf{X})p/m)$, for orthogonal projection $\mathbf{P}=\mathbf{X}\mathbf{X}^{\dagger}$ and residual projection $\mathbf{P}_{\perp}=\mathbf{I}_n-\mathbf{X}\mathbf{X}^{\dagger}$.

Proof. We combine the arguments of Theorems 3.1 and 4.4. In these analyses, the exact leverage scores enter only through the following auxiliary estimates: $\left\|\mathbb{E}_{\zeta}[\check{\mathbf{Q}}^2]-\mathbf{I}_p-\mathbf{X}_{\mathbf{H}}^{\top}\bar{\mathbf{F}}\mathbf{X}_{\mathbf{H}}\right\|$, $\|\mathbb{E}_{\zeta}[\check{\mathbf{Q}}^2]-\mathbb{E}_{\zeta}[\check{\mathbf{Q}}^2_{-s}]\|$, $\mathbb{E}_{\zeta'}\left[\max_{1\le i\le n}\left|\frac{F_{ii}}{\bar{\gamma}_i}-1\right|\right]$ and $\mathbb{E}_{\zeta''}\left[\max_{1\le i\le n}\left|\frac{F_{ii}}{\bar{\gamma}_{i_{k,s,i}}}-1\right|\right]$. For the approximate score correction in (32), Lemma F.2 provides the corresponding perturbed bounds, introducing an additional error of order $O\left(\frac{\omega\theta_{\max}(\mathbf{X})p}{m}\right)$. Substituting these estimates into the leave-one-out and leave-two-out analyses underlying Theorem 4.4 yields, for each vector $\mathbf{r}_l$,

$$\begin{aligned}
\|\mathbb{E}_{\zeta}[\check{\mathbf{Q}}\mathbf{X}_{\mathbf{H}}^{\top}\check{\mathbf{S}}^{\top}\check{\mathbf{S}}\mathbf{r}_l]\|^2 &= \epsilon^2\|\mathbf{r}_l\|^2,\\
\mathbb{E}_{\zeta}[\|\check{\mathbf{Q}}\mathbf{X}_{\mathbf{H}}^{\top}\check{\mathbf{S}}^{\top}\check{\mathbf{S}}\mathbf{r}_l\|^2] &= \mathbf{r}_l^{\top}\operatorname{diag}\left\{\frac{\ell_i(\mathbf{X})}{m\pi_i}\right\}_{i=1}^{n}\mathbf{r}_l + \epsilon\|\mathbf{r}_l\|_2^2.
\end{aligned}$$

where $\epsilon = O(\sqrt{(\log n/\operatorname{loglog} n)^2\cdot\theta_{\max}^3(\mathbf{X})p^3/m^3} + \omega\theta_{\max}(\mathbf{X})p/m)$. Applying these estimates within the proof of Theorem 3.1 yields the claimed bounds. □

F.3 Auxiliary lemma from [45]

The following auxiliary lemma, corresponding to Corollary E.1 in Section E.3.1 of [45], provides a key technical ingredient for proving Corollary F.1.

Lemma F.2 ([45, Corollary E.1])**.** *Under the settings and notations of Theorem 3.1, if one uses approximate leverage scores $\check{\ell}_{i_s}(\mathbf{X})$ in the debiasing matrix, instead of the exact leverage scores $\ell_{i_s}(\mathbf{X})$ in* (1) *in Theorem 3.1, that is*

$$\check{\mathbf{S}} = \operatorname{diag}\left\{1/\sqrt{1-\check{\ell}_{i_s}(\mathbf{X})/(m\pi_{i_s})}\right\}_{s=1}^{m}\cdot\mathbf{S},\quad (1-\omega)\ell_{i_s}(\mathbf{X}) \le \check{\ell}_{i_s}^{\mathbf{C}}(\mathbf{X}) \le (1+\omega)\ell_{i_s}(\mathbf{X}),\quad \omega\in[0,1].$$

Then, there exists $C>0$ independent of n,p so that for $m \ge C\theta_{\max}(\mathbf{X})p\log(p/\delta)$, $\delta \le m^{-3}$, $\theta_{\max}(\mathbf{X})$ in Definition 2.3, when conditioned on an event ζ that holds with probability at least $1-\delta$, we have

$$\begin{aligned}
\left\|\mathbb{E}_{\zeta}[\check{\mathbf{Q}}^2]-\mathbf{I}_p-\mathbf{X}_{\mathbf{H}}^{\top}\bar{\mathbf{F}}\mathbf{X}_{\mathbf{H}}\right\| &= O\left(\frac{\log n}{\operatorname{loglog} n}\sqrt{\frac{\theta_{\max}^3(\mathbf{X})p^3}{m^3}} + \frac{\omega\theta_{\max}(\mathbf{X})p}{m}\right),\\
\|\mathbb{E}_{\zeta}[\check{\mathbf{Q}}^2]-\mathbb{E}_{\zeta}[\check{\mathbf{Q}}^2_{-s}]\| &= O\left(\frac{1}{m}\right),\\
\mathbb{E}_{\zeta'}\left[\max_{1\le i\le n}\left|\frac{F_{ii}}{\bar{\gamma}_i}-1\right|\right] &= O\left(\frac{\log n}{\operatorname{loglog} n}\sqrt{\frac{\theta_{\max}^3(\mathbf{X})p^3}{m^3}} + \frac{\omega\theta_{\max}(\mathbf{X})p}{m}\right),
\end{aligned}$$

$$\mathbb{E}_{\zeta''}\left[\max_{1\le i\le n}\left|\frac{F_{ii}}{\widetilde{\gamma}_{i_{k,s,i}}}-1\right|\right]=O\left(\frac{\log n}{\operatorname{loglog} n}\sqrt{\frac{\theta^3_{\max}(\mathbf{X})p^3}{m^3}}+\frac{\omega\theta_{\max}(\mathbf{X})p}{m}\right).$$

G The full bias-variance characterizations for fast CUR

In this section, we provide the full bias–variance characterizations in Theorem G.1, supplementing Theorem 4.8 with the corresponding variance result. We then prove Theorem G.1 in Appendix G.1, which also includes the proof of Theorem 4.8. Additional results related to Theorem G.1 are provided in Appendix G.2.

Before showing Theorem G.1, we first present Algorithm 1, which summarizes the fast CUR decomposition procedure and the computation of the standard sampled CUR estimator $\tilde{\mathbf{U}}$.

Algorithm 1 Fast CUR Decomposition [57]

Input: $\mathbf{X}\in\mathbb{R}^{n\times p}$, $\mathbf{C}\in\mathbb{R}^{n\times c}$ consisting of c columns of $\mathbf{X}$, $\mathbf{R}\in\mathbb{R}^{r\times p}$ consisting of r rows of $\mathbf{X}$, sample sizes m_c and m_r, sampling probabilities $\{\pi_{\mathbf{C},i}\}_{i=1}^n$ and $\{\pi_{\mathbf{R},j}\}_{j=1}^p$ with $\pi_{\mathbf{C},i}\ge 0$, $\sum_{i=1}^n\pi_{\mathbf{C},i}=1$ and $\pi_{\mathbf{R},j}\ge 0$, $\sum_{j=1}^p\pi_{\mathbf{R},j}=1$.

Output: $\tilde{\mathbf{U}}\in\mathbb{R}^{c\times r}$.

1. Draw sampling matrices $\mathbf{S_C}\in\mathbb{R}^{m_c\times n}$ with probabilities $\{\pi_{\mathbf{C},i}\}_{i=1}^n$ and $\mathbf{S_R}\in\mathbb{R}^{m_r\times p}$ with probabilities $\{\pi_{\mathbf{R},j}\}_{j=1}^p$ as in Definition 2.1;
2. Compute the sampled matrices $\mathbf{S_C C}$, $\mathbf{S_R R}^\top$, $\mathbf{S_C X S_R}^\top$;
3. Compute the solution of the sketched problem $\min_{\mathbf{U}}\|\mathbf{S_C}(\mathbf{CUR}-\mathbf{X})\mathbf{S_R}^\top\|_F^2$ as

$$\tilde{\mathbf{U}}=(\mathbf{S_C C})^\dagger\mathbf{S_C X S_R}^\top(\mathbf{R S_R}^\top)^\dagger. \tag{33}$$

In the following, we provide precise statistical characterizations of the proposed debiased solution $\check{\mathbf{U}}$.

Theorem G.1 (Precise characterizations for debiased fast CUR decomposition). *For a data matrix $\mathbf{X}\in\mathbb{R}^{n\times p}$, let $\mathbf{C}\in\mathbb{R}^{n\times c}$ and $\mathbf{R}\in\mathbb{R}^{r\times p}$ be the pre-selected columns and rows of $\mathbf{X}$, and let $\mathbf{U}_{\mathrm{CUR}}$ be defined as in* (10)*. Assume that $\min_{i,\|\mathbf{c}_i\|>0}\|\mathbf{c}_i\|^2/\|\mathbf{C}\|_F^2\ge n^{-\alpha_c}$ and $\min_{j,\|\mathbf{r}_j\|>0}\|\mathbf{r}_j\|^2/\|\mathbf{R}\|_F^2\ge p^{-\alpha_r}$ for some constants $\alpha_c>0$ and $\alpha_r>0$. For standard random sampling matrices $\mathbf{S_C}$ and $\mathbf{S_R}$ used in Algorithm 1, define their corresponding* debiased *sampling matrices $\check{\mathbf{S}}_{\mathbf{C}}\in\mathbb{R}^{m_c\times n}$ and $\check{\mathbf{S}}_{\mathbf{R}}\in\mathbb{R}^{m_r\times p}$ as in* (1)*. Then, there exists $C>0$ independent of n,p,c,r so that for $m_c\ge C\theta_{\max}(\mathbf{C})c\log(c/\delta)$ and $m_r\ge C\theta_{\max}(\mathbf{R}^\top)r\log(r/\delta)$, $\delta\le\min\{n^{-(3+\alpha_c)},p^{-(3+\alpha_r)}\}$, when conditioned on an event ζ that holds with probability at least $1-\delta$, the debiased matrix $\check{\mathbf{U}}$ defined in* (12) *satisfies*

$$\mathrm{Bias}_\zeta(\check{\mathbf{U}})=\epsilon^2\cdot L(\mathbf{U}_{\mathrm{CUR}}),\quad \mathrm{Var}_\zeta(\check{\mathbf{U}})=\Delta_{\mathrm{CUR}}+\epsilon\cdot L(\mathbf{U}_{\mathrm{CUR}}),$$

with $\epsilon=O(\sqrt{(\log n/\operatorname{loglog} n)^2\cdot\theta^3_{\max}(\mathbf{C})c^3/m_c^3}+\sqrt{(\log p/\operatorname{loglog} p)^2\cdot\theta^3_{\max}(\mathbf{R}^\top)r^3/m_r^3})$, for $L(\mathbf{U})\equiv\|\mathbf{CUR}-\mathbf{X}\|_F^2$, the associated bias $\mathrm{Bias}_\zeta(\check{\mathbf{U}})\equiv L(\mathbb{E}_\zeta[\check{\mathbf{U}}])-L(\mathbf{U}_{\mathrm{CUR}})$ and variance $\mathrm{Var}_\zeta(\check{\mathbf{U}})\equiv\mathbb{E}_\zeta[L(\check{\mathbf{U}})]-L(\mathbf{U}_{\mathrm{CUR}})$, in a spirit similar to those of subsampled OLS in Definition 4.1, where we define $\Delta_{\mathrm{CUR}}=(\Delta_{\mathrm{CUR},1}+\Delta_{\mathrm{CUR},2})^2+\sqrt{\epsilon\cdot L(\mathbf{U}_{\mathrm{CUR}})}(\Delta_{\mathrm{CUR},1}+\Delta_{\mathrm{CUR},2})$, with $\Delta_{\mathrm{CUR},1}=\|\operatorname{diag}\{\sqrt{\ell_i(\mathbf{C})/(m_c\pi_{\mathbf{C},i})}\}_{i=1}^n\mathbf{B}_1\|_F$ and $\Delta_{\mathrm{CUR},2}=\|\operatorname{diag}\{\sqrt{\ell_j(\mathbf{R}^\top)/(m_r\pi_{\mathbf{R},j})}\}_{j=1}^p\mathbf{B}_2\|_F$ for $\mathbf{B}_1=(\mathbf{I}_n-\mathbf{W_C}\mathbf{W_C}^\top)\mathbf{X}\in\mathbb{R}^{n\times p}$ and $\mathbf{B}_2=(\mathbf{I}_p-\mathbf{V_R}\mathbf{V_R}^\top)\mathbf{X}^\top\in\mathbb{R}^{p\times n}$, $\mathbf{W_C}\in\mathbb{R}^{n\times k_c}$ left singular-vector matrix of $\mathbf{C}$, and $\mathbf{V_R}\in\mathbb{R}^{p\times k_r}$ right singular-vector matrix of $\mathbf{R}$. Here, $\mathbf{c}_i^\top$ is the ith row of $\mathbf{C}$, $\mathbf{r}_j^\top$ is the jth row of $\mathbf{R}^\top$, $\theta_{\max}(\mathbf{C})$ and $\theta_{\max}(\mathbf{R}^\top)$ as in Definition 2.3 denote the maximum importance sampling approximation factors for $\mathbf{C}$ and $\mathbf{R}^\top$, respectively.

G.1 Detailed proof of Theorem G.1

We first recall some notations. Let $k_c=\operatorname{rank}(\mathbf{C})\le$ c and $k_r=\operatorname{rank}(\mathbf{R})\le$ r.

For $\mathbf{C}$, consider its singular value decomposition (SVD) $\mathbf{C}=\mathbf{W_C}\Sigma_{\mathbf{C}}\mathbf{V_C}^\top$, where $\mathbf{W_C}\in\mathbb{R}^{n\times k_c}$ and $\mathbf{V_C}\in\mathbb{R}^{c\times k_c}$ contain the left and right singular vectors of $\mathbf{C}$, respectively, and $\Sigma_{\mathbf{C}}=$

$\mathrm{diag}(\sigma_{\mathbf{C},1}, \ldots, \sigma_{\mathbf{C},k_c})$, where $\sigma_{\mathbf{C},1} \geq \sigma_{\mathbf{C},2} \geq \ldots \geq \sigma_{\mathbf{C},k_c}$ are its nonzero singular values. Similarly, the SVD of $\mathbf{R}$ is $\mathbf{R} = \mathbf{W}_\mathbf{R}\Sigma_\mathbf{R}\mathbf{V}_\mathbf{R}^\top$, where $\mathbf{W}_\mathbf{R} \in \mathbb{R}^{r\times k_r}$ and $\mathbf{V}_\mathbf{R} \in \mathbb{R}^{p\times k_r}$ contain the left and right singular vectors of $\mathbf{R}$, and $\Sigma_\mathbf{R} = \mathrm{diag}(\sigma_{\mathbf{R},1}, \ldots, \sigma_{\mathbf{R},k_r})$, where $\sigma_{\mathbf{R},1} \geq \sigma_{\mathbf{R},2} \geq \ldots \geq \sigma_{\mathbf{R},k_r}$ are its nonzero singular values. We rewrite

$$\mathbf{C}\mathbf{U}_{\mathrm{CUR}}\mathbf{R} = \mathbf{C}\mathbf{C}^\dagger\mathbf{X}\mathbf{R}^\dagger\mathbf{R} = \mathbf{W}_\mathbf{C}\mathbf{W}_\mathbf{C}^\top\mathbf{X}\mathbf{V}_\mathbf{R}\mathbf{V}_\mathbf{R}^\top \tag{34}$$

and

$$\mathbf{C}\check{\mathbf{U}}\mathbf{R} = \mathbf{W}_\mathbf{C}\Sigma_\mathbf{C}\mathbf{V}_\mathbf{C}^\top(\check{\mathbf{S}}_\mathbf{C}\mathbf{W}_\mathbf{C}\Sigma_\mathbf{C}\mathbf{V}_\mathbf{C}^\top)^\dagger\check{\mathbf{S}}_\mathbf{C}\mathbf{X}\check{\mathbf{S}}_\mathbf{R}^\top(\mathbf{W}_\mathbf{R}\Sigma_\mathbf{R}\mathbf{V}_\mathbf{R}^\top\check{\mathbf{S}}_\mathbf{R}^\top)^\dagger\mathbf{W}_\mathbf{R}\Sigma_\mathbf{R}\mathbf{V}_\mathbf{R}^\top.$$

Following the proof strategies developed in Theorems 4.4 and 3.1, the proof of Theorem G.1 proceeds in two main steps:

1. construct a high probability event ζ, based on subspace-embedding-type results from Lemma B.3, under which the matrices $(\mathbf{W}_\mathbf{C}^\top\check{\mathbf{S}}_\mathbf{C}^\top\check{\mathbf{S}}_\mathbf{C}\mathbf{W}_\mathbf{C})^{-1}$ and $(\mathbf{V}_\mathbf{R}^\top\check{\mathbf{S}}_\mathbf{R}^\top\check{\mathbf{S}}_\mathbf{R}\mathbf{V}_\mathbf{R})^{-1}$ are well conditioned; and
2. conditioning on that event ζ, bound the quantities $\mathrm{Bias}_\zeta(\check{\mathbf{U}})$ and $\mathrm{Var}_\zeta(\check{\mathbf{U}}) - \Delta_{\mathrm{CUR}}$ using a "leave-one-out" type analysis.

For random sampling matrix $\mathbf{S}_\mathbf{C}$ in Algorithm 1, denote $\mathbf{z}_{\mathbf{C},s_c}^\top = \mathbf{e}_{i_{s_c}}^\top/\sqrt{\pi_{\mathbf{C},i_{s_c}}}\mathbf{W}_\mathbf{C} \in \mathbb{R}^{k_c}$ the $i_{s_c}^{th}$ row of the sketch $\mathbf{S}_\mathbf{C}\mathbf{W}_\mathbf{C}$, where $\{\pi_{\mathbf{C},i}\}_{i=1}^n$ are the sampling probabilities of $\mathbf{S}_\mathbf{C}$, so that $\mathbb{E}[\mathbf{z}_{\mathbf{C},s_c}\mathbf{z}_{\mathbf{C},s_c}^\top] = \mathbf{W}_\mathbf{C}^\top\mathbf{W}_\mathbf{C}$. Similarly, for random sampling matrix $\mathbf{S}_\mathbf{R}$ in Algorithm 1, denote $\mathbf{z}_{\mathbf{R},s_r}^\top = \mathbf{e}_{i_{s_r}}^\top/\sqrt{\pi_{\mathbf{R},i_{s_r}}}\mathbf{V}_\mathbf{R} \in \mathbb{R}^{k_r}$ the $i_{s_r}^{th}$ row of the sketch $\mathbf{S}_\mathbf{R}\mathbf{V}_\mathbf{R}$, where $\{\pi_{\mathbf{R},i}\}_{i=1}^p$ are the sampling probabilities of $\mathbf{S}_\mathbf{R}$, so that $\mathbb{E}[\mathbf{z}_{\mathbf{R},s_r}\mathbf{z}_{\mathbf{R},s_r}^\top] = \mathbf{V}_\mathbf{R}^\top\mathbf{V}_\mathbf{R}$. Without loss of generality, assume that $t_c = m_c/4$ and $t_r = m_r/4$ are integers, and define the event $\zeta = \zeta_\mathbf{C} \bigcap \zeta_\mathbf{R}$, where $\zeta_\mathbf{C}$ depends only on $\mathbf{S}_\mathbf{C}$ and $\zeta_\mathbf{R}$ depends only on $\mathbf{S}_\mathbf{R}$, constructed as in (16). Specifically, for $\zeta_\mathbf{C}$, we set

$$\zeta_{\mathbf{C},j} : \sum_{s_c=t_c(j-1)+1}^{t_cj} \frac{1}{t_c}\mathbf{z}_{\mathbf{C},s_c}\mathbf{z}_{\mathbf{C},s_c}^\top \succeq \frac{1}{2}\mathbf{W}_\mathbf{C}^\top\mathbf{W}_\mathbf{C}, \quad j = 1,2,3,4, \quad \zeta_\mathbf{C} = \bigcap_{j=1}^4 \zeta_{\mathbf{C},j}, \tag{35}$$

and analogously, for $\zeta_\mathbf{R}$,

$$\zeta_{\mathbf{R},j} : \sum_{s_r=t_r(j-1)+1}^{t_rj} \frac{1}{t_r}\mathbf{z}_{\mathbf{R},s_r}\mathbf{z}_{\mathbf{R},s_r}^\top \succeq \frac{1}{2}\mathbf{V}_\mathbf{R}^\top\mathbf{V}_\mathbf{R}, \quad j = 1,2,3,4, \quad \zeta_\mathbf{R} = \bigcap_{j=1}^4 \zeta_{\mathbf{R},j}. \tag{36}$$

Thus, we have, on $\zeta_{\mathbf{C},j}$

$$\sum_{s_c=t_c(j-1)+1}^{t_cj} \frac{F_{i_{s_c}i_{s_c}}}{t_c}\mathbf{z}_{\mathbf{C},s_c}\mathbf{z}_{\mathbf{C},s_c}^\top \succeq \sum_{s_c=t_c(j-1)+1}^{t_cj} \frac{1}{t_c}\mathbf{z}_{\mathbf{C},s_c}\mathbf{z}_{\mathbf{C},s_c}^\top \succeq \frac{1}{2}\mathbf{W}_\mathbf{C}^\top\mathbf{W}_\mathbf{C}, \quad j = 1,2,3,4,$$

and similarly, on $\zeta_{\mathbf{R},j}$,

$$\sum_{s_r=t_r(j-1)+1}^{t_rj} \frac{F_{i_{s_r}i_{s_r}}}{t_r}\mathbf{z}_{\mathbf{R},s_r}\mathbf{z}_{\mathbf{R},s_r}^\top \succeq \sum_{s_r=t_r(j-1)+1}^{t_rj} \frac{1}{t_r}\mathbf{z}_{\mathbf{R},s_r}\mathbf{z}_{\mathbf{R},s_r}^\top \succeq \frac{1}{2}\mathbf{V}_\mathbf{R}^\top\mathbf{V}_\mathbf{R}, \quad j = 1,2,3,4.$$

Since random sampling matrices $\mathbf{S}_\mathbf{C}$ and $\mathbf{S}_\mathbf{R}$ are independent, the events $\zeta_{\mathbf{C},1}$, $\zeta_{\mathbf{C},2}$, $\zeta_{\mathbf{C},3}$, $\zeta_{\mathbf{C},4}$, $\zeta_{\mathbf{R},1}$, $\zeta_{\mathbf{R},2}$, $\zeta_{\mathbf{R},3}$, and $\zeta_{\mathbf{R},4}$ are independent.

Viewing $\mathbf{C}\check{\mathbf{U}}\mathbf{R} - \mathbf{X}$ as a two-sided subsampled least-squares residual, we analyze its left and right oblique projections sequentially. Conditioning on either sketch allows us to apply Theorems 3.1 and 4.4 to the oblique projection induced by the other, while their independence decouples the interaction term in expectation. The triangle and Cauchy–Schwarz inequalities then bound the column, row, and interaction terms by the CUR residual, yielding the stated bias and variance bounds.

In the following, we begin by bounding $\|\mathbb{E}_\zeta[\mathbf{C}\check{\mathbf{U}}\mathbf{R}] - \mathbf{X}\|_F^2 - \|\mathbf{C}\mathbf{U}_{\mathrm{CUR}}\mathbf{R} - \mathbf{X}\|_F^2$. Note that conditioned on the event ζ, $\check{\mathbf{S}}_\mathbf{C}\mathbf{W}_\mathbf{C}$ and $\mathbf{W}_\mathbf{R}\Sigma_\mathbf{R}$ have full column rank, and $\Sigma_\mathbf{C}\mathbf{V}_\mathbf{C}^\top$ and $\mathbf{V}_\mathbf{R}^\top\check{\mathbf{S}}_\mathbf{R}^\top$ have full row rank. Together with Lemma B.2, we further rewrite

$$\mathbf{C}\check{\mathbf{U}}\mathbf{R} = \mathbf{W}_\mathbf{C}\Sigma_\mathbf{C}\mathbf{V}_\mathbf{C}^\top(\Sigma_\mathbf{C}\mathbf{V}_\mathbf{C}^\top)^\dagger(\check{\mathbf{S}}_\mathbf{C}\mathbf{W}_\mathbf{C})^\dagger\check{\mathbf{S}}_\mathbf{C}\mathbf{X}\check{\mathbf{S}}_\mathbf{R}^\top(\mathbf{V}_\mathbf{R}^\top\check{\mathbf{S}}_\mathbf{R}^\top)^\dagger(\mathbf{W}_\mathbf{R}\Sigma_\mathbf{R})^\dagger\mathbf{W}_\mathbf{R}\Sigma_\mathbf{R}\mathbf{V}_\mathbf{R}^\top \tag{37}$$

$$= \mathbf{W_C}(\check{\mathbf{S}}_\mathbf{C}\mathbf{W_C})^\dagger\check{\mathbf{S}}_\mathbf{C}\mathbf{X}\check{\mathbf{S}}_\mathbf{R}^\top(\mathbf{V}_\mathbf{R}^\top\check{\mathbf{S}}_\mathbf{R}^\top)^\dagger\mathbf{V}_\mathbf{R}^\top. \tag{38}$$

Using (34) and (37), and expanding the squared Frobenius norm of the residual, we obtain

$$\begin{aligned}
&\|\mathbf{C}\mathbf{U}_{\mathrm{CUR}}\mathbf{R}-\mathbf{X}\|_F^2 = \mathrm{tr}(\mathbf{V_R}\mathbf{V}_\mathbf{R}^\top\mathbf{X}^\top\mathbf{W_C}\mathbf{W}_\mathbf{C}^\top\mathbf{W_C}\mathbf{W}_\mathbf{C}^\top\mathbf{X}\mathbf{V_R}\mathbf{V}_\mathbf{R}^\top) - \mathrm{tr}(\mathbf{V_R}\mathbf{V}_\mathbf{R}^\top\mathbf{X}^\top\mathbf{W_C}\mathbf{W}_\mathbf{C}^\top\mathbf{X})\\
&- \mathrm{tr}(\mathbf{X}^\top\mathbf{W_C}\mathbf{W}_\mathbf{C}^\top\mathbf{X}\mathbf{V_R}\mathbf{V}_\mathbf{R}^\top) + \mathrm{tr}(\mathbf{X}^\top\mathbf{X}) = \mathrm{tr}(\mathbf{V_R}\mathbf{V}_\mathbf{R}^\top\mathbf{X}^\top\mathbf{W_C}\mathbf{W}_\mathbf{C}^\top\mathbf{X}\mathbf{V_R}\mathbf{V}_\mathbf{R}^\top)\\
&- \mathrm{tr}(\mathbf{V_R}\mathbf{V}_\mathbf{R}^\top\mathbf{V_R}\mathbf{V}_\mathbf{R}^\top\mathbf{X}^\top\mathbf{W_C}\mathbf{W}_\mathbf{C}^\top\mathbf{X}) - \mathrm{tr}(\mathbf{X}^\top\mathbf{W_C}\mathbf{W}_\mathbf{C}^\top\mathbf{X}\mathbf{V_R}\mathbf{V}_\mathbf{R}^\top\mathbf{V_R}\mathbf{V}_\mathbf{R}^\top)\\
&+ \mathrm{tr}(\mathbf{X}^\top\mathbf{X}) = \mathrm{tr}(\mathbf{V_R}\mathbf{V}_\mathbf{R}^\top\mathbf{X}^\top\mathbf{W_C}\mathbf{W}_\mathbf{C}^\top\mathbf{X}\mathbf{V_R}\mathbf{V}_\mathbf{R}^\top) - \mathrm{tr}(\mathbf{V_R}\mathbf{V}_\mathbf{R}^\top\mathbf{X}^\top\mathbf{W_C}\mathbf{W}_\mathbf{C}^\top\mathbf{X}\mathbf{V_R}\mathbf{V}_\mathbf{R}^\top)\\
&- \mathrm{tr}(\mathbf{V_R}\mathbf{V}_\mathbf{R}^\top\mathbf{X}^\top\mathbf{W_C}\mathbf{W}_\mathbf{C}^\top\mathbf{X}\mathbf{V_R}\mathbf{V}_\mathbf{R}^\top) + \mathrm{tr}(\mathbf{X}^\top\mathbf{X}) = \mathrm{tr}(\mathbf{X}^\top\mathbf{X}) - \mathrm{tr}(\mathbf{V_R}\mathbf{V}_\mathbf{R}^\top\mathbf{X}^\top\mathbf{W_C}\mathbf{W}_\mathbf{C}^\top\mathbf{X}\mathbf{V_R}\mathbf{V}_\mathbf{R}^\top)\\
&= \mathrm{tr}(\mathbf{X}^\top\mathbf{X}) - \mathrm{tr}(\mathbf{V_R}\mathbf{V}_\mathbf{R}^\top\mathbf{X}^\top\mathbf{W_C}\mathbf{W}_\mathbf{C}^\top\mathbf{W_C}\mathbf{W}_\mathbf{C}^\top\mathbf{X}\mathbf{V_R}\mathbf{V}_\mathbf{R}^\top)
\end{aligned}$$

and

$$\begin{aligned}
&\|\mathbb{E}_\zeta[\mathbf{C}\check{\mathbf{U}}\mathbf{R}]-\mathbf{X}\|_F^2 = \mathrm{tr}(\mathbb{E}_\zeta[\mathbf{V_R}(\check{\mathbf{S}}_\mathbf{R}\mathbf{V_R})^\dagger\check{\mathbf{S}}_\mathbf{R}\mathbf{X}^\top\check{\mathbf{S}}_\mathbf{C}^\top(\mathbf{W}_\mathbf{C}^\top\check{\mathbf{S}}_\mathbf{C}^\top)^\dagger\mathbf{W}_\mathbf{C}^\top]\mathbb{E}_\zeta[\mathbf{W_C}(\check{\mathbf{S}}_\mathbf{C}\mathbf{W_C})^\dagger\check{\mathbf{S}}_\mathbf{C}\mathbf{X}\check{\mathbf{S}}_\mathbf{R}^\top(\mathbf{V}_\mathbf{R}^\top\check{\mathbf{S}}_\mathbf{R}^\top)^\dagger\mathbf{V}_\mathbf{R}^\top])\\
&- \mathrm{tr}((\mathbb{E}_\zeta[\mathbf{V_R}(\check{\mathbf{S}}_\mathbf{R}\mathbf{V_R})^\dagger\check{\mathbf{S}}_\mathbf{R}\mathbf{X}^\top\check{\mathbf{S}}_\mathbf{C}^\top(\mathbf{W}_\mathbf{C}^\top\check{\mathbf{S}}_\mathbf{C}^\top)^\dagger\mathbf{W}_\mathbf{C}^\top]\mathbf{X}) - \mathrm{tr}(\mathbf{X}^\top\mathbb{E}_\zeta[\mathbf{W_C}(\check{\mathbf{S}}_\mathbf{C}\mathbf{W_C})^\dagger\check{\mathbf{S}}_\mathbf{C}\mathbf{X}\check{\mathbf{S}}_\mathbf{R}^\top(\mathbf{V}_\mathbf{R}^\top\check{\mathbf{S}}_\mathbf{R}^\top)^\dagger\mathbf{V}_\mathbf{R}^\top])\\
&+ \mathrm{tr}(\mathbf{X}^\top\mathbf{X}) = \mathrm{tr}(\mathbf{V_R}\mathbb{E}_\zeta[(\check{\mathbf{S}}_\mathbf{R}\mathbf{V_R})^\dagger\check{\mathbf{S}}_\mathbf{R}\mathbf{X}^\top\check{\mathbf{S}}_\mathbf{C}^\top(\mathbf{W}_\mathbf{C}^\top\check{\mathbf{S}}_\mathbf{C}^\top)^\dagger]\mathbb{E}_\zeta[(\check{\mathbf{S}}_\mathbf{C}\mathbf{W_C})^\dagger\check{\mathbf{S}}_\mathbf{C}\mathbf{X}\check{\mathbf{S}}_\mathbf{R}^\top(\mathbf{V}_\mathbf{R}^\top\check{\mathbf{S}}_\mathbf{R}^\top)^\dagger]\mathbf{V}_\mathbf{R}^\top)\\
&- \mathrm{tr}(\mathbf{V_R}\mathbf{V}_\mathbf{R}^\top\mathbf{V_R}(\mathbb{E}_\zeta[(\check{\mathbf{S}}_\mathbf{R}\mathbf{V_R})^\dagger\check{\mathbf{S}}_\mathbf{R}\mathbf{X}^\top\check{\mathbf{S}}_\mathbf{C}^\top(\mathbf{W}_\mathbf{C}^\top\check{\mathbf{S}}_\mathbf{C}^\top)^\dagger\mathbf{W}_\mathbf{C}^\top]\mathbf{X})\\
&- \mathrm{tr}(\mathbf{X}^\top\mathbb{E}_\zeta[\mathbf{W_C}(\check{\mathbf{S}}_\mathbf{C}\mathbf{W_C})^\dagger\check{\mathbf{S}}_\mathbf{C}\mathbf{X}\check{\mathbf{S}}_\mathbf{R}^\top(\mathbf{V}_\mathbf{R}^\top\check{\mathbf{S}}_\mathbf{R}^\top)^\dagger]\mathbf{V}_\mathbf{R}^\top\mathbf{V_R}\mathbf{V}_\mathbf{R}^\top)\\
&+ \mathrm{tr}(\mathbf{X}^\top\mathbf{X}) = \mathrm{tr}(\mathbf{V_R}\mathbb{E}_\zeta[(\check{\mathbf{S}}_\mathbf{R}\mathbf{V_R})^\dagger\check{\mathbf{S}}_\mathbf{R}\mathbf{X}^\top\check{\mathbf{S}}_\mathbf{C}^\top(\mathbf{W}_\mathbf{C}^\top\check{\mathbf{S}}_\mathbf{C}^\top)^\dagger\mathbf{W}_\mathbf{C}^\top]\mathbb{E}_\zeta[\mathbf{W_C}(\check{\mathbf{S}}_\mathbf{C}\mathbf{W_C})^\dagger\check{\mathbf{S}}_\mathbf{C}\mathbf{X}\check{\mathbf{S}}_\mathbf{R}^\top(\mathbf{V}_\mathbf{R}^\top\check{\mathbf{S}}_\mathbf{R}^\top)^\dagger]\mathbf{V}_\mathbf{R}^\top)\\
&- \mathrm{tr}(\mathbf{V_R}\mathbb{E}_\zeta[(\check{\mathbf{S}}_\mathbf{R}\mathbf{V_R})^\dagger\check{\mathbf{S}}_\mathbf{R}\mathbf{X}^\top\check{\mathbf{S}}_\mathbf{C}^\top(\mathbf{W}_\mathbf{C}^\top\check{\mathbf{S}}_\mathbf{C}^\top)^\dagger]\mathbf{W}_\mathbf{C}^\top\mathbf{W_C}\mathbf{W}_\mathbf{C}^\top\mathbf{X}\mathbf{V_R}\mathbf{V}_\mathbf{R}^\top)\\
&- \mathrm{tr}(\mathbf{V_R}\mathbf{V}_\mathbf{R}^\top\mathbf{X}^\top\mathbf{W_C}\mathbf{W}_\mathbf{C}^\top\mathbf{W_C}\mathbb{E}_\zeta[(\check{\mathbf{S}}_\mathbf{C}\mathbf{W_C})^\dagger\check{\mathbf{S}}_\mathbf{C}\mathbf{X}\check{\mathbf{S}}_\mathbf{R}^\top(\mathbf{V}_\mathbf{R}^\top\check{\mathbf{S}}_\mathbf{R}^\top)^\dagger]\mathbf{V}_\mathbf{R}^\top) + \mathrm{tr}(\mathbf{X}^\top\mathbf{X}).
\end{aligned}$$

This further gives

$$\begin{aligned}
&\mathrm{Bias}_\zeta(\check{\mathbf{U}}) = \|\mathbb{E}_\zeta[\mathbf{C}\check{\mathbf{U}}\mathbf{R}]-\mathbf{X}\|_F^2 - \|\mathbf{C}\mathbf{U}_{\mathrm{CUR}}\mathbf{R}-\mathbf{X}\|_F^2\\
&= \|\mathbb{E}_\zeta[\mathbf{W_C}(\check{\mathbf{S}}_\mathbf{C}\mathbf{W_C})^\dagger\check{\mathbf{S}}_\mathbf{C}\mathbf{X}\check{\mathbf{S}}_\mathbf{R}^\top(\mathbf{V}_\mathbf{R}^\top\check{\mathbf{S}}_\mathbf{R}^\top)^\dagger\mathbf{V}_\mathbf{R}^\top] - \mathbf{W_C}\mathbf{W}_\mathbf{C}^\top\mathbf{X}\mathbf{V_R}\mathbf{V}_\mathbf{R}^\top\|_F^2\\
&= \|\mathbb{E}_\zeta[\mathbf{W_C}(\check{\mathbf{S}}_\mathbf{C}\mathbf{W_C})^\dagger\check{\mathbf{S}}_\mathbf{C}(\mathbf{I}_n-\mathbf{W_C}\mathbf{W}_\mathbf{C}^\top)\mathbf{X}(\mathbf{I}_p-\mathbf{V_R}\mathbf{V}_\mathbf{R}^\top)\check{\mathbf{S}}_\mathbf{R}^\top(\mathbf{V}_\mathbf{R}^\top\check{\mathbf{S}}_\mathbf{R}^\top)^\dagger\mathbf{V}_\mathbf{R}^\top]\\
&+ \mathbb{E}_\zeta[\mathbf{W_C}\mathbf{W}_\mathbf{C}^\top\mathbf{X}(\mathbf{I}_p-\mathbf{V_R}\mathbf{V}_\mathbf{R}^\top)\check{\mathbf{S}}_\mathbf{R}^\top(\mathbf{V}_\mathbf{R}^\top\check{\mathbf{S}}_\mathbf{R}^\top)^\dagger\mathbf{V}_\mathbf{R}^\top] + \mathbb{E}_\zeta[\mathbf{W_C}(\check{\mathbf{S}}_\mathbf{C}\mathbf{W_C})^\dagger\check{\mathbf{S}}_\mathbf{C}(\mathbf{I}_n-\mathbf{W_C}\mathbf{W}_\mathbf{C}^\top)\mathbf{X}\mathbf{V_R}\mathbf{V}_\mathbf{R}^\top]\|_F^2\\
&\overset{(a)}{\leq} \|\mathbb{E}_\zeta[\mathbf{W_C}(\check{\mathbf{S}}_\mathbf{C}\mathbf{W_C})^\dagger\check{\mathbf{S}}_\mathbf{C}(\mathbf{I}_n-\mathbf{W_C}\mathbf{W}_\mathbf{C}^\top)\mathbf{X}(\mathbf{I}_p-\mathbf{V_R}\mathbf{V}_\mathbf{R}^\top)\check{\mathbf{S}}_\mathbf{R}^\top(\mathbf{V}_\mathbf{R}^\top\check{\mathbf{S}}_\mathbf{R}^\top)^\dagger\mathbf{V}_\mathbf{R}^\top]\|_F^2\\
&+ \|\mathbb{E}_\zeta[\mathbf{W_C}\mathbf{W}_\mathbf{C}^\top\mathbf{X}(\mathbf{I}_p-\mathbf{V_R}\mathbf{V}_\mathbf{R}^\top)\check{\mathbf{S}}_\mathbf{R}^\top(\mathbf{V}_\mathbf{R}^\top\check{\mathbf{S}}_\mathbf{R}^\top)^\dagger\mathbf{V}_\mathbf{R}^\top]\|_F^2 + \|\mathbb{E}_\zeta[\mathbf{W_C}(\check{\mathbf{S}}_\mathbf{C}\mathbf{W_C})^\dagger\check{\mathbf{S}}_\mathbf{C}(\mathbf{I}_n-\mathbf{W_C}\mathbf{W}_\mathbf{C}^\top)\mathbf{X}\mathbf{V_R}\mathbf{V}_\mathbf{R}^\top]\|_F^2\\
&+ 2\|\mathbb{E}_\zeta[\mathbf{W_C}(\check{\mathbf{S}}_\mathbf{C}\mathbf{W_C})^\dagger\check{\mathbf{S}}_\mathbf{C}(\mathbf{I}_n-\mathbf{W_C}\mathbf{W}_\mathbf{C}^\top)\mathbf{X}(\mathbf{I}_p-\mathbf{V_R}\mathbf{V}_\mathbf{R}^\top)\check{\mathbf{S}}_\mathbf{R}^\top(\mathbf{V}_\mathbf{R}^\top\check{\mathbf{S}}_\mathbf{R}^\top)^\dagger\mathbf{V}_\mathbf{R}^\top]\|_F\\
&\cdot \|\mathbb{E}_\zeta[\mathbf{W_C}\mathbf{W}_\mathbf{C}^\top\mathbf{X}(\mathbf{I}_p-\mathbf{V_R}\mathbf{V}_\mathbf{R}^\top)\check{\mathbf{S}}_\mathbf{R}^\top(\mathbf{V}_\mathbf{R}^\top\check{\mathbf{S}}_\mathbf{R}^\top)^\dagger\mathbf{V}_\mathbf{R}^\top]\|_F\\
&+ 2\|\mathbb{E}_\zeta[\mathbf{W_C}(\check{\mathbf{S}}_\mathbf{C}\mathbf{W_C})^\dagger\check{\mathbf{S}}_\mathbf{C}(\mathbf{I}_n-\mathbf{W_C}\mathbf{W}_\mathbf{C}^\top)\mathbf{X}(\mathbf{I}_p-\mathbf{V_R}\mathbf{V}_\mathbf{R}^\top)\check{\mathbf{S}}_\mathbf{R}^\top(\mathbf{V}_\mathbf{R}^\top\check{\mathbf{S}}_\mathbf{R}^\top)^\dagger\mathbf{V}_\mathbf{R}^\top]\|_F\\
&\cdot \|\mathbb{E}_\zeta[\mathbf{W_C}(\check{\mathbf{S}}_\mathbf{C}\mathbf{W_C})^\dagger\check{\mathbf{S}}_\mathbf{C}(\mathbf{I}_n-\mathbf{W_C}\mathbf{W}_\mathbf{C}^\top)\mathbf{X}\mathbf{V_R}\mathbf{V}_\mathbf{R}^\top]\|_F\\
&+ 2\|\mathbb{E}_\zeta[\mathbf{W_C}\mathbf{W}_\mathbf{C}^\top\mathbf{X}(\mathbf{I}_p-\mathbf{V_R}\mathbf{V}_\mathbf{R}^\top)\check{\mathbf{S}}_\mathbf{R}^\top(\mathbf{V}_\mathbf{R}^\top\check{\mathbf{S}}_\mathbf{R}^\top)^\dagger\mathbf{V}_\mathbf{R}^\top]\|_F\|\mathbb{E}_\zeta[\mathbf{W_C}(\check{\mathbf{S}}_\mathbf{C}\mathbf{W_C})^\dagger\check{\mathbf{S}}_\mathbf{C}(\mathbf{I}_n-\mathbf{W_C}\mathbf{W}_\mathbf{C}^\top)\mathbf{X}\mathbf{V_R}\mathbf{V}_\mathbf{R}^\top]\|_F,
\end{aligned}$$

where the inequality (a) follows from the triangle inequality. By (15), (17), and (31), $\|\mathbb{E}_\zeta[\mathbf{W_C}(\check{\mathbf{S}}_\mathbf{C}\mathbf{W_C})^\dagger\check{\mathbf{S}}_\mathbf{C}(\mathbf{I}_n-\mathbf{W_C}\mathbf{W}_\mathbf{C}^\top)\mathbf{X}\mathbf{V_R}\mathbf{V}_\mathbf{R}^\top]\|_F^2$ corresponds to the bias of subsampled OLS with design matrix $\mathbf{W_C}$ and response matrix $\mathbf{X}\mathbf{V_R}\mathbf{V}_\mathbf{R}^\top$. Applying Theorems 3.1 and 4.4 thus gives

$$\begin{aligned}
&\|\mathbb{E}_\zeta[\mathbf{W_C}(\check{\mathbf{S}}_\mathbf{C}\mathbf{W_C})^\dagger\check{\mathbf{S}}_\mathbf{C}(\mathbf{I}_n-\mathbf{W_C}\mathbf{W}_\mathbf{C}^\top)\mathbf{X}\mathbf{V_R}\mathbf{V}_\mathbf{R}^\top]\|_F^2 = O\left(\frac{(\log n)^2\theta_{\max}^3(\mathbf{C})c^3}{(\log\log n)^2m_c^3}\right)\|(\mathbf{I}_n-\mathbf{W_C}\mathbf{W}_\mathbf{C}^\top)\mathbf{X}\mathbf{V_R}\mathbf{V}_\mathbf{R}^\top\|_F^2\\
&= O\left(\frac{(\log n)^2\theta_{\max}^3(\mathbf{C})c^3}{(\log\log n)^2m_c^3}\right)\|(\mathbf{I}_n-\mathbf{W_C}\mathbf{W}_\mathbf{C}^\top)\mathbf{X}\|_F^2\|\mathbf{V_R}\mathbf{V}_\mathbf{R}^\top\|^2 = O\left(\frac{(\log n)^2\theta_{\max}^3(\mathbf{C})c^3}{(\log\log n)^2m_c^3}\right)\|\mathbf{C}\mathbf{U}_{\mathrm{CUR}}\mathbf{R}-\mathbf{X}\|_F^2
\end{aligned}$$

and

$$\begin{aligned}
&\|\mathbb{E}_\zeta[\mathbf{W_C}\mathbf{W}_\mathbf{C}^\top\mathbf{X}(\mathbf{I}_p-\mathbf{V_R}\mathbf{V}_\mathbf{R}^\top)\check{\mathbf{S}}_\mathbf{R}^\top(\mathbf{V}_\mathbf{R}^\top\check{\mathbf{S}}_\mathbf{R}^\top)^\dagger\mathbf{V}_\mathbf{R}^\top]\|_F^2 = \|\mathbb{E}_\zeta[\mathbf{V_R}(\check{\mathbf{S}}_\mathbf{R}\mathbf{V_R})^\dagger\check{\mathbf{S}}_\mathbf{R}(\mathbf{I}_p-\mathbf{V_R}\mathbf{V}_\mathbf{R}^\top)\mathbf{X}^\top\mathbf{W_C}\mathbf{W}_\mathbf{C}^\top]\|_F^2\\
&= O\left(\frac{(\log p)^2\theta_{\max}^3(\mathbf{R}^\top)r^3}{(\log\log p)^2m_r^3}\right)\|(\mathbf{I}_p-\mathbf{V_R}\mathbf{V}_\mathbf{R}^\top)\mathbf{X}^\top\mathbf{W_C}\mathbf{W}_\mathbf{C}^\top\|_F^2
\end{aligned}$$

$$= O\left(\frac{(\log p)^2\theta^3_{\max}(\mathbf{R}^\top)r^3}{(\log\log p)^2 m_r^3}\right)\|\mathbf{W_C}\mathbf{W_C}^\top\mathbf{X}(\mathbf{I}_p - \mathbf{V_R}\mathbf{V_R}^\top)\|_F^2$$
$$= O\left(\frac{(\log p)^2\theta^3_{\max}(\mathbf{R}^\top)r^3}{(\log\log p)^2 m_r^3}\right)\|\mathbf{X}(\mathbf{I}_p - \mathbf{V_R}\mathbf{V_R}^\top)\|_F^2 = O\left(\frac{(\log p)^2\theta^3_{\max}(\mathbf{R}^\top)r^3}{(\log\log p)^2 m_r^3}\right)\|\mathbf{C}\mathbf{U}_{\mathrm{CUR}}\mathbf{R} - \mathbf{X}\|_F^2,$$

where we use the inequalities

$$\|(\mathbf{W_C}\mathbf{W_C}^\top - \mathbf{I}_n)\mathbf{X}\|_F^2 \le \|\mathbf{C}\mathbf{U}_{\mathrm{CUR}}\mathbf{R} - \mathbf{X}\|_F^2,\ \ \|\mathbf{X}(\mathbf{V_R}\mathbf{V_R}^\top - \mathbf{I}_p)\|_F^2 \le \|\mathbf{C}\mathbf{U}_{\mathrm{CUR}}\mathbf{R} - \mathbf{X}\|_F^2.$$

Taking Theorems 3.1 and 4.4 twice, we also similarly obtain

$$\|\mathbb{E}_\zeta[\mathbf{W_C}(\check{\mathbf{S}}_\mathbf{C}\mathbf{W_C})^\dagger\check{\mathbf{S}}_\mathbf{C}(\mathbf{I}_n - \mathbf{W_C}\mathbf{W_C}^\top)\mathbf{X}(\mathbf{I}_p - \mathbf{V_R}\mathbf{V_R}^\top)\check{\mathbf{S}}_\mathbf{R}^\top(\mathbf{V_R}^\top\check{\mathbf{S}}_\mathbf{R}^\top)^\dagger\mathbf{V_R}^\top]\|_F^2$$
$$= \|\mathbb{E}_{\zeta_\mathbf{C}}[\mathbf{W_C}(\check{\mathbf{S}}_\mathbf{C}\mathbf{W_C})^\dagger\check{\mathbf{S}}_\mathbf{C}]\mathbb{E}_{\zeta_\mathbf{R}}[(\mathbf{I}_n - \mathbf{W_C}\mathbf{W_C}^\top)\mathbf{X}(\mathbf{I}_p - \mathbf{V_R}\mathbf{V_R}^\top)\check{\mathbf{S}}_\mathbf{R}^\top(\mathbf{V_R}^\top\check{\mathbf{S}}_\mathbf{R}^\top)^\dagger\mathbf{V_R}^\top]\|_F^2$$
$$= O\left(\frac{(\log n)^2\theta^3_{\max}(\mathbf{C})c^3}{(\log\log n)^2 m_c^3}\right)\|\mathbb{E}_{\zeta_\mathbf{R}}[(\mathbf{I}_n - \mathbf{W_C}\mathbf{W_C}^\top)\mathbf{X}(\mathbf{I}_p - \mathbf{V_R}\mathbf{V_R}^\top)\check{\mathbf{S}}_\mathbf{R}^\top(\mathbf{V_R}^\top\check{\mathbf{S}}_\mathbf{R}^\top)^\dagger\mathbf{V_R}^\top]\|_F^2$$
$$= O\left(\frac{(\log n\log p)^2\theta^3_{\max}(\mathbf{C})\theta^3_{\max}(\mathbf{R}^\top)c^3r^3}{(\log\log n\log\log p)^2 m_c^3 m_r^3}\right)\|(\mathbf{I}_n - \mathbf{W_C}\mathbf{W_C}^\top)\mathbf{X}(\mathbf{I}_p - \mathbf{V_R}\mathbf{V_R}^\top)\|_F^2$$
$$= O\left(\frac{(\log n\log p)^2\theta^3_{\max}(\mathbf{C})\theta^3_{\max}(\mathbf{R}^\top)c^3r^3}{(\log\log n\log\log p)^2 m_c^3 m_r^3}\right)(\|\mathbf{X}(\mathbf{I}_p - \mathbf{V_R}\mathbf{V_R}^\top)\|_F^2 + \|\mathbf{W_C}\mathbf{W_C}^\top\mathbf{X}(\mathbf{I}_p - \mathbf{V_R}\mathbf{V_R}^\top)\|_F^2$$
$$+ 2\|\mathbf{W_C}\mathbf{W_C}^\top\mathbf{X}(\mathbf{I}_p - \mathbf{V_R}\mathbf{V_R}^\top)\|_F\|\mathbf{X}(\mathbf{I}_p - \mathbf{V_R}\mathbf{V_R}^\top)\|_F)$$
$$= O\left(\frac{(\log n\log p)^2\theta^3_{\max}(\mathbf{C})\theta^3_{\max}(\mathbf{R}^\top)c^3r^3}{(\log\log n\log\log p)^2 m_c^3 m_r^3}\right)\|\mathbf{C}\mathbf{U}_{\mathrm{CUR}}\mathbf{R} - \mathbf{X}\|_F^2,$$

where in the first equality we take $\zeta = \zeta_\mathbf{C} \cap \zeta_\mathbf{R}$, with $\zeta_\mathbf{C}$ and $\zeta_\mathbf{R}$ independent. Thus, we conclude

$$\mathrm{Bias}_\zeta(\check{\mathbf{U}}) = O\left(\frac{(\log n\log p)^2\theta^3_{\max}(\mathbf{C})\theta^3_{\max}(\mathbf{R}^\top)c^3r^3}{(\log\log n\log\log p)^2 m_c^3 m_r^3} + \frac{(\log n)^2\theta^3_{\max}(\mathbf{C})c^3}{(\log\log n)^2 m_c^3}\sqrt{\frac{(\log p)^2\theta^3_{\max}(\mathbf{R}^\top)r^3}{(\log\log p)^2 m_r^3}}\right.$$
$$+\frac{(\log p)^2\theta^3_{\max}(\mathbf{R}^\top)r^3}{(\log\log p)^2 m_r^3}\sqrt{\frac{(\log n)^2\theta^3_{\max}(\mathbf{C})c^3}{(\log\log n)^2 m_c^3}} + \frac{\log n\log p}{\log\log n\log\log p}\sqrt{\frac{\theta^3_{\max}(\mathbf{C})\theta^3_{\max}(\mathbf{R}^\top)c^3r^3}{m_c^3 m_r^3}}$$
$$\left.+\frac{(\log n)^2\theta^3_{\max}(\mathbf{C})c^3}{(\log\log n)^2 m_c^3} + \frac{(\log p)^2\theta^3_{\max}(\mathbf{R}^\top)r^3}{(\log\log p)^2 m_r^3}\right)\|\mathbf{C}\mathbf{U}_{\mathrm{CUR}}\mathbf{R} - \mathbf{X}\|_F^2.$$

Subsequently, we proceed to bound the other term $\mathrm{Var}_\zeta(\check{\mathbf{U}}) - \Delta_{\mathrm{CUR}}$. Following the decomposition used to bound $\mathrm{Bias}_\zeta(\check{\mathbf{U}})$, we first rewrite

$$\mathrm{Var}_\zeta(\check{\mathbf{U}}) = \mathbb{E}_\zeta[\|\mathbf{W_C}(\check{\mathbf{S}}_\mathbf{C}\mathbf{W_C})^\dagger\check{\mathbf{S}}_\mathbf{C}\mathbf{X}\check{\mathbf{S}}_\mathbf{R}^\top(\mathbf{V_R}^\top\check{\mathbf{S}}_\mathbf{R}^\top)^\dagger\mathbf{V_R}^\top - \mathbf{W_C}\mathbf{W_C}^\top\mathbf{X}\mathbf{V_R}\mathbf{V_R}^\top\|_F^2]$$
$$= \mathbb{E}_\zeta[\|\mathbf{W_C}(\check{\mathbf{S}}_\mathbf{C}\mathbf{W_C})^\dagger\check{\mathbf{S}}_\mathbf{C}(\mathbf{I}_n - \mathbf{W_C}\mathbf{W_C}^\top)\mathbf{X}(\mathbf{I}_p - \mathbf{V_R}\mathbf{V_R}^\top)\check{\mathbf{S}}_\mathbf{R}^\top(\mathbf{V_R}^\top\check{\mathbf{S}}_\mathbf{R}^\top)^\dagger\mathbf{V_R}^\top$$
$$+ \mathbf{W_C}\mathbf{W_C}^\top\mathbf{X}(\mathbf{I}_p - \mathbf{V_R}\mathbf{V_R}^\top)\check{\mathbf{S}}_\mathbf{R}^\top(\mathbf{V_R}^\top\check{\mathbf{S}}_\mathbf{R}^\top)^\dagger\mathbf{V_R}^\top + \mathbf{W_C}(\check{\mathbf{S}}_\mathbf{C}\mathbf{W_C})^\dagger\check{\mathbf{S}}_\mathbf{C}(\mathbf{I}_n - \mathbf{W_C}\mathbf{W_C}^\top)\mathbf{X}\mathbf{V_R}\mathbf{V_R}^\top\|_F^2]$$
$$\le \mathbb{E}_\zeta[\|\mathbf{W_C}(\check{\mathbf{S}}_\mathbf{C}\mathbf{W_C})^\dagger\check{\mathbf{S}}_\mathbf{C}(\mathbf{I}_n - \mathbf{W_C}\mathbf{W_C}^\top)\mathbf{X}(\mathbf{I}_p - \mathbf{V_R}\mathbf{V_R}^\top)\check{\mathbf{S}}_\mathbf{R}^\top(\mathbf{V_R}^\top\check{\mathbf{S}}_\mathbf{R}^\top)^\dagger\mathbf{V_R}^\top\|_F^2]$$
$$+ \mathbb{E}_\zeta[\|\mathbf{W_C}\mathbf{W_C}^\top\mathbf{X}(\mathbf{I}_p - \mathbf{V_R}\mathbf{V_R}^\top)\check{\mathbf{S}}_\mathbf{R}^\top(\mathbf{V_R}^\top\check{\mathbf{S}}_\mathbf{R}^\top)^\dagger\mathbf{V_R}^\top\|_F^2] + \mathbb{E}_\zeta[\|\mathbf{W_C}(\check{\mathbf{S}}_\mathbf{C}\mathbf{W_C})^\dagger\check{\mathbf{S}}_\mathbf{C}(\mathbf{I}_n - \mathbf{W_C}\mathbf{W_C}^\top)\mathbf{X}\mathbf{V_R}\mathbf{V_R}^\top\|_F^2]$$
$$+ 2\mathbb{E}_\zeta[\|\mathbf{W_C}(\check{\mathbf{S}}_\mathbf{C}\mathbf{W_C})^\dagger\check{\mathbf{S}}_\mathbf{C}(\mathbf{I}_n - \mathbf{W_C}\mathbf{W_C}^\top)\mathbf{X}(\mathbf{I}_p - \mathbf{V_R}\mathbf{V_R}^\top)\check{\mathbf{S}}_\mathbf{R}^\top(\mathbf{V_R}^\top\check{\mathbf{S}}_\mathbf{R}^\top)^\dagger\mathbf{V_R}^\top\|_F$$
$$\cdot\|\mathbf{W_C}\mathbf{W_C}^\top\mathbf{X}(\mathbf{I}_p - \mathbf{V_R}\mathbf{V_R}^\top)\check{\mathbf{S}}_\mathbf{R}^\top(\mathbf{V_R}^\top\check{\mathbf{S}}_\mathbf{R}^\top)^\dagger\mathbf{V_R}^\top\|_F]$$
$$+ 2\mathbb{E}_\zeta[\|\mathbf{W_C}(\check{\mathbf{S}}_\mathbf{C}\mathbf{W_C})^\dagger\check{\mathbf{S}}_\mathbf{C}(\mathbf{I}_n - \mathbf{W_C}\mathbf{W_C}^\top)\mathbf{X}(\mathbf{I}_p - \mathbf{V_R}\mathbf{V_R}^\top)\check{\mathbf{S}}_\mathbf{R}^\top(\mathbf{V_R}^\top\check{\mathbf{S}}_\mathbf{R}^\top)^\dagger\mathbf{V_R}^\top\|_F$$
$$\cdot\|\mathbf{W_C}(\check{\mathbf{S}}_\mathbf{C}\mathbf{W_C})^\dagger\check{\mathbf{S}}_\mathbf{C}(\mathbf{I}_n - \mathbf{W_C}\mathbf{W_C}^\top)\mathbf{X}\mathbf{V_R}\mathbf{V_R}^\top\|_F]$$
$$+ 2\mathbb{E}_\zeta[\|\mathbf{W_C}\mathbf{W_C}^\top\mathbf{X}(\mathbf{I}_p - \mathbf{V_R}\mathbf{V_R}^\top)\check{\mathbf{S}}_\mathbf{R}^\top(\mathbf{V_R}^\top\check{\mathbf{S}}_\mathbf{R}^\top)^\dagger\mathbf{V_R}^\top\|_F\|\mathbf{W_C}(\check{\mathbf{S}}_\mathbf{C}\mathbf{W_C})^\dagger\check{\mathbf{S}}_\mathbf{C}(\mathbf{I}_n - \mathbf{W_C}\mathbf{W_C}^\top)\mathbf{X}\mathbf{V_R}\mathbf{V_R}^\top\|_F]$$
$$\overset{(b)}{\le} \mathbb{E}_\zeta[\|\mathbf{W_C}(\check{\mathbf{S}}_\mathbf{C}\mathbf{W_C})^\dagger\check{\mathbf{S}}_\mathbf{C}(\mathbf{I}_n - \mathbf{W_C}\mathbf{W_C}^\top)\mathbf{X}(\mathbf{I}_p - \mathbf{V_R}\mathbf{V_R}^\top)\check{\mathbf{S}}_\mathbf{R}^\top(\mathbf{V_R}^\top\check{\mathbf{S}}_\mathbf{R}^\top)^\dagger\mathbf{V_R}^\top\|_F^2]$$
$$+ \mathbb{E}_\zeta[\|\mathbf{W_C}\mathbf{W_C}^\top\mathbf{X}(\mathbf{I}_p - \mathbf{V_R}\mathbf{V_R}^\top)\check{\mathbf{S}}_\mathbf{R}^\top(\mathbf{V_R}^\top\check{\mathbf{S}}_\mathbf{R}^\top)^\dagger\mathbf{V_R}^\top\|_F^2] + \mathbb{E}_\zeta[\|\mathbf{W_C}(\check{\mathbf{S}}_\mathbf{C}\mathbf{W_C})^\dagger\check{\mathbf{S}}_\mathbf{C}(\mathbf{I}_n - \mathbf{W_C}\mathbf{W_C}^\top)\mathbf{X}\mathbf{V_R}\mathbf{V_R}^\top\|_F^2]$$
$$+ 2\sqrt{\mathbb{E}_\zeta[\|\mathbf{W_C}(\check{\mathbf{S}}_\mathbf{C}\mathbf{W_C})^\dagger\check{\mathbf{S}}_\mathbf{C}(\mathbf{I}_n - \mathbf{W_C}\mathbf{W_C}^\top)\mathbf{X}(\mathbf{I}_p - \mathbf{V_R}\mathbf{V_R}^\top)\check{\mathbf{S}}_\mathbf{R}^\top(\mathbf{V_R}^\top\check{\mathbf{S}}_\mathbf{R}^\top)^\dagger\mathbf{V_R}^\top\|_F^2]}$$

$$\cdot \sqrt{\mathbb{E}_\zeta[\|\mathbf{W_C}\mathbf{W_C}^\top\mathbf{X}(\mathbf{I}_p - \mathbf{V_R}\mathbf{V_R}^\top)\check{\mathbf{S}}_\mathbf{R}^\top(\mathbf{V_R}^\top\check{\mathbf{S}}_\mathbf{R}^\top)^\dagger\mathbf{V_R}^\top\|_F^2]}$$
$$+ 2\sqrt{\mathbb{E}_\zeta[\|\mathbf{W_C}(\check{\mathbf{S}}_\mathbf{C}\mathbf{W_C})^\dagger\check{\mathbf{S}}_\mathbf{C}(\mathbf{I}_n - \mathbf{W_C}\mathbf{W_C}^\top)\mathbf{X}(\mathbf{I}_p - \mathbf{V_R}\mathbf{V_R}^\top)\check{\mathbf{S}}_\mathbf{R}^\top(\mathbf{V_R}^\top\check{\mathbf{S}}_\mathbf{R}^\top)^\dagger\mathbf{V_R}^\top\|_F^2]}$$
$$\cdot \sqrt{\mathbb{E}_\zeta[\|\mathbf{W_C}(\check{\mathbf{S}}_\mathbf{C}\mathbf{W_C})^\dagger\check{\mathbf{S}}_\mathbf{C}(\mathbf{I}_n - \mathbf{W_C}\mathbf{W_C}^\top)\mathbf{X}\mathbf{V_R}\mathbf{V_R}^\top\|_F^2]}$$
$$+ 2\sqrt{\mathbb{E}_\zeta[\|\mathbf{W_C}\mathbf{W_C}^\top\mathbf{X}(\mathbf{I}_p - \mathbf{V_R}\mathbf{V_R}^\top)\check{\mathbf{S}}_\mathbf{R}^\top(\mathbf{V_R}^\top\check{\mathbf{S}}_\mathbf{R}^\top)^\dagger\mathbf{V_R}^\top\|_F^2]}$$
$$\cdot \sqrt{\mathbb{E}_\zeta[\|\mathbf{W_C}(\check{\mathbf{S}}_\mathbf{C}\mathbf{W_C})^\dagger\check{\mathbf{S}}_\mathbf{C}(\mathbf{I}_n - \mathbf{W_C}\mathbf{W_C}^\top)\mathbf{X}\mathbf{V_R}\mathbf{V_R}^\top\|_F^2]},$$

where in inequality (b) we consider the Cauchy–Schwarz inequality. Using Theorems 3.1 and 4.4, it follows that

$$\mathbb{E}_\zeta[\|\mathbf{W_C}(\check{\mathbf{S}}_\mathbf{C}\mathbf{W_C})^\dagger\check{\mathbf{S}}_\mathbf{C}(\mathbf{I}_n - \mathbf{W_C}\mathbf{W_C}^\top)\mathbf{X}\mathbf{V_R}\mathbf{V_R}^\top\|_F^2] \le \mathbb{E}_\zeta[\|\mathbf{W_C}(\check{\mathbf{S}}_\mathbf{C}\mathbf{W_C})^\dagger\check{\mathbf{S}}_\mathbf{C}(\mathbf{I}_n - \mathbf{W_C}\mathbf{W_C}^\top)\mathbf{X}\|_F^2]$$
$$= O\left(\frac{\log n}{\operatorname{loglog} n}\sqrt{\frac{\theta_{\max}^3(\mathbf{C})c^3}{m_c^3}}\right)\|(\mathbf{I}_n - \mathbf{W_C}\mathbf{W_C}^\top)\mathbf{X}\|_F^2 + \left\|\operatorname{diag}\left\{\frac{(\ell_i(\mathbf{C}))^{1/2}}{\sqrt{m_c\pi_{\mathbf{C},i}}}\right\}_{i=1}^n(\mathbf{I}_n - \mathbf{W_C}\mathbf{W_C}^\top)\mathbf{X}\right\|_F^2$$
$$= O\left(\frac{\log n}{\operatorname{loglog} n}\sqrt{\frac{\theta_{\max}^3(\mathbf{C})c^3}{m_c^3}}\right)\|\mathbf{C}\mathbf{U}_{\mathrm{CUR}}\mathbf{R} - \mathbf{X}\|_F^2 + \left\|\operatorname{diag}\left\{\frac{(\ell_i(\mathbf{C}))^{1/2}}{\sqrt{m_c\pi_{\mathbf{C},i}}}\right\}_{i=1}^n(\mathbf{I}_n - \mathbf{W_C}\mathbf{W_C}^\top)\mathbf{X}\right\|_F^2$$
$$= O\left(\frac{\log n}{\operatorname{loglog} n}\sqrt{\frac{\theta_{\max}^3(\mathbf{C})c^3}{m_c^3}} + \frac{\theta_{\max}(\mathbf{C})c}{m_c}\right)\|\mathbf{C}\mathbf{U}_{\mathrm{CUR}}\mathbf{R} - \mathbf{X}\|_F^2$$

and

$$\mathbb{E}_\zeta[\|\mathbf{W_C}\mathbf{W_C}^\top\mathbf{X}(\mathbf{I}_p - \mathbf{V_R}\mathbf{V_R}^\top)\check{\mathbf{S}}_\mathbf{R}^\top(\mathbf{V_R}^\top\check{\mathbf{S}}_\mathbf{R}^\top)^\dagger\mathbf{V_R}^\top\|_F^2] = \mathbb{E}_\zeta[\|\mathbf{V_R}(\check{\mathbf{S}}_\mathbf{R}\mathbf{V_R})^\dagger\check{\mathbf{S}}_\mathbf{R}(\mathbf{I}_p - \mathbf{V_R}\mathbf{V_R}^\top)\mathbf{X}^\top\mathbf{W_C}\mathbf{W_C}^\top\|_F^2]$$
$$\le \mathbb{E}_\zeta[\|\mathbf{V_R}(\check{\mathbf{S}}_\mathbf{R}\mathbf{V_R})^\dagger\check{\mathbf{S}}_\mathbf{R}(\mathbf{I}_p - \mathbf{V_R}\mathbf{V_R}^\top)\mathbf{X}^\top\|_F^2]$$
$$= O\left(\frac{\log p}{\operatorname{loglog} p}\sqrt{\frac{\theta_{\max}^3(\mathbf{R}^\top)r^3}{m_r^3}}\right)\|(\mathbf{I}_p - \mathbf{V_R}\mathbf{V_R}^\top)\mathbf{X}^\top\|_F^2 + \left\|\operatorname{diag}\left\{\frac{(\ell_j(\mathbf{R}^\top))^{1/2}}{\sqrt{m_r\pi_{\mathbf{R},j}}}\right\}_{j=1}^p(\mathbf{I}_p - \mathbf{V_R}\mathbf{V_R}^\top)\mathbf{X}^\top\right\|_F^2$$
$$= O\left(\frac{\log p}{\operatorname{loglog} p}\sqrt{\frac{\theta_{\max}^3(\mathbf{R}^\top)r^3}{m_r^3}}\right)\|\mathbf{C}\mathbf{U}_{\mathrm{CUR}}\mathbf{R} - \mathbf{X}\|_F^2 + \left\|\operatorname{diag}\left\{\frac{(\ell_j(\mathbf{R}^\top))^{1/2}}{\sqrt{m_r\pi_{\mathbf{R},j}}}\right\}_{j=1}^p(\mathbf{I}_p - \mathbf{V_R}\mathbf{V_R}^\top)\mathbf{X}^\top\right\|_F^2$$
$$= O\left(\frac{\log p}{\operatorname{loglog} p}\sqrt{\frac{\theta_{\max}^3(\mathbf{R}^\top)r^3}{m_r^3}} + \frac{\theta_{\max}(\mathbf{R}^\top)r}{m_r}\right)\|\mathbf{C}\mathbf{U}_{\mathrm{CUR}}\mathbf{R} - \mathbf{X}\|_F^2,$$

where we apply the inequalities

$$\|(\mathbf{W_C}\mathbf{W_C}^\top - \mathbf{I}_n)\mathbf{X}\|_F^2 \le \|\mathbf{C}\mathbf{U}_{\mathrm{CUR}}\mathbf{R} - \mathbf{X}\|_F^2,\ \ \|\mathbf{X}(\mathbf{V_R}\mathbf{V_R}^\top - \mathbf{I}_p)\|_F^2 \le \|\mathbf{C}\mathbf{U}_{\mathrm{CUR}}\mathbf{R} - \mathbf{X}\|_F^2.$$

Recalling Theorems 3.1 and 4.4 again, we similarly obtain,

$$\mathbb{E}_\zeta[\|\mathbf{W_C}(\check{\mathbf{S}}_\mathbf{C}\mathbf{W_C})^\dagger\check{\mathbf{S}}_\mathbf{C}(\mathbf{I}_n - \mathbf{W_C}\mathbf{W_C}^\top)\mathbf{X}(\mathbf{I}_p - \mathbf{V_R}\mathbf{V_R}^\top)\check{\mathbf{S}}_\mathbf{R}^\top(\mathbf{V_R}^\top\check{\mathbf{S}}_\mathbf{R}^\top)^\dagger\mathbf{V_R}^\top\|_F^2]$$
$$= O\left(\frac{\log n}{\operatorname{loglog} n}\sqrt{\frac{\theta_{\max}^3(\mathbf{C})c^3}{m_c^3}}\right)\mathbb{E}_\zeta[\|\mathbf{X}(\mathbf{I}_p - \mathbf{V_R}\mathbf{V_R}^\top)\check{\mathbf{S}}_\mathbf{R}^\top(\mathbf{V_R}^\top\check{\mathbf{S}}_\mathbf{R}^\top)^\dagger\mathbf{V_R}^\top\|_F^2]$$
$$+ \mathbb{E}_\zeta\left[\left\|\operatorname{diag}\left\{\frac{(\ell_i(\mathbf{C}))^{1/2}}{\sqrt{m_c\pi_{\mathbf{C},i}}}\right\}_{i=1}^n(\mathbf{I}_n - \mathbf{W_C}\mathbf{W_C}^\top)\mathbf{X}(\mathbf{I}_p - \mathbf{V_R}\mathbf{V_R}^\top)\check{\mathbf{S}}_\mathbf{R}^\top(\mathbf{V_R}^\top\check{\mathbf{S}}_\mathbf{R}^\top)^\dagger\mathbf{V_R}^\top\right\|_F^2\right]$$
$$= O\left(\frac{\log n}{\operatorname{loglog} n}\sqrt{\frac{\theta_{\max}^3(\mathbf{C})c^3}{m_c^3}}\right)\mathbb{E}_\zeta[\|\mathbf{X}(\mathbf{I}_p - \mathbf{V_R}\mathbf{V_R}^\top)\check{\mathbf{S}}_\mathbf{R}^\top(\mathbf{V_R}^\top\check{\mathbf{S}}_\mathbf{R}^\top)^\dagger\mathbf{V_R}^\top\|_F^2]$$
$$+ \frac{4\theta_{\max}(\mathbf{C})c}{m_c}\mathbb{E}_\zeta[\|\mathbf{X}(\mathbf{I}_p - \mathbf{V_R}\mathbf{V_R}^\top)\check{\mathbf{S}}_\mathbf{R}^\top(\mathbf{V_R}^\top\check{\mathbf{S}}_\mathbf{R}^\top)^\dagger\mathbf{V_R}^\top\|_F^2].$$

Applying Theorems 3.1 and 4.4 twice further yields

$$\mathbb{E}_\zeta[\|\mathbf{W_C}(\check{\mathbf{S}}_\mathbf{C}\mathbf{W_C})^\dagger\check{\mathbf{S}}_\mathbf{C}(\mathbf{I}_n - \mathbf{W_C}\mathbf{W_C}^\top)\mathbf{X}(\mathbf{I}_p - \mathbf{V_R}\mathbf{V_R}^\top)\check{\mathbf{S}}_\mathbf{R}^\top(\mathbf{V_R}^\top\check{\mathbf{S}}_\mathbf{R}^\top)^\dagger\mathbf{V_R}^\top\|_F^2]$$

$$= O\left(\frac{\log p}{\operatorname{loglog} p}\sqrt{\frac{\theta_{\max}^3(\mathbf{R}^\top)r^3}{m_r^3}}\right)\mathbb{E}_\zeta[\|\mathbf{W_C}(\check{\mathbf{S}}_\mathbf{C}\mathbf{W_C})^\dagger\check{\mathbf{S}}_\mathbf{C}(\mathbf{I}_n-\mathbf{W_C}\mathbf{W_C^\top})\mathbf{X}\|_F^2]$$

$$+\mathbb{E}_\zeta\left[\left\|\mathbf{W_C}(\check{\mathbf{S}}_\mathbf{C}\mathbf{W_C})^\dagger\check{\mathbf{S}}_\mathbf{C}(\mathbf{I}_n-\mathbf{W_C}\mathbf{W_C^\top})\mathbf{X}(\mathbf{I}_p-\mathbf{V_R}\mathbf{V_R^\top})\operatorname{diag}\left\{\frac{(\ell_j(\mathbf{R}^\top))^{1/2}}{\sqrt{m_r\pi_{\mathbf{R},j}}}\right\}_{j=1}^p\right\|_F^2\right]$$

$$= O\left(\frac{\log p}{\operatorname{loglog} p}\sqrt{\frac{\theta_{\max}^3(\mathbf{R}^\top)r^3}{m_r^3}}\right)\mathbb{E}_\zeta[\|\mathbf{W_C}(\check{\mathbf{S}}_\mathbf{C}\mathbf{W_C})^\dagger\check{\mathbf{S}}_\mathbf{C}(\mathbf{I}_n-\mathbf{W_C}\mathbf{W_C^\top})\mathbf{X}\|_F^2]$$

$$+\frac{4\theta_{\max}(\mathbf{R}^\top)r}{m_r}\mathbb{E}_\zeta[\|\mathbf{W_C}(\check{\mathbf{S}}_\mathbf{C}\mathbf{W_C})^\dagger\check{\mathbf{S}}_\mathbf{C}(\mathbf{I}_n-\mathbf{W_C}\mathbf{W_C^\top})\mathbf{X}\|_F^2].$$

Putting the above together, we conclude

$$\operatorname{Var}_\zeta(\check{\mathbf{U}}) = O\left(\frac{\log n}{\operatorname{loglog} n}\sqrt{\frac{\theta_{\max}^3(\mathbf{C})c^3}{m_c^3}}+\frac{\log p}{\operatorname{loglog} p}\sqrt{\frac{\theta_{\max}^3(\mathbf{R}^\top)r^3}{m_r^3}}\right)\|\mathbf{C}\mathbf{U}_{\mathrm{CUR}}\mathbf{R}-\mathbf{X}\|_F^2$$

$$+\left\|\operatorname{diag}\left\{\frac{(\ell_i(\mathbf{C}))^{1/2}}{\sqrt{m_c\pi_{\mathbf{C},i}}}\right\}_{i=1}^n(\mathbf{I}_n-\mathbf{W_C}\mathbf{W_C^\top})\mathbf{X}\right\|_F^2+\left\|\operatorname{diag}\left\{\frac{(\ell_j(\mathbf{R}^\top))^{1/2}}{\sqrt{m_r\pi_{\mathbf{R},j}}}\right\}_{j=1}^p(\mathbf{I}_p-\mathbf{V_R}\mathbf{V_R^\top})\mathbf{X}^\top\right\|_F^2$$

$$+O\left(\sqrt[4]{\frac{(\log n)^2\theta_{\max}^3(\mathbf{C})c^3}{(\operatorname{loglog} n)^2m_c^3}}\right)\|\mathbf{C}\mathbf{U}_{\mathrm{CUR}}\mathbf{R}-\mathbf{X}\|_F\left\|\operatorname{diag}\left\{\frac{(\ell_j(\mathbf{R}^\top))^{1/2}}{\sqrt{m_r\pi_{\mathbf{R},j}}}\right\}_{j=1}^p(\mathbf{I}_p-\mathbf{V_R}\mathbf{V_R^\top})\mathbf{X}^\top\right\|_F$$

$$+O\left(\sqrt[4]{\frac{(\log p)^2\theta_{\max}^3(\mathbf{R}^\top)r^3}{(\operatorname{loglog} p)^2m_r^3}}\right)\|\mathbf{C}\mathbf{U}_{\mathrm{CUR}}\mathbf{R}-\mathbf{X}\|_F\left\|\operatorname{diag}\left\{\frac{(\ell_i(\mathbf{C}))^{1/2}}{\sqrt{m_c\pi_{\mathbf{C},i}}}\right\}_{i=1}^n(\mathbf{I}_n-\mathbf{W_C}\mathbf{W_C^\top})\mathbf{X}\right\|_F$$

$$+2\left\|\operatorname{diag}\left\{\frac{(\ell_j(\mathbf{R}^\top))^{1/2}}{\sqrt{m_r\pi_{\mathbf{R},j}}}\right\}_{j=1}^p(\mathbf{I}_p-\mathbf{V_R}\mathbf{V_R^\top})\mathbf{X}^\top\right\|_F\left\|\operatorname{diag}\left\{\frac{(\ell_i(\mathbf{C}))^{1/2}}{\sqrt{m_c\pi_{\mathbf{C},i}}}\right\}_{i=1}^n(\mathbf{I}_n-\mathbf{W_C}\mathbf{W_C^\top})\mathbf{X}\right\|_F.$$

This completes the proof.

G.2 Special cases of Theorem G.1 under leverage score sampling and SRHT

In this section, we establish fine-grained accuracy characterizations for fast CUR under leverage score sampling and the SRHT. The proofs follow directly by combining the arguments from Theorem 4.4 and Corollaries 4.5 and 4.7, and are therefore omitted.

Corollary G.2 (Bias–variance characterizations for fast CUR under leverage score sampling)**.** *Under the settings and notations of Theorem G.1, standard random sampling matrices $\mathbf{S_C}\in\mathbb{R}^{m_c\times n}$ with sampling probabilities $\pi_{\mathbf{C},i}\in[\ell_i(\mathbf{C})/(c\theta_{\max}(\mathbf{C})),\ell_i(\mathbf{C})/(c\theta_{\min}(\mathbf{C}))]$ with $\theta_{\min}(\mathbf{C})\in[1/2,1]$ and $\mathbf{S_R}\in\mathbb{R}^{m_r\times p}$ with sampling probabilities $\pi_{\mathbf{R},i}\in[\ell_i(\mathbf{R}^\top)/(r\theta_{\max}(\mathbf{R}^\top)),\ell_i(\mathbf{R}^\top)/(r\theta_{\min}(\mathbf{R}^\top))]$ with $\theta_{\min}(\mathbf{R}^\top)\in[1/2,1]$, there exists $C>0$, $\alpha_c>0$, $\alpha_r>0$, independent of n,p,c,r so that for $m_c\geq C\theta_{\max}(\mathbf{C})c\log(c/\delta)$, $m_r\geq C\theta_{\max}(\mathbf{R}^\top)r\log(r/\delta)$, $\delta\leq\min\{n^{-(3+\alpha_c)},p^{-(3+\alpha_r)}\}$, when conditioned on an event ζ that holds with the probability at least $1-\delta$, the standard sampled estimator $\tilde{\mathbf{U}}$ in* (33) *satisfies*

$$\operatorname{Bias}_\zeta(\tilde{\mathbf{U}})=\epsilon^2\cdot L(\mathbf{U}_{\mathrm{CUR}}),\ \operatorname{Var}_\zeta(\tilde{\mathbf{U}})=\Delta_{\mathrm{CUR}}+\epsilon\cdot L(\mathbf{U}_{\mathrm{CUR}}),$$

with $\epsilon = O(\sqrt{(\log n/\operatorname{loglog} n)^2\cdot\theta_{\max}^3(\mathbf{C})c^3/m_c^3}+\epsilon_{\theta,c}\cdot\theta_{\max}(\mathbf{C})c/m_c+\sqrt{(\log p/\operatorname{loglog} p)^2\cdot\theta_{\max}^3(\mathbf{R}^\top)r^3/m_r^3}+\epsilon_{\theta,r}\cdot\theta_{\max}(\mathbf{R}^\top)r/m_r)$, $\epsilon_{\theta,c}=\max\{\theta_{\min}^{-1}(\mathbf{C})-1,1-\theta_{\max}^{-1}(\mathbf{C})\}$, *and* $\epsilon_{\theta,r}=\max\{\theta_{\min}^{-1}(\mathbf{R}^\top)-1,1-\theta_{\max}^{-1}(\mathbf{R}^\top)\}$. *Here,* $(\theta_{\max}(\mathbf{C}),\theta_{\min}(\mathbf{C}))$ *and* $(\theta_{\max}(\mathbf{R}^\top),\theta_{\min}(\mathbf{R}^\top))$ *in Definition 2.3 denote the maximum and minimum importance sampling approximation factors for* $\mathbf{C}$ *and* $\mathbf{R}^\top$*, respectively.*

Corollary G.3 (Bias–variance characterizations for fast CUR decomposition under SRHT)**.** *Under the setting and notations of Theorem G.1, let $\tilde{\mathbf{C}}_{\mathrm{SRHT}}=\mathbf{S_C}\mathbf{C}_{\mathrm{HD}}\in\mathbb{R}^{m_c\times c}$ the SRHT of $\mathbf{C}$ with $\mathbf{C}_{\mathrm{HD}}=\mathbf{H}_n\mathbf{D}_n\mathbf{C}/\sqrt{n}$, $\tilde{\mathbf{R}}_{\mathrm{SRHT}}^\top=\mathbf{S_R}\mathbf{R}_{\mathrm{HD}}^\top\in\mathbb{R}^{m_r\times r}$ the SRHT of $\mathbf{R}^\top$ with $\mathbf{R}_{\mathrm{HD}}^\top=\mathbf{H}_p\mathbf{D}_p\mathbf{R}^\top/\sqrt{p}$, $\tilde{\mathbf{X}}_{\mathrm{SRHT}_{\mathbf{C},\mathbf{R}}}=\mathbf{S_C}\mathbf{H}_n\mathbf{D}_n\mathbf{X}\mathbf{D}_p\mathbf{H}_p\mathbf{S_R}^\top/\sqrt{np}$. Assume that $\min_{i,\|\mathbf{c}_{\mathrm{HD},i}\|>0}\|\mathbf{c}_{\mathrm{HD},i}\|^2/\|\mathbf{C}_{\mathrm{HD}}\|_F^2\geq n^{-\alpha_c}$ and $\min_{j,\|\mathbf{r}_{\mathrm{HD},j}\|>0}\|\mathbf{r}_{\mathrm{HD},j}\|^2/\|\mathbf{R}_{\mathrm{HD}}\|_F^2\geq p^{-\alpha_r}$ for some constants $\alpha_c>0$ and $\alpha_r>0$.*

Then, there exists $C > 0$, $\max\{n\exp(-c), p\exp(-r)\} < \delta < \min\{n^{-(3+\alpha_c)}, p^{-(3+\alpha_r)}\}$ *such that for* $m_c \geq Cc\log(c/\delta)$, $m_r \geq Cr\log(r/\delta)$, *when conditioned on an event* ζ *that holds with the probability at least* $1-\delta$, *the sampled estimator* $\tilde{\mathbf{U}} = \tilde{\mathbf{C}}^{\dagger}_{\mathrm{SRHT}}\tilde{\mathbf{X}}_{\mathrm{SRHT}_{\mathbf{C},\mathbf{R}}}\tilde{\mathbf{R}}^{\dagger}_{\mathrm{SRHT}}$ *satisfies*

$$\mathrm{Bias}_{\zeta}(\tilde{\mathbf{U}}) = \epsilon^2 \cdot L(\mathbf{U}_{\mathrm{CUR}}),\ \mathrm{Var}_{\zeta}(\tilde{\mathbf{U}}) = \Delta_{\mathrm{CUR}} + \epsilon \cdot L(\mathbf{U}_{\mathrm{CUR}}),$$

with $\epsilon = O(\sqrt{(\log n/\operatorname{loglog} n)^2 \cdot c^3/m_c^3} + \sqrt{c\log(n/\delta)}/m_c + \sqrt{(\log p/\operatorname{loglog} p)^2 \cdot r^3/m_r^3} + \sqrt{r\log(p/\delta)}/m_r)$, *where we define* $\Delta_{\mathrm{CUR}} = (\Delta_{\mathrm{CUR},1} + \Delta_{\mathrm{CUR},2})^2 + \sqrt{\epsilon \cdot L(\mathbf{U}_{\mathrm{CUR}})}(\Delta_{\mathrm{CUR},1} + \Delta_{\mathrm{CUR},2})$, *with* $\Delta_{\mathrm{CUR},1} = \|\operatorname{diag}\{\sqrt{\ell_i(\mathbf{C}_{\mathrm{HD}})/m_c\pi_{\mathbf{C},i}}\}_{i=1}^{n}\mathbf{B}_1\|_F$ *and* $\Delta_{\mathrm{CUR},2} = \|\operatorname{diag}\{\sqrt{\ell_j(\mathbf{R}_{\mathrm{HD}}^{\top})/m_r\pi_{\mathbf{R},j}}\}_{j=1}^{p}\mathbf{B}_2\|_F$ *for* $\mathbf{B}_1 = \mathbf{H}_n\mathbf{D}_n(\mathbf{I}_n - \mathbf{W}_{\mathbf{C}}\mathbf{W}_{\mathbf{C}}^{\top})\mathbf{X}/\sqrt{n} \in \mathbb{R}^{n\times p}$ *and* $\mathbf{B}_2 = \mathbf{H}_p\mathbf{D}_p(\mathbf{I}_p - \mathbf{V}_{\mathbf{R}}\mathbf{V}_{\mathbf{R}}^{\top})\mathbf{X}^{\top}/\sqrt{p} \in \mathbb{R}^{p\times n}$. *Here,* $\mathbf{S}_{\mathbf{C}} \in \mathbb{R}^{m_c\times n}$ *and* $\mathbf{S}_{\mathbf{R}} \in \mathbb{R}^{m_r\times p}$ *are the uniform random sampling matrices applied to* $\mathbf{C}_{\mathrm{HD}}$ *and* $\mathbf{R}_{\mathrm{HD}}^{\top}$, *respectively.*

Table 2 summarizes the existing upper bounds on the bias of fast CUR decomposition established in prior work under uniform sampling, leverage score sampling, and SRHT [57, 59], and contrasts them with our results on debiased fast CUR decomposition. We see that the proposed debiased approach consistently achieves a substantially smaller bias across all three sampling schemes.

Table 2: Bias characterizations for classical fast CUR established in previous efforts versus that for the proposed debiased fast CUR in Theorem G.1, under uniform sampling (**UNI**), (exact/ approximated) leverage score sampling (**Lev**) with $\theta_{\min}(\mathbf{C})/(2\theta_{\min}(\mathbf{C}) - 1) \leq \theta_{\max}(\mathbf{C}) \leq 1/(1 - \sqrt{\theta_{\max}(\mathbf{C})c/m_c})$ and $\theta_{\min}(\mathbf{R}^{\top})/(2\theta_{\min}(\mathbf{R}^{\top}) - 1) \leq \theta_{\max}(\mathbf{R}^{\top}) \leq 1/(1 - \sqrt{\theta_{\max}(\mathbf{R}^{\top})r/m_r})$, and **SRHT** with $c > \log(n/\delta)$ and $r > \log(p/\delta)$. We report the regime $m_c/m_r \leq \min\{\nu_1, \nu_2, \nu_3, \nu_4, \nu_5, \nu_6\}$, where $\nu_1 = c/r$, $\nu_2 = c\theta_{\max}(\mathbf{C})/(r\theta_{\max}(\mathbf{R}^{\top}))$, $\nu_3 = c\theta_{\max}(\mathbf{C})(\log n\log\log p)^{2/3}/(r\theta_{\max}(\mathbf{R}^{\top})(\log p\log\log n)^{2/3})$, $\nu_4 = \sqrt{cn(\log n)^2/(rp(\log p)^2)}$, $\nu_5 = c(\log n\log\log p)^{2/3}/(r(\log p\log\log n)^{2/3})$, $\nu_6 = \sqrt{c\log(n/\delta)/(r\log(p/\delta))}$; the complementary regime, where the ratio m_c/m_r exceeds this minimum, yields analogous conclusions and is omitted for brevity.

	UNI	**Lev**	**SRHT**
[57]	$O(\theta_{\max}(\mathbf{C})^2c^2q/m_c^2)$	$O(c^2q/m_c^2)$	$O(cn(\log n)^2/m_c^2)$
[59]	—	$O(\rho c/m_c)$	$O(\rho c/m_c)$
This work	$O\Big(\frac{(\log n)^2\theta_{\max}^3(\mathbf{C})c^3}{(\operatorname{loglog} n)^2m_c^3}\Big)$	$O\Big(\frac{(\log n)^2c^3}{(\operatorname{loglog} n)^2m_c^3}\Big)$	$O\Big(\frac{(\log n)^2c^3}{\operatorname{loglog} n)^2m_c^3} + \frac{c\log(n/\delta)}{m_c^2}\Big)$

[a] Here, q in the error bounds of [57] is defined as $q = \min\{n, p\}$. Moreover, the bias error bounds in [59] are established under the additional condition $\rho = \big(\|(\mathbf{I}_n - \mathbf{C}\mathbf{C}^{\dagger})\mathbf{X}\mathbf{R}^{\dagger}\mathbf{R}\|_F + \|\mathbf{C}\mathbf{C}^{\dagger}\mathbf{X}(\mathbf{I}_p - \mathbf{R}^{\dagger}\mathbf{R})\|_F\big)^2/\|\mathbf{C}\mathbf{C}^{\dagger}\mathbf{X}\mathbf{R}^{\dagger}\mathbf{R} - \mathbf{X}\|_F^2 \geq \epsilon$.

H Additional numerical experiments and implementation details

In this section, we first provide implementation details for the numerical experiments on subsampled OLS and fast CUR decomposition in Appendices H.1 and H.2. We then present additional numerical results for subsampled OLS in Appendix H.3, followed by numerical results for fast CUR decomposition in Appendix H.4.

H.1 Sketching matrices

Given a data matrix $\mathbf{X} \in \mathbb{R}^{n\times p}$, leverage scores are estimated following the procedure described in [24]. The LESS-uniform sketching matrix is generated according to the construction detailed in [16, Section E.1]. Implementation specifics, including those for the SRHT, follow [16].

Experiments were conducted on a server with an AMD EPYC 7452 32-Core Processor, 256GB RAM, and NVIDIA GeForce RTX 3090 GPUs. Code is publicly available at `https://github.com/chengmeiniu/debiased-oblique-projections.git`.

H.2 Datasets

The Flight Delay dataset used in Appendix H.3 is obtained from the U.S. Department of Transportation. The dataset contains 539747 U.S. domestic weekday flights in January 2025, with five recorded variables per flight: arrival delay (difference in minutes between scheduled and actual arrival time, and early arrivals show negative numbers), arrival taxi in time (in minutes), departure taxi out time (in minutes), departure delays (difference in minutes between scheduled and actual departure time, and early departures show negative numbers), and computer reservation system based elapsed time of the flight (in minutes; a measure for the distance of the flight). Following [40], we take arrival delay as the response and the remaining four variables as predictors, and further augment the design with all quadratic terms and pairwise interactions, yielding 14 predictors in total. All variables are standardized to zero mean and unit variance. In the experiments, we randomly select $n = 2^{13}$ samples from the full dataset to construct the regression design matrix.

The MSD (Million Song Year Prediction) dataset consists of 515,345 songs released between 1922 and 2011. Each song is represented by multiple audio segments, and each segment is described by 12 timbre features that capture perceptual properties such as brightness and spectral flatness. Our goal is to predict the release year using the complete set of timbre-based descriptors. Following [40], we model the logarithm of the release year as the response variable and use all derived timbre features as predictors. All predictors are standardized to zero mean and unit variance. In the experiments, we randomly sample $n = 2^{14}$ observations from the full dataset to form the regression design matrix.

The CIFAR-10 dataset used in Appendix H.4 contains 60,000 images of size $32 \times 32 \times 3$. Each image is flattened into a vector. We construct the data matrix using $n = 2^{13}$ samples and $p = 2^{11}$ features.

The ImageNet dataset used in Appendix H.4 is based on ImageNet-64, a downsampled version of ImageNet. We use images from `Imagenet64_train_part1`, specifically `train_data_batch_1`. Each image has resolution $64 \times 64 \times 3$ and is flattened into a vector representation. We construct the data matrix using $n = 2^{14}$ samples and $p = 2^{11}$ features.

H.3 Additional numerical experiments for subsampled OLS

In this section, we use the Flight Delay (Airline) dataset, obtained from the website of the U.S. Department of Transportation[5], to provide additional numerical evidence supporting our theoretical findings on classical and debiased subsampled OLS. The data matrix $\mathbf{X} \in \mathbb{R}^{n \times p}$ is constructed from the Airline dataset, and $\mathbf{y} \in \mathbb{R}^n$ denotes the corresponding response vector.

Figure 2 examines how the sketch size m affects the bias and variance of subsampled OLS on the Airline dataset. We report the bias by $(L(\hat{\mathbb{E}}[\hat{\boldsymbol{\beta}}]) - L(\boldsymbol{\beta}_{\mathrm{OLS}}))/L(\boldsymbol{\beta}_{\mathrm{OLS}})$ and the variance by $(\hat{\mathbb{E}}[L(\hat{\boldsymbol{\beta}})] - L(\boldsymbol{\beta}_{\mathrm{OLS}}))/L(\boldsymbol{\beta}_{\mathrm{OLS}})$, where $L(\boldsymbol{\beta}) = \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|^2$ is as in Equation (5). We compare uniform sampling (**UNI**), approximate leverage score sampling (**Lev**), **SRHT** (see Definition 4.6), and LESS [16, 29], together with their debiased variants **DUNI**, **DLev**, and **DSRHT**, using exact leverage scores[6].

The results on the Airline dataset in Figure 2 closely match those on the MSD dataset in Figure 1. UNI is the most biased, while DUNI substantially reduces its bias and approaches the comparable bias levels of Lev, DLev, SRHT, DSRHT, and LESS. Debiasing has almost no visible effect on approximate leverage score sampling and SRHT, consistent with Corollaries 4.5 and 4.7. For variance, approximate leverage sampling, SRHT, their debiased variants, and LESS are comparable and uniformly outperform UNI and DUNI. Moreover, the debiased sampling schemes introduce essentially no variance increase relative to their corresponding standard counterparts. These results further support that debiasing improves accuracy without sacrificing variance. Table 3 shows that standard UNI, Lev, and SRHT are faster than their debiased counterparts on both datasets, reflecting the additional cost of computing exact leverage scores for debiasing. All debiased variants nevertheless

[5] Airline On-Time Performance Data, `https://www.transtats.bts.gov/DL_SelectFields.asp?Table_ID=236`.

[6] We use exact leverage scores in (1) to isolate the statistical effect of debiasing. For DLev, reusing the approximate scores that define the sampling probabilities makes the correction scalar, which cancels across the pseudoinverse and yields $\check{\mathbf{P}} = \tilde{\mathbf{P}}$. Likewise, the nearly uniform leverage scores of $\mathbf{H}_n\mathbf{D}_n\mathbf{X}/\sqrt{n}$ make the DSRHT correction approximately scalar. Using exact scores avoids these cancellations and enables a controlled evaluation of debiasing. In practice, they can be replaced by efficiently computed approximations [24].

remain faster than LESS. With accuracy comparable to their debiased counterparts and shorter runtimes, standard Lev and SRHT offer a more favorable accuracy–efficiency trade-off.

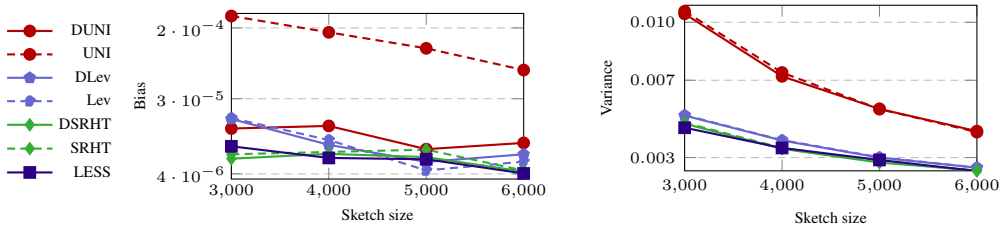

Figure 2: Bias and variance as functions of the sketch size m, comparing debiased sampling (in **solid** lines) and standard sampling (in **dashed** lines) on Airline dataset. **DUNI**, **DLev**, **DSRHT** are the corresponding debiased versions. Expectations are estimated from 500 independent runs.

Table 3: Wall-clock time (s) versus sketch size m for various methods on Airline and MSD datasets. Results are obtained by averaging over 500 independent runs.

Dataset	**Method**	$m = 3000$	$m = 5000$	**Dataset**	Method	$m = 5000$	$m = 7000$
Airline	**UNI**	**0.0013**	**0.0017**	**MSD**	**UNI**	**0.0133**	**0.0175**
	DUNI	0.0026	0.0030		**DUNI**	0.0467	0.0515
	Lev	**0.0021**	**0.0028**		**Lev**	**0.0269**	**0.0315**
	DLev	0.0035	0.0042		**DLev**	0.0602	0.0646
	SRHT	**0.2596**	**0.3306**		**SRHT**	**0.4986**	**0.5964**
	DSRHT	0.3532	0.3544		**DSRHT**	0.7836	0.7870
	LESS	0.3613	0.3694		**LESS**	0.8031	0.8294

H.4 Numerical results for fast CUR decomposition

In this section, we present numerical experiments for fast CUR decomposition to support our theoretical findings on both classical and debiased CUR estimators. We evaluate CUR approximations on matrices $\mathbf{X}$ constructed from the CIFAR-10 dataset [37] and the ImageNet dataset [13].

In Figures 3 and 4, we study the effect of sketch size m on the bias and variance of fast CUR decomposition on CIFAR-10 and ImageNet datasets. We measure the bias by $\|\hat{\mathbb{E}}[\mathbf{X} - \mathbf{CUR}]\|_F^2 / \|\mathbf{X}\|_F^2$ and the variance by $\hat{\mathbb{E}}[\|\mathbf{X} - \mathbf{CUR}\|_F^2] / \|\mathbf{X}\|_F^2$. We compare uniform sampling (**UNI**), approximate leverage score sampling (**Lev**), and **SRHT** (see Definition 4.6), together with their debiased counterparts **DUNI**, **DLev**, and **DSRHT**. The debiased variants use exact leverage scores as explained in Footnote 6. For each method, we first sample columns and rows of $\mathbf{X}$ to form $\mathbf{C} \in \mathbb{R}^{n \times c}$ and $\mathbf{R} \in \mathbb{R}^{r \times p}$, using column-norm and row-norm sampling probabilities, respectively. We fix $c = 30$ and $r = 60$ for CIFAR-10, and $c = 40$ and $r = 100$ for ImageNet. We then compute the core matrix $\mathbf{U}$ in (11) and (12), with the standard sampling matrices $\mathbf{S_C} \in \mathbb{R}^{m_c \times n}$ and $\mathbf{S_R} \in \mathbb{R}^{m_r \times p}$, and the debiased sampling matrices $\check{\mathbf{S}}_\mathbf{C} \in \mathbb{R}^{m_c \times n}$ and $\check{\mathbf{S}}_\mathbf{R} \in \mathbb{R}^{m_r \times p}$. Tables 4 and 5 use the same choices of c and r as in Figures 3 and 4.

As shown in Figures 3 and 4, Lev and SRHT, together with their debiased variants, consistently outperform UNI and DUNI in the bias–variance performance, with SRHT achieving the best overall variance both with and without debiasing. Approximate leverage score sampling exhibits nearly identical bias and variance with and without debiasing; the same behavior is observed for SRHT. This is consistent with Corollaries G.2 and G.3. In contrast, DUNI consistently reduces the bias of UNI, moving its performance closer to that of Lev, DLev, SRHT, and DSRHT, while maintaining nearly the same variance. Overall, debiasing improves accuracy with negligible change in variance. The timing results in Table 4 complement these observations. Debiasing incurs additional runtime for all three schemes because of the exact leverage score computation used in these experiments. For Lev and SRHT, this overhead brings little measurable accuracy gain, making their standard versions preferable in terms of the accuracy–efficiency trade-off. Accordingly, we compare standard Lev and SRHT with the OSP methods of [46, Algorithm 2] below.

In Table 5, we compare the accuracy–efficiency trade-off of our method with the OSP methods of [46, Algorithm 2] on CIFAR-10 and ImageNet datasets, where accuracy is measured by $\hat{\mathbb{E}}[\|\mathbf{X} - \mathbf{CUR}\|_F^2/\|\mathbf{X}\|_F^2]$, denoted as variance in Figures 3 and 4. The reported wall-clock time includes the computation of $\mathbf{C}$, $\mathbf{R}$, $\tilde{\mathbf{U}}$ in (33) and $\mathbf{C}\tilde{\mathbf{U}}\mathbf{R}$ for Lev and SRHT, and the computation of $\mathbf{C}$, $\mathbf{R}$, $\mathbf{U}^* = \mathbf{X}_{R,C}^{\dagger}$ and $\mathbf{C}\mathbf{U}^*\mathbf{R}$ for OSP-SRHT and OSP-SS. These OSP methods use randomized column-pivoted QR with oversampling to improve the stability and accuracy of the core matrix estimator $\mathbf{U}^\star = \mathbf{X}_{R,C}^{\dagger}$. Specifically, for OSP, we implement randomized column-pivoted QR using SRHT and sparse sign sketches, denoted by **OSP-SRHT** and **OSP-SS**, respectively. The results show that Lev achieves a better accuracy–efficiency trade-off than both OSP baselines, indicating that our CUR construction effectively balances computational cost and numerical stability. Although SRHT is slower than OSP-SS because SRHT-based sketching is more expensive than sparse sign sketching, it is still more accurate. Moreover, SRHT consistently outperforms OSP-SRHT in terms of the overall accuracy–efficiency trade-off, further demonstrating the advantage of our approach.

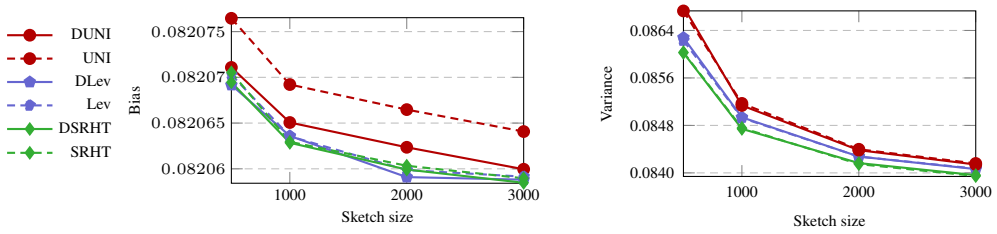

Figure 3: Bias and variance as functions of the sketch size m, comparing debiased sampling (in **solid** lines) and standard sampling (in **dashed** lines) on CIFAR-10 dataset. **DUNI**, **DLev**, **DSRHT** are the corresponding debiased versions. Expectations are estimated from 200 independent runs.

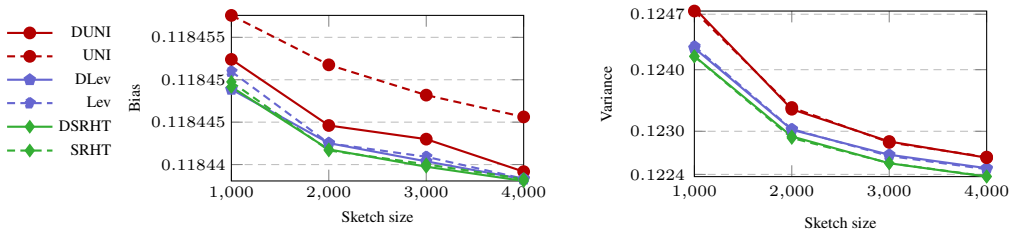

Figure 4: Bias and variance as functions of the sketch size m, comparing debiased sampling (in **solid** lines) and standard sampling (in **dashed** lines) on ImageNet dataset. **DUNI**, **DLev**, **DSRHT** are the corresponding debiased versions. Expectations are estimated from 200 independent runs.

Table 4: Wall-clock time (s) versus sketch size for various methods on CIFAR-10 and ImageNet datasets. We vary the sketch size m_c of $\mathbf{S}_c$ and fix the sketch size $m_r = 500$ of $\mathbf{S}_\mathbf{R}$. Results are obtained by averaging over 200 independent runs.

Dataset	Method	$m_c = 2000$	$m_c = 3000$	Dataset	Method	$m_c = 2000$	$m_c = 3000$
CIFAR-10	**UNI**	**0.0899**	**0.1008**	**ImageNet**	**UNI**	**0.1665**	**0.1774**
	DUNI	0.0956	0.1077		**DUNI**	0.1841	0.1954
	Lev	**0.0908**	**0.1057**		**Lev**	**0.1777**	**0.1869**
	DLev	0.0976	0.1112		**DLev**	0.1920	0.2024
	SRHT	**0.5708**	**0.6707**		**SRHT**	**0.9424**	**1.0695**
	DSRHT	1.0158	1.0791		**DSRHT**	1.9502	1.9952

Table 5: Wall-clock time (s) versus relative error for Lev, SRHT, OSP-SRHT, and OSP-SS on CIFAR-10 and ImageNet datasets. Expectations are estimated from 200 independent runs. For **Lev** and **SRHT**, we vary the sketch size m_c of $\mathbf{S}_c$ and fix the sketch size $m_r = 500$ of $\mathbf{S}_\mathbf{R}$, with $m_c \in \{2000, 3000, 4000\}$. For **OSP-SRHT** and **OSP-SS**, we vary the numbers of selected columns and rows of $\mathbf{X}$: on CIFAR-10, we use $c \in \{30, 50, 80\}$ and $r = c + r_0$ with row oversampling parameter $r_0 = 30$; on ImageNet, we use $c \in \{40, 70, 100\}$ and $r = c + r_0$ with $r_0 = 70$.

Dataset	**Lev**		**SRHT**		**OSP-SRHT**		**OSP-SS**	
	Time(s)	Error	Time(s)	Error	Time(s)	Error	Time(s)	Error
CIFAR-10	0.0908	0.0843	0.5708	0.0842	0.9493	0.1042	0.1918	0.1044
	0.1057	0.0841	0.6707	0.0840	1.0516	0.0946	0.2198	0.0950
	0.1152	0.0840	0.7477	0.0839	1.0695	0.0885	0.2438	0.0888
ImageNet	0.1777	0.1230	0.9424	0.1229	1.2669	0.1487	0.2307	0.1491
	0.1869	0.1226	1.0695	0.1225	1.3560	0.1405	0.3145	0.1411
	0.1989	0.1224	1.1753	0.1223	1.4259	0.1398	0.3895	0.1405

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: We have clearly stated the contributions and scopes of this work in the abstract and introduction.

   Guidelines:

   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [No].

   Justification: The assumptions underlying our theoretical results are commonly used in randomized numerical linear algebra (RandNLA) and are explicitly stated in the corresponding theorem statements. To keep the presentation concise, we do not further discuss how these assumptions restrict the applicability of our theoretical guarantees.

   Guidelines:

   - The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.

- The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
- The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
- If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
- While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [Yes]

   Justification: All theoretical results have been proved. All proofs are provided in the appendix.

   Guidelines:

   - The answer [N/A] means that the paper does not include theoretical results.
   - All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
   - All assumptions should be clearly stated or referenced in the statement of any theorems.
   - The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
   - Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
   - Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes].

   Justification: Parameter choices and dataset information are provided in the appendix.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.

- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
  (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
  (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
  (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
  (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [Yes].

   Justification: A supplementary zip file containing the main experimental code is provided to reproduce the reported results. One experiment uses the Million Song Year Prediction dataset, which is publicly available from the UCI Machine Learning Repository at `https://archive.ics.uci.edu/dataset/203/yearpredictionmsd`.

   Guidelines:

   - The answer [N/A] means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
   - The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
   - At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
   - Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

   Answer:[Yes].

   Justification: The paper specifies the experimental parameters, including sketch sizes and data dimensions, as well as the dataset details needed to understand the results.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.

- The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
- The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [Yes].

   Justification: The error metrics follow from our bias-variance definition in Definition 4.1, and are commonly used in the related work.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
   - The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
   - The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
   - The assumptions made should be given (e.g., Normally distributed errors).
   - It should be clear whether the error bar is the standard deviation or the standard error of the mean.
   - It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
   - For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
   - If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer:[Yes].

   Justification: The experiments were run on a server equipped with an AMD EPYC 7452 32-Core Processor, 256GB of RAM, and NVIDIA GeForce RTX 3090 GPUs.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes].

   Justification: This work aligns with the NeurIPS code of Ethics.

Guidelines:

- The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
- If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.
- The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [N/A].

    Justification: This work provides a bias-aware analysis of sketching-based estimators, which may advance the statistical understanding of randomized numerical linear algebra. This work is theoretical and algorithmic, and does not lead to any negative societal impact.

    Guidelines:

    - The answer [N/A] means that there is no societal impact of the work performed.
    - If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.
    - Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
    - The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
    - The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
    - If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

    Answer:[N/A].

    Justification: This work is purely theoretical and algorithmic, and does not involve datasets or models with a high risk of misuse.

    Guidelines:

    - The answer [N/A] means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.

- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

    Justification: We test our theoretical results on the publicly available datasets that have been cited in the paper as well.

    Guidelines:

    - The answer [N/A] means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
    - If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
    - For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
    - If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [N/A].

    Justification: This work does not release any new assets.

    Guidelines:

    - The answer [N/A] means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [N/A].

    Justification: This work does not involve crowdsourcing or research with human subjects.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.

- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [N/A].

    Justification: This work does not involve crowdsourcing or research with human subjects, and therefore does not require IRB approval.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
    - We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
    - For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

    Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

    Answer: [N/A]

    Justification: LLMs are not used as part of the core methodology in this work.

    Guidelines:

    - The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
    - Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.